Syntax Literate: Jurnal Ilmiah Indonesia p–ISSN: 2541-0849
e-ISSN: 2548-1398
Vol. 10, No. 3, Maret 2025

PERBANDINGAN TREE BASED MODEL UNTUK KLASIFIKASI TIPE KEPRIBADIAN MBTI

Rani Asriya Effendy*, Irving Vitra Paputungan
Universitas Islam Indonesia, Yogyakarta, Indonesia
Email: 20523191@students.uii.ac.id*, 045230101@uii.ac.id

Abstrak

Penelitian ini menguji penggunaan Model Berbasis Pohon untuk mengklasifikasikan tipe kepribadian *Myers-Briggs Type Indicator* (MBTI). Data dikumpulkan melalui kuesioner yang mengacu pada tes 16Personalities, diikuti dengan langkah-langkah preprocessing untuk menangani nilai yang hilang, memisahkan label MBTI, dan menganalisis ringkasan statistik. Tujuh algoritma, termasuk Decision Tree, Random Forest, Extra Trees, dan Gradient Boosting, digunakan untuk mengevaluasi akurasi klasifikasi di seluruh dimensi MBTI (E-I, S-N, T-F, J-P). Random Forest muncul sebagai model terbaik dengan akurasi 72,5% saat menganalisis pertanyaan dengan korelasi tinggi, menyoroti ketangguhannya dalam mengelola interaksi data. Penelitian ini menekankan efektivitas pembelajaran mesin dalam klasifikasi kepribadian dan memberikan wawasan praktis untuk mengintegrasikan penilaian MBTI ke dalam aplikasi. Penelitian di masa depan menyarankan untuk memanfaatkan set data yang lebih besar dan lebih beragam serta mengeksplorasi integrasi pembelajaran mendalam untuk meningkatkan kinerja model.

Kata Kunci : MBTI, Model Berbasis Pohon, Klasifikasi Kepribadian, Pembelajaran Mesin.

Abstract

This study examines the use of Tree-Based Models for classifying Myers- Briggs Type Indicator (MBTI) personality types. Data was collected through questionnaires referencing the 16Personalities test, followed by preprocessing steps to handle missing values, split MBTI labels, and analyze statistical summaries. Seven algorithms, including Decision Tree, Random Forest, Extra Trees, and Gradient Boosting, were applied to evaluate classification accuracy across MBTI dimensions (E-I, S-N, T-F, J-P). Random Forest emerged as the best model with an accuracy of 72.5% when analyzing high-correlation questions, highlighting its robustness in managing data interactions. This research emphasizes the effectiveness of machine learning in personality classification and provides practical insights for integrating MBTI assessments into applications. Future work suggests leveraging larger, more diverse datasets and exploring deep learning integrations for enhanced model performance.

Keywords: *MBTI, Tree-Based Model, Personality Classification, Machine Learning*

Pendahuluan

Kepribadian memegang peran penting dalam berbagai aspek kehidupan seperti pendidikan, pekerjaan, dan hubungan sosial (Pulungan et al., 2024). WHO (2023) mencatat peningkatan 15% kasus gangguan kesehatan mental sejak pandemi, menunjukkan hubungan erat antara kesehatan mental dan kepribadian. Gallup (2024) melaporkan 60% pekerja mengalami burnout akibat ketidaksesuaian kepribadian dengan tuntutan pekerjaan. Teknologi memungkinkan pengukuran kepribadian yang lebih efisien melalui *Myers-Briggs Type Indicator* (MBTI), yang mengklasifikasikan individu ke dalam 16 tipe berdasarkan empat dimensi utama: *Ekstrovert-Introvert* (E-I), *Sensing-iNtuition* (S-N), *Thinking-Feeling* (T-F), dan *Judging-Perceiving* (J- P) (Fitri et al., 2023; Wahyudi et al., 2022).

MBTI yang sebelumnya diukur melalui kuesioner kini dapat diotomatisasi dengan algoritma machine learning, khususnya Tree-Based Model (Devlin et al., 2019). Beberapa algoritma seperti *Decision Tree, Random Forest, Extra Trees, Gradient Boosting, High Gradient Boosting, Bagging Classifier, dan AdaBoost Classifier* digunakan dalam klasifikasi kepribadian (Ryan et al., 2023). Tree-Based Model memiliki keunggulan dalam menangkap interaksi antar fitur, kemudahan interpretasi, dan tidak memerlukan asumsi distribusi data (Alifah et al., 2024). Penelitian Serrano (2024) menunjukkan performa baik algoritma Decision Tree dan C4.5 dalam klasifikasi MBTI.

Penelitian terdahulu oleh Effendy (2025) Hasil penelitian menunjukkan bahwa algoritma Tree Based Model yang paling akurat adalah Extra Trees dengan nilai akurasi 0.728, sementara algoritma lain seperti Random Forest, XGBoost, AdaBoost, dan Gradient Boosting juga mencapai akurasi di atas 0.70. Hal ini menunjukkan bahwa beberapa algoritma turunan Tree Based Model dapat digunakan secara efektif dalam klasifikasi tipe kepribadian berbasis MBTI (Hadiwandra, 2019).

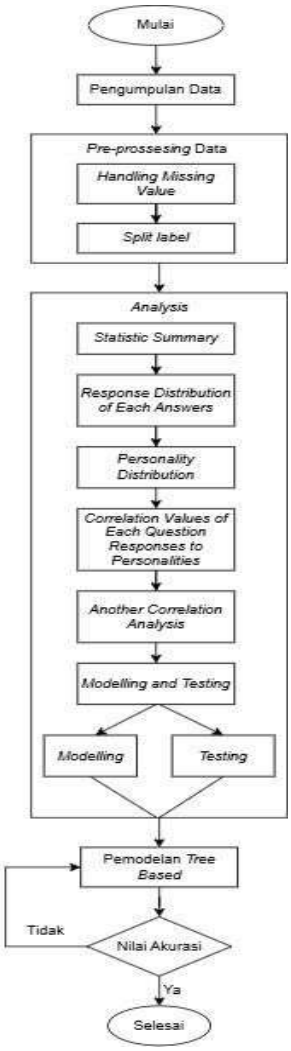
Penelitian ini bertujuan mengimplementasikan dan menganalisis kinerja Tree-Based Model dalam klasifikasi tipe kepribadian MBTI. Algoritma yang diterapkan dievaluasi menggunakan data berbeda untuk menentukan tingkat akurasi terbaik. Fokus penelitian ini pada penerapan *Tree-Based Model* dengan rumusan masalah mencakup cara penggunaan algoritma dalam klasifikasi MBTI serta relevansi algoritma yang diterapkan. Penelitian diharapkan memberikan manfaat dalam pengembangan ilmu machine learning, khususnya pada klasifikasi tipe kepribadian MBTI. Implementasi algoritma *Tree-Based* juga diharapkan memberikan rekomendasi praktis untuk aplikasi pengenalan kepribadian yang efektif, sekaligus mengidentifikasi algoritma terbaik untuk tujuan tersebut.

Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen, yang bertujuan untuk mengetahui pengaruh variabel independen terhadap variabel dependen dalam kondisi terkendali. Penelitian eksperimen mencoba, mencari, dan mengkonfirmasi hubungan kausal. Metode ini digunakan untuk mengidentifikasi efek perlakuan (*treatment*) terhadap hasil (*outcome*), sesuai dengan prinsip dasar penelitian kuantitatif.

Variabel adalah atribut atau sifat dari seseorang, objek, atau kegiatan yang memiliki variasi tertentu untuk dipelajari dan disimpulkan. Penelitian ini memiliki dua variabel, yaitu variabel bebas (*independent variable*) dan variabel terikat (dependent variable). Variabel bebas adalah tipe kepribadian MBTI yang terdiri dari empat dimensi: *Ekstroversi-Introversi* (E-I), *Sensing- iNtuition* (S-N), *Thinking-Feeling* (T-F), dan Judging-Perceiving (J-P). Variabel terikat adalah klasifikasi tipe kepribadian, yang

merupakan hasil prediksi model berdasarkan tipe kepribadian MBTI. Dalam penelitian ini, subjek penelitian terdiri dari individu yang berpartisipasi dalam pengisan kuesioner MBTI yang dipilih secara random.



Gambar 1. Diagram Penelitian

Pada penelitian ini data yang digunakan didapakan dari pengisian kuesioner, dimana pertanyaan yang di tanyakan mengacu pada salah satu website tes kepribadian MBTI 16Personalities. Sebelum mengisi kuesioner yang ada peserta diharuskan untuk menjawab tes terlebih dahulu pada website tersebut kemudian mengisi hal yanng serupa pada kuesioner yang ada dan di akhiri dengan tangkapan layar hasil dari tes kepribadian pada website 16Personalities.

Pre-processing data dalam penelitian ini bertujuan menyiapkan data untuk pemodelan. Penanganan missing value dilakukan untuk menggantikan data yang hilang akibat tidak tersedianya informasi tertentu. Penilaian MBTI mengidentifikasi individu berdasarkan empat dimensi inti: *Ekstrovert-Introvert* (E-I), *Sensing- iNtuition* (S-N), *Thinking-Feeling* (T-F), dan *Judging-Perceiving* (J-P). Statistik deskriptif, distribusi respons, distribusi kepribadian, dan nilai korelasi antara respons dan dimensi kepribadian dianalisis untuk memahami hubungan setiap pertanyaan dengan tipe kepribadian yang diukur.

Modelling dilakukan dengan menggunakan *Tree-Based Model*, seperti *Decision Tree Classifier, Random Forest Classifier, Extra Trees Classifier, Gradient Boosting Classifier*, dan lainnya. Dataset dibagi menjadi train set (80%) dan test set (20%) dengan dimensi kepribadian dianalisis secara terpisah. Pengujian dilakukan menggunakan Custom Accuracy Score Formula untuk mengevaluasi performa model pada setiap dimensi kepribadian, menghasilkan skor akurasi untuk dimensi E-I, S-N, T-F, dan J-P.

Analisis lanjut dilakukan dengan hanya menggunakan pertanyaan- pertanyaan yang memiliki korelasi tinggi terhadap setiap dimensi kepribadian. Model diulang menggunakan subset pertanyaan tersebut untuk meningkatkan akurasi prediksi. Hasil pengujian akhir menunjukkan efektivitas model dalam memprediksi tipe kepribadian MBTI dengan akurasi lebih baik dibandingkan pengujian awal menggunakan keseluruhan pertanyaan. Data hasil pengujian dan analisis disajikan dalam tabel dan visualisasi untuk setiap dimensi.

Hasil dan Pembahasan

Analisis *Pre-Processing Data*

Analisis pre-processing data bertujuan menyiapkan data untuk pemodelan. Penanganan missing value dilakukan dengan memeriksa keberadaan nilai null pada dataset menggunakan perintah df_rename. isnull ().sum ().sum (). Setelah itu, nilai null diganti menggunakan modus setiap kolom dengan perintah df_miss_handled=df_rename. Fillna (df_rename.mode ().iloc[0]), memastikan data lengkap dan siap digunakan. Setelah penggantian, jumlah nilai null diperiksa kembali untuk memastikan tidak ada data yang hilang. Split label dilakukan untuk memisahkan komponen MBTI, menghilangkan tambahan (-A/-T), seperti pada ENFP-A atau INTJ-T, dengan perintah df_splitted['MBTI'] = df_splitted['MBTI']. apply (lambda x: x.split('-')[0]). Komponen MBTI selanjutnya dipisah menjadi empat dimensi utama (E-I, S-N, T-F, J-P) menggunakan fungsi def split_mbti(label) untuk mengekstrak setiap karakter MBTI dalam bentuk pandas series. Hasil pemisahan ini ditampilkan dalam tabel, mempermudah analisis lanjutan berdasarkan dimensi MBTI yang terstruktur.

Analisis Hasil *Statistic Summary*

	q1	q2	q3	q4	q5	q6	q7	q8	q9	q10	...	q51	
count	174.000000	174.000000	174.000000	174.000000	174.000000	174.000000	174.000000	174.000000	174.000000	174.000000	...	174.000000	174.0
mean	4.402299	4.712644	4.143678	4.557471	4.362069	3.948276	4.442529	4.287356	4.396552	3.804598	...	4.091954	3.9
std	2.053953	1.964471	1.864496	1.966372	1.887270	2.091179	2.173042	2.084393	1.895596	1.937396	...	2.037976	1.9
min	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	...	1.000000	1.0
25%	3.000000	3.000000	3.000000	3.000000	3.000000	2.000000	2.000000	3.000000	3.000000	2.000000	...	2.000000	2.0
50%	5.000000	5.000000	4.000000	5.000000	4.000000	4.000000	5.000000	4.000000	4.000000	4.000000	...	4.000000	4.0
75%	6.000000	6.750000	6.000000	6.000000	6.000000	6.000000	7.000000	6.000000	6.000000	5.000000	...	6.000000	5.0
max	7.000000	7.000000	7.000000	7.000000	7.000000	7.000000	7.000000	7.000000	7.000000	7.000000	...	7.000000	7.0

8 rows × 60 columns

Gambar 2. Hasil Statistic Summary

Pertanyaan q1 memiliki 174 data dengan nilai minimum 1, maksimum 7, rata-rata 4,40, dan standar deviasi 2,05. Kuartil 1 (25%) bernilai 3, kuartil 2 (50%) bernilai 5, dan kuartil 3 (75%) bernilai 6, menunjukkan persebaran data yang sempit dengan keragaman rendah. Polanya serupa pada variabel lain, dengan statistik deskriptif berupa nilai

minimum, maksimum, rata-rata, standar deviasi, dan kuartil, yang memberikan gambaran distribusi dan persebaran data untuk analisis dimensi kepribadian.

Analisis hasil Response Distribution

Pada tahap ini yang dilakukan adalah menampilkan Histogram untuk setiap jawaban responden di setiap pertanyaan. Histogram menampilkan jawaban dari 1-7 sesuai jawaban yang di sediakan untuk setiap pertanyaan.

Analisis hasil *Correlation Value*

Pada analisis korelasi menggunakan *Pearson Correlation,* variabel *personalities* diubah menjadi boolean untuk menghasilkan *correlation matrix*. Pada dimensi E-I, pertanyaan dengan korelasi positif menunjukkan kecenderungan responden ke arah *Ekstrovert*, sedangkan korelasi negatif mengarah ke *Introvert.* Dimensi ini mencerminkan bagaimana individu berinteraksi dengan lingkungan, dengan beberapa pertanyaan seperti q43 dan q36 menunjukkan korelasi tinggi terhadap *Ekstrovert.* Pada dimensi S-N, korelasi positif menunjukkan kecenderungan ke arah *Sensing*, sedangkan korelasi negatif mengarah ke *iNtuition*. Dimensi ini merepresentasikan cara individu memproses informasi, dengan pertanyaan seperti q17 dan q57 memiliki korelasi tinggi terhadap *Sensing.*

Pada dimensi T-F, korelasi positif mengindikasikan kecenderungan ke arah *Thinking,* sementara korelasi negatif mengarah ke *Feeling*, yang mencerminkan pendekatan individu dalam pengambilan keputusan dan pengelolaan emosi. Pertanyaan seperti q25 dan q33 memiliki korelasi tinggi terhadap *Thinking.* Pada dimensi J-P, korelasi positif menunjukkan kecenderungan ke arah Judging, sedangkan korelasi negatif mengarah ke *Perceiving.* Dimensi ini mencerminkan pendekatan individu terhadap perencanaan dan pengambilan keputusan, dengan pertanyaan seperti q9 dan q56 menunjukkan korelasi tinggi terhadap *Judging*. Analisis ini membantu memahami hubungan pertanyaan terhadap dimensi kepribadian.

Analisis Hasil *Another Correlation*

Analisis hasil korelasi menunjukkan bahwa pertanyaan dengan korelasi tinggi terhadap dimensi kepribadian memberikan wawasan penting untuk meningkatkan akurasi model. Pada pengujian model untuk dimensi MBTI, model *Extra Trees* memiliki nilai akurasi tertinggi (0.70) karena pemilihan fitur dan pemisahan acak yang membantu mengurangi *overfitting* dan meningkatkan generalisasi. Model *Gradient Boosting* dan *Ada Boost* mengikuti dengan akurasi (0.69), menunjukkan kemampuan yang baik dalam menangkap interaksi dan memperbaiki kesalahan klasifikasi. Model lain seperti *Hist Gradient Boosting* (0.68), *Random Forest* (0.65), dan *Decision Tree* (0.64) menunjukkan performa yang lebih rendah, dengan *Bagging* menjadi model dengan akurasi terendah (0.63), menunjukkan efektivitas yang kurang dibanding model lainnya.

Ketika hanya menggunakan pertanyaan dengan korelasi tinggi, model *Random Forest* mencapai akurasi tertinggi (0.725), menunjukkan stabilitas yang baik dalam pengambilan keputusan dan kemampuan menangkap interaksi non- linier. Model lain seperti *Hist Gradient Boosting* (0.71) dan *Gradient Boosting* (0.707) juga menunjukkan performa yang kuat. Sebaliknya, *Decision Tree* memiliki akurasi terendah (0.61) karena rentan terhadap *overfitting* dan keterbatasannya dalam menangkap pola *non-linier*. Analisis ini menegaskan bahwa pemilihan model dan fitur yang relevan sangat penting untuk meningkatkan akurasi prediksi tipe kepribadian MBTI.

Kesimpulan

Kesimpulan dari penelitian ini menunjukkan bahwa Tree-Based Model dapat digunakan untuk mengklasifikasikan tipe kepribadian MBTI. Prosesnya meliputi pengumpulan data, pre-processing data, implementasi model, dan analisis kinerja masing-masing model dalam klasifikasi tipe kepribadian. Berdasarkan prediksi yang dilakukan menggunakan dataset dengan pembagian train set dan test set sebesar 8:2, model terbaik adalah Random Forest, yang mencapai akurasi tertinggi sebesar 0.725. Hal ini menunjukkan kemampuan model Random Forest dalam menangkap pola data dan memberikan hasil prediksi yang lebih akurat dibandingkan model lainnya.

DAFTAR PUSTAKA

Alifah, R. N., Najib, M. K., Nurdiati, S., Sari, A. P., Herlambang, K., Ginting, D. T. P. B., & Sya'adah, S. N. (2024). Perbandingan Metode Tree Based Classification untuk Masalah Klasifikasi Data Body Mass Index. *Indonesian Journal of Mathematics and Natural Sciences*, *47*(1), 49–65.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research, 16, 321–357.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 4171–4186.

Effendy, R. A. (2025). *Eksplorasi Algoritma Tree Based Model Untuk Kasus Tipe Kepribadian dengan Myers Briggs Type Indicator (MBTI)*. Universitas Islam Indonesia.

Fitri, N. A., Sutanto, T. E., & Liebenlito, M. (2023). Deteksi Kepribadian MBTI pada Diskusi Agama Islam di Twitter Indonesia 2009-2019. *The Indonesian Journal of Computer Science*, *12*(5).

Faisal Amri, C., Prasetyowati, S., & Sibaroni, Y. (2022). SOCIAL MEDIA USER PERSONALITY CLASSIFICATION BASED ON HOW USER LIVE AND MAKE DECISION. In Jurnal Ilmiah Penelitian dan Pembelajaran Informatika) (Vol. 07). https://www.youtube.com/watch?v=iz7lE1g4XM4

Hadiwandra, T. Y. (2019). Perbandingan Kinerja Model Klasifikasi Decission Tree, Bayesian Classifier, Instance Base, Linear Function Base, Rule Base pada 4 Dataset Berbeda. *Sains Dan Teknologi Informasi*, *5*(1), 70–78.

Herlambang, B. A., Soebroto, A. A., & Perdana, R. S. (2019). Journal of Information Technology and Computer Science, 4(1), 1-10.

Ikhram, F., & Ramsari, N. (2023). PENGGUNAAN INDIKATOR MYERS- BRIGGS TYPE INDICATOR (MBTI) UNTUK. In SOSIERA): Vol. II.

Pulungan, M. P., Purnomo, A., & Kurniasih, A. (2024). Penerapan SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Kepribadian MBTI Menggunakan Naive Bayes Classifier. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, *11*(5), 1033–1042.

Ryan, G., Katarina, P., & Suhartono, D. (2023). Mbti personality prediction using machine learning and smote for balancing data based on statement sentences. *Information*, *14*(4), 217.

Serrano-Guerrero, J., Alshouha, B., Bani-Doumi, M., Chiclana, F., Romero, F. P., & Olivas, J. A. (2024). Combining machine learning algorithms for personality trait prediction. *Egyptian Informatics Journal*, *25*, 100439.

Wahyudi, A. K., Azizah, N., & Saputro, H. (2022). Data Mining Klasifikasi Kepribadian Siswa SMP Negeri 5 Jepara Menggunakan Metode Decision Tree Algoritma C4. 5. *Journal of Information System and Computer*, *2*(2), 8–13.

WHO. (2023). World Health Statistics 2023 Monitoring Health for the SDGs, Sustainable Development Goals. World Health Organization.

Zhang, H. (2023). MBTI Personality Prediction Based on BERT Classification. In Highlights in Science, Engineering and Technology CSIC (Vol. 2022).

Copyright holder:
Rani Asriya Effendy, Irving Vitra Paputungan (2025)

First publication right:
Syntax Literate: Jurnal Ilmiah Indonesia

This article is licensed under:

