

Pendekatan Machine Learning untuk Klasifikasi Kepribadian: Studi Logistic Regression dan Naive Bayes

Laurensius Marcellano Adi Pradana ^{1,*}, Fikri Budiman ¹, Defri Kurniawan ¹

¹ Program Studi Teknik Informatika, Universitas Dian Nuswantoro, Indonesia

* Correspondence: 111202113597@mhs.dinus.ac.id

Copyright: © 2025 by the authors

Received: 1 September 2025 | Revised: 13 September 2025 | Accepted: 6 Oktober 2025 | Published: 3 Desember 2025

Abstrak

Klasifikasi kepribadian memiliki peran penting dalam bidang pendidikan, konseling, dan rekrutmen. Penelitian ini menggunakan pendekatan kuantitatif berupa eksperimen komputasi dengan algoritma pembelajaran mesin, yaitu *Logistic Regression* dan *Naive Bayes*. Data yang digunakan berasal dari dataset publik Kaggle yang terdiri dari 2.900 data, terbagi dalam 1.491 data untuk tipe kepribadian Extrovert dan 1.409 data untuk Introvert. Proses persiapan data meliputi konversi variabel kategori dengan *LabelEncoder*, normalisasi fitur numerik menggunakan *StandardScaler*, serta pembagian data menjadi 80% untuk pelatihan dan 20% untuk pengujian dengan metode validasi silang 5-fold. Evaluasi performa model dilakukan menggunakan empat metrik yaitu *Accuracy*, *Precision*, *Recall*, dan *F1-Score*, sehingga penilaian lebih lengkap dibandingkan penelitian sebelumnya yang hanya fokus pada akurasi. Hasil temuan kami menunjukkan bahwa *Logistic Regression* mampu mencapai akurasi sebesar 91%, sedangkan *Naive Bayes* memberikan akurasi lebih tinggi yaitu 92% serta lebih baik dalam mendeteksi kategori Introvert. Implikasi ini menunjukkan bahwa algoritma sederhana masih relevan dan efektif dalam klasifikasi kepribadian. Penerapan hasil penelitian ini dapat membantu konselor dalam mengenali tipe kepribadian siswa, mendukung pengembangan strategi pembelajaran yang sesuai dengan karakter siswa, serta meningkatkan efektivitas proses rekrutmen dengan menyesuaikan karakter individu terhadap kebutuhan organisasi.

Kata kunci: *extrovert; introvert; klasifikasi kepribadian; logistic regression; naive bayes*

Abstract

Personality classification plays an important role in education, counseling, and recruitment. This study applies a quantitative approach in the form of computational experiments using machine learning algorithms, namely *Logistic Regression* and *Naive Bayes*. The data were obtained from a public Kaggle dataset consisting of 2,900 records, with 1,491 labeled as *Extrovert* and 1,409 as *Introvert*. The preprocessing stage involved converting categorical variables with *LabelEncoder*, normalizing numerical features using *StandardScaler*, and splitting the dataset into 80% for training and 20% for testing with 5-fold cross-validation. Model performance was evaluated using four metrics: *Accuracy*, *Precision*, *Recall*, and *F1-Score*, providing a more comprehensive assessment compared to previous studies that focused only on accuracy. Our findings show that *Logistic Regression* achieved an accuracy of 91%, while *Naive Bayes* yielded a higher accuracy of 92% and performed better in detecting the *Introvert* category. These results imply that simple algorithms remain relevant and effective in personality classification. The application of this research can assist counselors in identifying students' personality types, support the development of learning strategies tailored to individual characteristics, and enhance the effectiveness of recruitment processes by aligning candidates' personalities with organizational needs.

Keywords: *extrovert; introvert; logistic regression; naive bayes; personality classification*



PENDAHULUAN

Kepribadian adalah cara seseorang menunjukkan diri kepada orang lain (Putri et al., 2025). Selain itu, kepribadian juga mencakup sifat-sifat dan cara bertindak yang khas, serta membantu seseorang beradaptasi dengan lingkungan sekitarnya (Nisa & Mirawati, 2022). Dengan memahami sifat dan kepribadian seseorang, kita bisa menemukan bakat dan kemampuan yang dimiliki, sehingga kemampuan yang terpendam dalam diri orang tersebut dapat dikembangkan (Karim, 2020; Yulianti et al., 2024). Saat seseorang berinteraksi dengan orang-orang di sekitarnya, mereka memiliki jenis kepribadian masing-masing untuk bisa beradaptasi atau menyesuaikan diri dengan lingkungannya (Masitoh et al., 2023; Nugraha & Zuhriah, 2023; Ristyawati, 2020).

Salah satu pendekatan populer untuk mengukur kepribadian adalah *Myers-Briggs Type Indicator* (MBTI), yang membagi kepribadian menjadi beberapa tipe, termasuk *Introvert* dan *Extrovert*. Seseorang yang memiliki kepribadian introvert cenderung tenang, lebih suka menyendiri, kurang ramah, dan sering merasa takut. Mereka biasanya merencanakan terlebih dahulu sebelum melakukan sesuatu, mudah tersinggung, sangat peka terhadap perasaan sendiri, dan selalu berada dalam kontrol yang tetap. Mereka lebih peka terhadap hukuman daripada hadiah. Sementara itu, orang ekstrovert lebih mudah berinteraksi dengan orang lain, memiliki banyak teman, memanfaatkan kesempatan, dan sering memperhatikan hal-hal di luar diri sendiri. Tindakan mereka cenderung tidak dipikirkan terlebih dahulu, mereka suka perubahan, lebih suka bertindak langsung, dan tidak terlalu terikat oleh kontrol yang ketat. Fenomena terbaru pada bulan Mei 2025 menunjukkan bahwa MBTI bisa membantu memperbaiki dinamika kelompok, cara berkomunikasi, pemahaman diri sendiri, sampai pengambilan keputusan dan pertumbuhan pribadi (Laksono & Astuti, 2020). Hal ini menegaskan bahwa metode pengklasifikasian seperti *Logistic Regression* dan *Naive Bayes* perlu digunakan untuk mengenali jenis kepribadian dengan cepat, adil, dan hemat waktu, terutama di bidang pendidikan, konseling, serta organisasi (Cahyani & Budiman, 2025; Meilana et al., 2021; Setiawan, 2022).

Mengatasi tantangan dalam mengenali jenis kepribadian secara cepat, adil, dan efisien, penelitian ini menggunakan dua algoritma klasifikasi, yaitu *Logistic Regression* dan *Naive Bayes*, dalam mengenali jenis kepribadian seseorang, terutama antara *Introvert* dan *Extrovert* (Khotimah & Saputri, 2022; Pangestu & Yuniarta, 2020; Prayitno, 2023). *Logistic Regression* adalah metode statistik yang digunakan untuk menentukan hubungan antara variabel bebas dengan variabel hasil yang memiliki dua kategori (Liu et al., 2023; Song et al., 2021). Algoritma ini bekerja dengan baik karena bisa memberikan kemungkinan atau peluang suatu data (Itou et al., 2021) termasuk dalam kategori tertentu (Bailly et al., 2022; Wang et al., 2021; Zhao et al., 2021), serta hasilnya mudah dipahami, sehingga sering digunakan dalam masalah klasifikasi dua kelas. Sementara itu, *Naive Bayes* adalah algoritma yang berbasis pada probabilitas dan menggunakan Teorema Bayes (Chen et al., 2021; Peretz et al., 2024) dengan mengasumsikan bahwa setiap fitur dalam data adalah independent (Li et al., 2022; Reddy et al., 2022; Shang, 2024). Penelitian terbaru menunjukkan bahwa meskipun dianggap sederhana, kedua algoritma ini tetap kompetitif dan sering menghasilkan kinerja yang setara atau bahkan lebih baik dibandingkan metode yang lebih kompleks (Aji, 2020; Mukharyahya et al., 2025; Putra & Setiawan, 2025; Ridho et al., 2023).

Kendati demikian, kajian sebelumnya menunjukkan sejumlah keterbatasan. Pertama, banyak penelitian hanya menggunakan satu algoritma tertentu tanpa melakukan perbandingan langsung. Pada Prameswari & Setiawan (2020) menggunakan *Logistic Regression* dengan pendekatan TF-IDF, sementara Pulungan et al. (2024) menerapkan *Naive Bayes* dengan teknik SMOTE untuk mengatasi ketidakseimbangan data. Kedua, sebagian besar penelitian sebelumnya masih menjadikan akurasi sebagai metrik evaluasi utama, padahal akurasi tidak cukup representatif untuk menilai performa model pada kasus data yang seimbang maupun

tidak seimbang. Evaluasi tunggal berbasis akurasi berisiko mengabaikan dimensi penting lain, seperti kemampuan model dalam mengidentifikasi kelas minoritas (*Recall*) atau ketepatan klasifikasi positif (*Precision*). Dengan demikian, diperlukan kerangka evaluasi yang lebih komprehensif, misalnya melalui penggunaan *F1-Score* sebagai metrik harmonisasi *Precision* dan *Recall*.

Penelitian ini bertujuan untuk membandingkan *Logistic Regression* dan *Naive bayes* dalam mengklasifikasikan kepribadian *Introvert* dan *Extrovert* menggunakan dataset dari Kaggle. Kontribusi dari penelitian ini terletak pada penggunaan berbagai metrik evaluasi yang lebih lengkap, yaitu *Accuracy*, *Precision*, *Recall*, dan *F1-Score*, bukan hanya akurasi saja seperti penelitian sebelumnya. *Logistic Regression* dipilih karena memiliki kemudahan dalam penjelasan dan efisiensi, sementara *Naive bayes* dikenal sederhana namun tetap efektif dalam klasifikasi teks. Perbandingan kedua metode ini penting untuk mengetahui kelebihan dan kelemahan masing-masing algoritma serta menguji sejauh mana algoritma sederhana dapat digunakan dalam klasifikasi kepribadian. Implikasi dari hasil penelitian ini tidak hanya berkontribusi pada literatur akademis di bidang klasifikasi kepribadian, tetapi juga relevan dalam aplikasi nyata seperti personalisasi strategi pembelajaran, asesmen konseling, dan proses rekrutmen berbasis kepribadian.

METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan cara melakukan eksperimen komputasi berbasis algoritma machine learning. Data yang digunakan berasal dari dataset publik di Kaggle (<https://www.kaggle.com/datasets/rakeshkapilavai/Extrovert-vs-Introvert-behavior-data>), terdiri dari 2.900 entri survei, yaitu 1.491 data berlabel *Extrovert* dan 1.409 data berlabel *Introvert*. Data tersebut berupa kombinasi numerik seperti *Time_spent_Alone*, *Social_event_attendance*, *Going_outside*, *Friends_circle_size*, dan *Post_frequency*, serta kategori seperti *Stage_fear* dan *Drained_after_socializing*, dengan variabel target berupa *Personality*. Sebelum memulai pemodelan, data diolah terlebih dahulu untuk memastikan tidak terdapat nilai kosong atau data yang sama. Variabel kategorikal diubah menjadi bentuk numerik menggunakan *LabelEncoder*, sedangkan semua fitur dinormalisasi dengan *StandardScaler* agar memiliki skala yang sama, sehingga memperbaiki kinerja model. Setelah itu, dataset dibagi menjadi 80% untuk latihan dan 20% untuk pengujian secara stratifikasi agar keseimbangan antar kelas tetap terjaga. Penggunaan *random state* dilakukan untuk memudahkan reproduksi hasil. Selain itu, evaluasi tambahan juga dilakukan melalui metode *cross-validation* 5-fold agar hasil yang diperoleh lebih akurat.

Algoritma pertama yang digunakan adalah *Logistic Regression*, yaitu metode untuk memilah dua kategori dengan menghitung kemungkinan suatu kategori menggunakan fungsi logistik. *Logistic Regression* efektif dalam membedakan dua kelompok dan mudah dipahami, dalam penelitian ini model dijalankan dengan regularisasi L2 dan batas iterasi maksimal 1000 agar model bisa mencapai kestabilan. Algoritma kedua adalah *Naive bayes*, metode klasifikasi yang didasarkan pada Teorema Bayes dan mengasumsikan bahwa semua fitur saling bebas. Jenis yang digunakan adalah *GaussianNB* karena sesuai dengan fitur numerik kontinu hasil preprocessing, dan parameter digunakan sesuai pengaturan default dari *scikit-learn*.

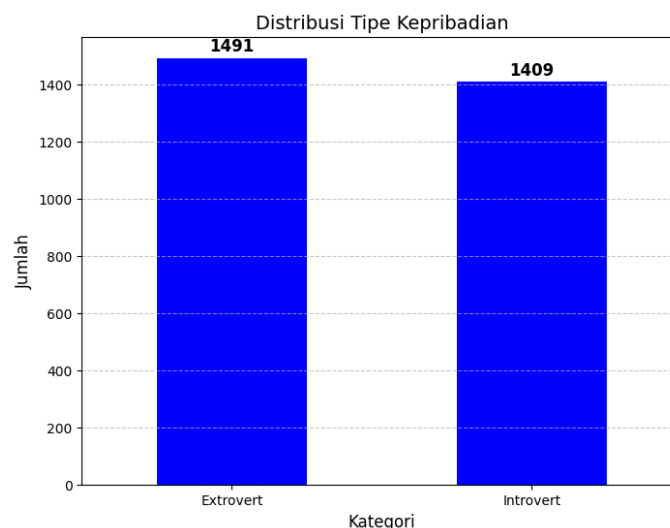
Evaluasi performa model dilakukan dengan menggunakan beberapa ukuran, yaitu akurasi, *Precision*, *Recall*, dan *F1-Score*. Akurasi digunakan untuk mengetahui seberapa baik model bekerja secara umum dengan melihat persentase dari prediksi yang benar terhadap seluruh data uji. *Precision* digunakan untuk mengukur seberapa tepat model dalam memprediksi kategori positif, sedangkan *Recall* mengukur seberapa baik model mampu mendeteksi semua data yang sebenarnya termasuk dalam kategori tersebut. *F1-Score* adalah kombinasi dari *Precision* dan *Recall*, sehingga berguna terutama ketika data tidak seimbang.

Selain itu, evaluasi juga menggunakan *Confusion matrix* untuk melihat kesalahan klasifikasi antar kelas, sehingga memudahkan dalam mengetahui bagian mana yang lemah dari model.

HASIL DAN PEMBAHASAN

Hasil

Dataset penelitian ini diambil dari Kaggle dengan judul Data Perilaku Extrovert versus Introvert, terdiri dari 2.900 data, yaitu 1.491 data Extrovert dan 1.409 data Introvert. Fitur yang ada mencakup data berupa angka seperti waktu yang dihabiskan sendirian, partisipasi dalam acara sosial, keluar rumah, ukuran lingkaran teman, serta frekuensi posting, serta data kategorikal seperti rasa takut berbicara di depan umum dan rasa kelelahan setelah berinteraksi sosial. Tujuan dari dataset ini adalah untuk mengklasifikasikan kepribadian seseorang. Preprocessing dilakukan menggunakan bahasa pemrograman Python di Google Colab. Beberapa langkah yang dilakukan adalah memeriksa adanya data yang kosong atau berulang, mengubah data kategorikal menjadi angka dengan LabelEncoder, serta menormalisasi data numerik menggunakan StandardScaler. Setelah itu, data dibagi menjadi dua bagian, yaitu 80% untuk latihan dan 20% untuk pengujian dengan metode stratifikasi agar distribusi kelas tetap seimbang. Untuk menjaga konsistensi hasil, digunakan nilai random state. Evaluasi model dilakukan dengan metode *5-fold cross-validation* agar hasil yang diperoleh lebih stabil dan dapat dipercaya.

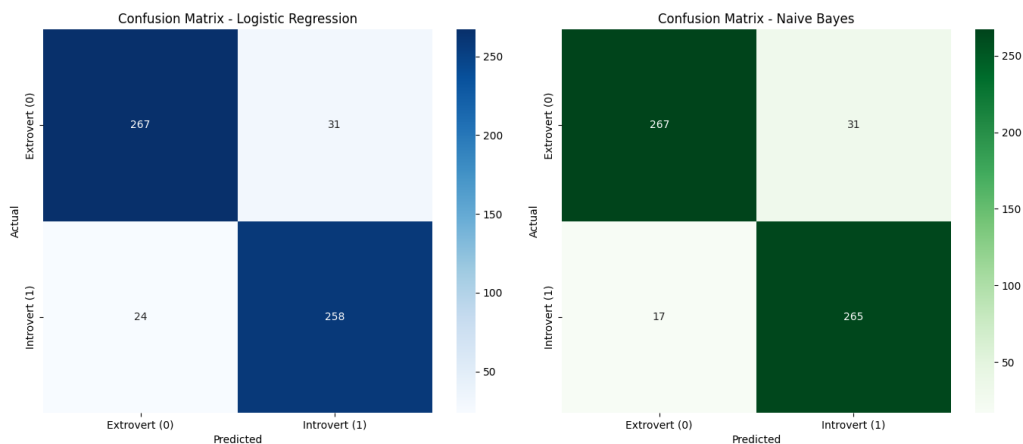


Gambar 1. Distribusi antar kelas

Pada gambar 1 terlihat bagaimana data dalam dataset penelitian ini dibagi menjadi dua kategori, yaitu *Extrovert* dan *Introvert*. Jumlah data di kedua kategori tampaknya hampir sama, yaitu 1.491 data untuk kategori *Extrovert* dan 1.409 data untuk kategori *Introvert*. Karena jumlah data kedua kategori hampir seimbang, model tidak perlu menggunakan teknik oversampling untuk memperbaiki keseimbangan data. Hal ini memungkinkan proses klasifikasi berjalan lebih baik dan optimal.

Selanjutnya gambar 2 menunjukkan *Confusion Matrix* dari model *Logistic Regression* dan *Naive Bayes*. *Confusion Matrix* ini berfungsi untuk memperlihatkan perbedaan hasil klasifikasi antara kedua model. Pada *Logistic Regression*, terdapat 31 data *Extrovert* yang salah diklasifikasikan sebagai *Introvert* (*False Negative*) dan 24 data *Introvert* yang salah diklasifikasikan sebagai *Extrovert* (*False Positive*). Di sisi lain, pada *Naive Bayes* jumlah kesalahan pada data *Extrovert* sama yaitu 31, tetapi kesalahan pada kategori *Introvert* lebih sedikit, hanya 17 data. Jadi, meskipun kinerja kedua algoritma tidak terlalu jauh berbeda, *Naive*

Bayes lebih baik dalam mendeteksi kategori *Introvert* karena menghasilkan kesalahan yang lebih sedikit dibandingkan *Logistic Regression*.



Gambar 2. *Confusion matrix*

Tabel 1. Hasil klasifikasi

Model	Accuracy	Precision	Recall	F1-Score
<i>Logistic Regression</i>	91,00%	91,00%	91,00%	91,00%
<i>Naive bayes</i>	92,00%	92,00%	92,00%	92,00%

Tabel 1 menampilkan hasil penelitian ini, di mana model *Logistic Regression* memiliki performa keseluruhan sebesar 91% berdasarkan metrik *Accuracy*, *Precision*, *Recall*, dan *F1-Score*. Model *Naive Bayes* sedikit lebih unggul dengan nilai konsisten sebesar 92% pada semua metrik yang diukur. Perbedaan antara kedua model memang tidak terlalu besar, namun temuan ini menunjukkan bahwa *Naive Bayes* lebih efektif dalam mengklasifikasikan tipe kepribadian dibandingkan *Logistic Regression*. Secara umum, kedua algoritma ini mampu mencapai performa yang baik, yaitu di atas 90%, sehingga keduanya layak digunakan sebagai metode klasifikasi dalam penelitian ini.

Pembahasan

Hasil penelitian ini menunjukkan bahwa baik *Logistic Regression* maupun *Naive Bayes* mampu mencapai kinerja yang tinggi dalam klasifikasi tipe kepribadian dengan akurasi di atas 90%. *Logistic Regression* mencapai akurasi 91%, sementara *Naive Bayes* lebih unggul sedikit dengan 92%. Perbedaan kecil ini tetap signifikan untuk dianalisis, mengingat konteks data yang digunakan bersifat relatif seimbang antara kategori *Introvert* dan *Extrovert*.

Secara teknis, perbedaan performa tersebut dapat dijelaskan melalui karakteristik algoritma. *Logistic Regression* bekerja dengan model linear yang memproyeksikan hubungan antar variabel ke dalam fungsi logistik. Kelebihannya terletak pada interpretabilitas dan kestabilan model, tetapi keterbatasannya muncul ketika pola data tidak sepenuhnya linear. Hal inilah yang terlihat pada kasus klasifikasi tipe kepribadian, di mana hubungan antar fitur seperti *social event attendance*, *friends circle size*, dan *drained after socializing* tidak sepenuhnya linear. Sebaliknya, *Naive Bayes* yang berasumsi independensi antar fitur ternyata lebih cocok untuk dataset ini. Fitur-fitur kepribadian relatif berdiri sendiri dan tidak terlalu saling bergantung, sehingga asumsi independensi pada *Naive Bayes* justru menjadi keunggulan. Temuan ini sejalan dengan Meilana et al. (2021) dan Ridho et al. (2023) yang menunjukkan bahwa *Naive Bayes* tetap kompetitif meski dibandingkan algoritma yang lebih kompleks.

Hasil penelitian kami jika dibandingkan dengan penelitian sebelumnya, kontribusi penelitian ini menjadi lebih jelas. Prameswari & Setiawan (2020) hanya menggunakan Logistic Regression dengan pendekatan TF-IDF dan AHP pada data Twitter dan menghasilkan akurasi yang relatif rendah (33,5%), terutama karena tidak adanya pembanding algoritma lain dan penggunaan metrik yang terbatas pada akurasi. Pulungan et al. (2024) memang menerapkan Naive Bayes dengan teknik SMOTE untuk mengatasi ketidakseimbangan data MBTI, namun tetap terbatas pada akurasi sebagai satu-satunya ukuran performa. Penelitian ini melengkapi kekurangan tersebut dengan tiga hal penting: pertama, membandingkan dua algoritma populer pada dataset yang relatif seimbang, kedua, menggunakan empat metrik evaluasi (*Accuracy*, *Precision*, *Recall*, *F1-Score*) yang memberikan gambaran lebih komprehensif, dan yang ketiga (3) menunjukkan performa Naive Bayes yang lebih konsisten dalam mendeteksi kategori *Introvert*. Dengan demikian, penelitian ini menegaskan bahwa algoritma sederhana tidak hanya efisien tetapi juga tetap relevan untuk permasalahan klasifikasi berbasis kepribadian.

Implikasi temuan ini penting baik secara teoretis maupun praktis. Dari sisi teori, penelitian ini memperkuat pandangan bahwa kesesuaian algoritma dengan struktur data lebih menentukan performa daripada kompleksitas model. Hal ini membuka peluang kajian lebih lanjut terkait hubungan karakteristik dataset dengan algoritma yang digunakan dalam klasifikasi psikometri. Dari sisi praktik, sistem klasifikasi berbasis machine learning dapat dimanfaatkan di bidang pendidikan untuk mempersonalisasi strategi pembelajaran sesuai tipe kepribadian siswa. Dalam konseling, hasil klasifikasi dapat mempercepat proses asesmen awal, sehingga konselor dapat merancang intervensi lebih tepat. Sementara dalam rekrutmen, pemetaan kepribadian berbasis algoritma dapat membantu organisasi menempatkan kandidat sesuai dengan kebutuhan budaya kerja dan dinamika tim.

Namun demikian, penelitian ini tidak lepas dari keterbatasan. Pertama, dataset yang digunakan masih bersifat publik dengan variabel yang terbatas sehingga mungkin belum merepresentasikan kompleksitas kepribadian dalam konteks nyata. Kedua, penelitian ini hanya membandingkan dua algoritma sederhana tanpa melibatkan model yang lebih kompleks seperti Support Vector Machine atau Random Forest, yang potensial memberikan perbandingan lebih luas. Ketiga, meskipun performa model cukup tinggi, generalisasi ke data lain belum diuji secara ekstensif di luar dataset Kaggle.

Arah penelitian berikutnya dapat diarahkan pada eksplorasi algoritma yang lebih canggih, penggunaan teknik *feature engineering* untuk memperkaya representasi kepribadian, serta pengujian pada dataset lapangan yang lebih kontekstual, misalnya dari respon mahasiswa atau karyawan dalam situasi nyata. Selain itu, integrasi metode explainable AI dapat memberikan pemahaman lebih transparan terkait faktor-faktor yang berkontribusi pada klasifikasi, sehingga hasil lebih dapat diterima oleh praktisi pendidikan, konselor, maupun pihak HRD. Secara keseluruhan, penelitian ini menegaskan bahwa meskipun algoritma sederhana seperti Logistic Regression dan Naive Bayes sering dipandang kurang unggul dibandingkan model kompleks, keduanya tetap kompetitif dan memiliki nilai praktis tinggi, terutama ketika diterapkan pada data yang sesuai dengan asumsi dasar algoritmanya.

SIMPULAN

Penelitian ini menunjukkan bahwa *Logistic Regression* dan *Naive Bayes* sama-sama efektif dalam klasifikasi kepribadian, dengan akurasi 91% dan 92%. Naive Bayes terbukti lebih konsisten, khususnya dalam mendeteksi kategori *Introvert*. Kontribusi utama penelitian ini adalah perbandingan dua algoritma dengan evaluasi komprehensif menggunakan empat metrik (*Accuracy*, *Precision*, *Recall*, *F1-Score*), melengkapi penelitian sebelumnya yang terbatas pada satu algoritma dan akurasi semata. Implikasi praktisnya, hasil penelitian dapat digunakan untuk mendukung asesmen kepribadian dalam pendidikan, konseling, dan rekrutmen. Namun,

penelitian ini masih terbatas pada dataset publik dan dua algoritma sederhana. Penelitian selanjutnya perlu menguji model yang lebih kompleks, memperluas variabel, dan menerapkan pendekatan *explainable AI* agar hasil klasifikasi lebih akurat sekaligus dapat dipahami oleh praktisi.

REFERENSI

- Aji, K. (2020). Sistem Pakar Tes Kepribadian Menggunakan Metode Naive Bayes. *JOINTECS (Journal of Information Technology and Computer Science)*, 4(2), 75–78. <https://doi.org/10.31328/jointecs.v4i2.1010>
- Bailly, A., Blanc, C., Francis, É., Guillotin, T., Jamal, F., Wakim, B., & Roy, P. (2022). Effects of dataset size and interactions on the prediction performance of logistic regression and deep learning models. *Computer Methods and Programs in Biomedicine*, 213, 106504. <https://doi.org/10.1016/j.cmpb.2021.106504>
- Cahyani, O. N., & Budiman, F. (2025). Performa Logistic Regression dan Naive Bayes dalam Klasifikasi Berita Hoax di Indonesia. *Edumatic: Jurnal Pendidikan Informatika*, 9(1), 60–68. <https://doi.org/10.29408/edumatic.v9i1.28987>
- Chen, H., Hu, S., Hua, R., & Zhao, X. (2021). Improved naive Bayes classification algorithm for traffic risk management. *EURASIP Journal on Advances in Signal Processing*, 2021(1), 30. <https://doi.org/10.1186/s13634-021-00742-6>
- Itoo, F., Meenakshi, & Singh, S. (2021). Comparison and analysis of logistic regression, Naïve Bayes and KNN machine learning algorithms for credit card fraud detection. *International Journal of Information Technology*, 13(4), 1503-1511. <https://doi.org/10.1007/s41870-020-00430-y>
- Karim, B. A. (2020). Teori Kepribadian dan Perbedaan Individu. *Education and Learning Journal*, 1(1), 40-49. <https://doi.org/10.33096/eljour.v1i1.45>
- Khotimah, H., & Saputri, T. (2022). Correlation Between Introvert-Extrovert Personality and Students' Speaking Ability. *Education and Human Development Journal*, 6(3), 61–72. <https://doi.org/10.33086/ehdj.v6i3.2416>
- Laksono, W. A., & Astuti, Y. (2020). Metode Myer Briggs Type Indicator (MBTI) Untuk Tes Kepribadian Sebagai Media Pengembangan Diri. *Journal of Information System Management (JOISM)*, 1(2), 22–27. <https://doi.org/10.24076/JOISM.2020v1i2.443>
- Li, L., Zhou, Z., Bai, N., Wang, T., Xue, K. H., Sun, H., ... & Miao, X. (2022). Naive Bayes classifier based on memristor nonlinear conductance. *Microelectronics Journal*, 129, 105574. <https://doi.org/10.1016/j.mejo.2022.105574>
- Liu, B., Wang, J., Li, Y. Y., Li, K. P., & Zhang, Q. (2023). The association between systemic immune-inflammation index and rheumatoid arthritis: evidence from NHANES 1999–2018. *Arthritis research & therapy*, 25(1), 34. <https://doi.org/10.1186/s13075-023-03018-6>
- Masitoh, I., Supriadi, P., & Marliani, R. (2023). Dampak Kepribadian Introvert dalam Interaksi Sosial. *Jurnal Pelita Nusantara*, 1(2), 245–249. <https://doi.org/10.59996/jurnalpelitanusantara.v1i2.203>
- Meilana, M., Astuti, Y., Wulandari, I. R., Sulistyowati, I., & Mimartiningtyas, B. A. (2021). Algoritma Naive Bayes untuk Mengklasifikasikan Kepribadian Siswa SMP Berdasarkan Tipologi Hippocrates-Galenus. *SISTEMASI: Jurnal Sistem Informasi*, 10(2), 480–489. <https://doi.org/10.32520/stmsi.v10i2.1339>
- Mukharyahya, Z. A., Astuti, Y. P., & Cahyani, O. N. (2025). Perbandingan Naive Bayes dan Support Vector Machine dalam Klasifikasi Tingkat Kemiskinan di Indonesia. *Edumatic: Jurnal Pendidikan Informatika*, 9(1), 119–128. <https://doi.org/10.29408/edumatic.v9i1.29512>

- Nisa, K., & Mirawati, M. (2022). Kepribadian Introvert Pada Remaja. *Educativo: Jurnal Pendidikan*, 1(2), 606–613. <https://doi.org/10.56248/educativo.v1i2.79>
- Nugraha, G., & Zuhriah, Z. (2023). Kepribadian Introvert Dalam Kemampuan Bersosialisasi Pada Mahasiswa Ilmu Komunikasi. *Jurnal Ilmu Komunikasi UHO : Jurnal Penelitian Kajian Ilmu Komunikasi Dan Informasi*, 8(2), 223–231. <https://doi.org/10.52423/jikuho.v8i2.39>
- Pangestu, N. S., & Yunianta, T. N. H. (2020). Proses Berpikir Kreatif Matematis Siswa Extrovert dan Introvert SMP Kelas VIII Berdasarkan Tahapan Wallas. *Mosharafa: Jurnal Pendidikan Matematika*, 8(2), 215–226. <https://doi.org/10.31980/mosharafa.v8i2.472>
- Peretz, O., Koren, M., & Koren, O. (2024). Naive Bayes classifier—An ensemble procedure for recall and precision enrichment. *Engineering Applications of Artificial Intelligence*, 136, 108972. <https://doi.org/10.1016/j.engappai.2024.108972>
- Prameswari, K., & Setiawan, E. B. (2020). Analisis Kepribadian Melalui Twitter Menggunakan Metode Logistic Regression dengan Pembobotan TF-IDF dan AHP. *E-Proceeding of Engineering*, 6(2), 9667–9682.
- Prayitno, S. H. (2023). Pengaruh Kepribadian Introvert-Extrovert terhadap Kepercayaan Diri dan Kecemasan pada Mahasiswa. *Jurnal Ilmiah Kesehatan Rustida*, 10(1), 8–20. <https://doi.org/10.55500/jikr.v10i1.171>
- 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. <https://doi.org/10.25126/jtiik.2024117989>
- Putra, O. P., & Setiawan, B. D. (2025). Klasifikasi Kepribadian Model Big Five (OCEAN) Pada Esai Berbahasa Inggris Menggunakan Metode Naïve Bayes Dengan Seleksi Fitur Information Gain. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 9(7), 1–6.
- Putri, N. D., Putri, Z. N. A., Alvindo, N. A., Rahmawati, A. D., & Dewi, R. S. (2025). Efektivitas Tes MBTI Terhadap Keberhasilan Proses Konseling Kelompok. *Guruku: Jurnal Pendidikan Dan Sosial Humaniora*, 3(2), 75–93. <https://doi.org/10.59061/guruku.v3i2.976>
- Reddy, E. M. K., Gurralla, A., Hasitha, V. B., & Kumar, K. V. R. (2022). Introduction to Naive Bayes and a review on its subtypes with applications. *Bayesian reasoning and gaussian processes for machine learning applications*, 1-14. <https://doi.org/10.1201/9781003164265-1>
- Ridho, M. R., Arnomo, S. A., Fifi, F., Khisal, K., & Fariska, V. (2023). Prediksi Kepribadian Mahasiswa Menggunakan Naïve Bayes. *Prosiding Seminar Nasional Ilmu Sosial Dan Teknologi (SNISTEK)*, 5, 8–14. <https://doi.org/10.33884/psnistek.v5i.8056>
- Ristyawati, A. (2020). Efektifitas Kebijakan Pembatasan Sosial Berskala Besar Dalam Masa Pandemi Corona Virus 2019 oleh Pemerintah Sesuai Amanat UUD NRI Tahun 1945. *Administrative Law and Governance Journal*, 3(2), 240–249. <https://doi.org/10.14710/alj.v3i2.240-249>
- Setiawan, A. (2022). Klasifikasi Kepribadian Seseorang Berdasarkan Postingan Twitter Dengan Algoritma Naïve Bayes Classification Studi Kasus: CV. Wilis Elektronik. *INDEXIA*, 4(2), 1. <https://doi.org/10.30587/indexia.v4i2.4301>
- Shang, Y. (2024). Prevention and detection of DDOS attack in virtual cloud computing environment using Naive Bayes algorithm of machine learning. *Measurement: Sensors*, 31, 100991. <https://doi.org/10.1016/j.measen.2023.100991>
- Song, X., Liu, X., Liu, F., & Wang, C. (2021). Comparison of machine learning and logistic regression models in predicting acute kidney injury: A systematic review and meta-

- analysis. *International journal of medical informatics*, 151, 104484. <https://doi.org/10.1016/j.ijmedinf.2021.104484>
- Wang, Z., Huang, S., Wang, J., Sulaj, D., Hao, W., & Kuang, A. (2021). Risk factors affecting crash injury severity for different groups of e-bike riders: A classification tree-based logistic regression model. *Journal of safety research*, 76, 176-183. <https://doi.org/10.1016/j.jsr.2020.12.009>
- Yulianti, Y., Ar-Roufu, T. M., Pratama, A. W., Subagja, R., Darmawan, A. A., & Wibowo, E. P. (2024). Implementasi Test Kepribadian Untuk Mengenal Diri. *Menara Ilmu*, 18(2), 78–84. <https://doi.org/10.31869/mi.v18i2.5325>
- Zhao, J., Li, Z., Gao, Q., Zhao, H., Chen, S., Huang, L., ... & Wang, T. (2021). A review of statistical methods for dietary pattern analysis. *Nutrition journal*, 20(1), 37. <https://doi.org/10.1186/s12937-021-00692-7>