"Jurnal TRANSFORMASI (Informasi & Pengembangan Iptek)" (STMIK BINA PATRIA)

E-ISSN : 2827-8550
P- ISSN : 1978-5569

KEPUTUSAN MANUSIA VS KEPUTUSAN MESIN: STUDI KOMPARATIF TERHADAP AKURASI DAN KONSISTENSI DALAM PENGAMBILAN KEPUTUSAN

Rifki Nurfalah[1)], **Helfy Susilawati**[2)], **Ica Khoerunnisa**[3)]
[1)] *"Teknik Elektro" UNIVERSITAS GARUT*
[2)] *"Teknik Elektro" UNIVERSITAS GARUT*
[3)] *"Teknik Elektro" UNIVERSITAS GARUT*

Email : rifki.nurpalah@uniga.ac.id[1)],
helfy.susilawati@uniga.ac.id[2)],
icakhoerunnisa14@gmail.com[3)]

Abstract

This study aims to analyze and compare the accuracy, consistency, and decision-making efficiency between humans and machine learning (ML) algorithms in tabular data classification tasks. The dataset comprises 50 classification cases containing both numerical and categorical features with binary decision labels. Two groups were compared: 10 human participants, and six ML algorithms—Logistic Regression, Support Vector Machine, Random Forest, Decision Tree, k-Nearest Neighbors, and Naive Bayes. ML models were trained on 80% of the data and tested on the remaining 20%, while human participants manually classified all 50 test cases. The results showed that the average human accuracy was 76.2%, while ML algorithms achieved between 78.9% and 91.8%, with Random Forest yielding the highest performance. Human decision-making took an average of 18 seconds per case, significantly slower than the algorithmic predictions completed within milliseconds. Additionally, high variability in human responses indicated lower consistency compared to deterministic outputs from ML models. These findings support the integration of ML algorithms as a decision support or replacement tool in data-driven domains, with the potential to reduce human error in high-stakes environments. Nevertheless, human involvement remains essential in contexts requiring ethical consideration and interpretability.

Keywords *: Human Decision, Machine Learning, Human Error, Classification, Automation*

Abstrak

Penelitian ini bertujuan untuk menganalisis dan membandingkan akurasi, konsistensi, serta efisiensi waktu dalam pengambilan keputusan antara manusia dan algoritma machine learning (ML) dalam konteks klasifikasi data tabular. Dataset yang digunakan terdiri dari 50 kasus klasifikasi yang mengandung atribut numerik dan kategorikal, dengan label keputusan biner. Dua kelompok dibandingkan dalam penelitian ini: 10 partisipan manusia, dan enam algoritma ML yaitu Logistic Regression, Support Vector Machine, Random Forest, Decision Tree, k-Nearest Neighbors, dan Naive Bayes. Model ML dilatih menggunakan 80% data dan diuji pada 20% sisanya, sedangkan partisipan manusia mengevaluasi 50 data uji secara manual. Hasil menunjukkan bahwa rata-rata akurasi manusia adalah 76.2%, sedangkan algoritma ML mencapai akurasi antara 78.9% hingga 91.8%, dengan Random Forest sebagai model paling unggul. Waktu pengambilan keputusan rata-rata manusia adalah 18 detik per kasus, jauh lebih lambat dari algoritma yang menyelesaikan prediksi dalam hitungan milidetik. Selain itu, variabilitas antar keputusan manusia menunjukkan rendahnya konsistensi dibandingkan prediksi deterministik oleh mesin. Hasil penelitian ini mendukung penggunaan algoritma ML sebagai alat bantu atau alternatif dalam pengambilan keputusan berbasis data, dengan potensi mengurangi human error, terutama dalam lingkungan berisiko tinggi. Namun demikian, keterlibatan manusia tetap penting dalam konteks etika dan interpretabilitas.

Kata kunci : Keputusan Manusia, Machine Learning, Human Error, Klasifikasi, Otomatisasi

E-ISSN : 2827-8550
P- ISSN : 1978-5569

1. Pendahuluan

Dalam proses pengambilan keputusan, manusia sangat bergantung pada pengalaman dan intuisi, namun juga rentan terhadap bias, kelelahan, dan keterbatasan kognitif (Kahneman, 2011). Dengan meningkatnya kompleksitas data, kebutuhan akan pendekatan yang sistematis dan presisi semakin mendesak. Machine Learning (ML) sebagai bagian dari kecerdasan buatan telah banyak digunakan untuk meningkatkan efisiensi dan akurasi pengambilan keputusan (Esteva et al., 2017).

Permasalahan utama dalam penelitian ini adalah apakah algoritma ML dapat menggantikan atau setidaknya mendukung keputusan manusia secara signifikan dalam konteks klasifikasi data. Penelitian ini bertujuan untuk menganalisis secara kuantitatif perbandingan antara keputusan manusia dan keluaran algoritma ML terhadap akurasi, konsistensi, dan efisiensi.

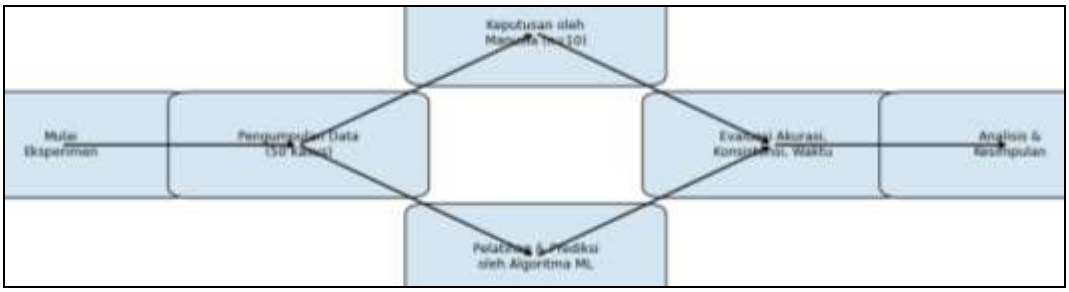
Penelitian ini bersifat eksperimental dan menggunakan pendekatan kuantitatif dengan data simulasi. Fokus utama adalah mengukur sejauh mana human error dapat diminimalkan dengan pendekatan algoritmik. Selain itu, penelitian ini juga mengkaji implikasi dari otomatisasi keputusan dalam sektor kritikal seperti keuangan dan kesehatan.

2. Metode Penelitian

2.1. Rancangan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif eksperimental untuk membandingkan hasil keputusan antara manusia dan algoritma machine learning (ML). Rancangan penelitian terdiri atas tiga tahapan utama, yaitu: pengumpulan data, pelaksanaan eksperimen, dan evaluasi hasil keputusan. Penelitian dilakukan secara simulasi berbasis data klasifikasi tabular yang telah dimodifikasi untuk menjaga kerahasiaan dan relevansi konteks.

Eksperimen dilakukan terhadap 50 data kasus klasifikasi yang melibatkan atribut numerik dan kategorikal. Dataset diolah dari data publik yang dimodifikasi untuk kepentingan eksperimen (misalnya data risiko kredit atau diagnosis penyakit).



Gambar 1. *Flowchart Metode Penelitian*

2.2. Proses Eksperimen

Sebanyak 10 partisipan manusia dilibatkan dalam eksperimen, dengan kriteria:

- Berusia 21–35 tahun
- Latar belakang pendidikan non-teknis
- Tidak memiliki pengetahuan sebelumnya tentang algoritma ML

Sementara itu, algoritma machine learning yang digunakan dalam eksperimen terdiri dari:

- Logistic Regression
- Support Vector Machine (SVM)
- Random Forest
- Decision Tree
- k-Nearest Neighbors (k-NN)
- Naive Bayes

E-ISSN : 2827-8550
P- ISSN : 1978-5569

Model-model tersebut dipilih karena mewakili berbagai pendekatan (linier, ensemble, probabilistik, instance-based) yang umum digunakan dalam studi klasifikasi.

2.3. Prosedur Penelitian

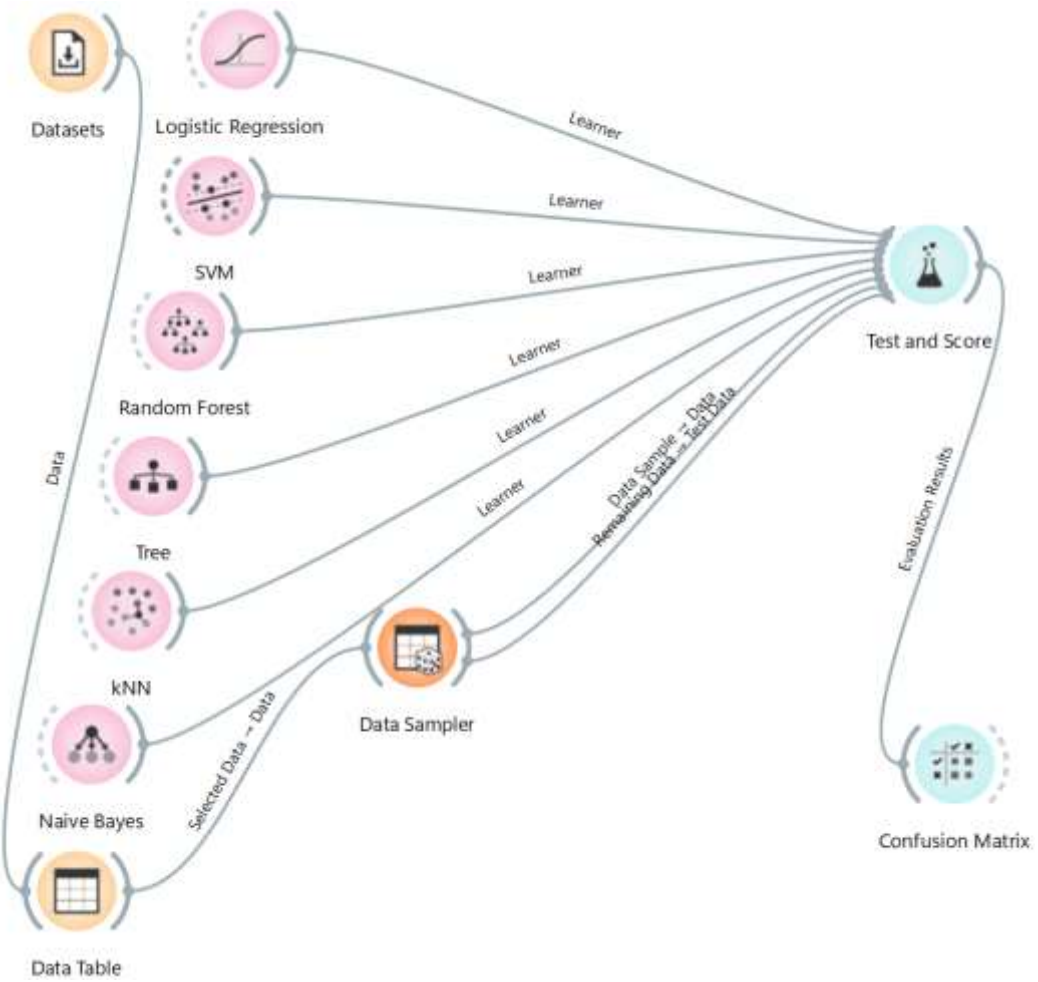
Tahapan eksperimen dirancang sebagai berikut:

1. Pengumpulan Data
   Dataset terdiri dari 50 kasus klasifikasi dengan atribut seperti usia, skor nilai, riwayat, dan label keputusan (ya/tidak). Dataset tersebut telah diolah dan dibagi menjadi data pelatihan (80%) dan data pengujian (20%).
2. Pembuatan Keputusan oleh Manusia
   Setiap partisipan diberikan 50 kasus data uji tanpa label dan diminta memberikan keputusan klasifikasi berdasarkan pemahamannya sendiri. Mereka diberi waktu maksimal 30 menit untuk menyelesaikan seluruh kasus.
3. Pelatihan dan Prediksi oleh Algoritma ML
   Model ML dilatih menggunakan data latih yang telah disiapkan. Setelah pelatihan selesai, model menghasilkan prediksi atas 50 data uji yang sama seperti yang diberikan kepada partisipan manusia.
4. Evaluasi dan Analisis
   Semua keputusan (manusia dan mesin) dibandingkan dengan label ground truth untuk memperoleh nilai akurasi. Variabilitas keputusan manusia diukur melalui simpangan baku antar partisipan. Waktu pengambilan keputusan juga dicatat dan dibandingkan.

2.4. Instrumen Penelitian

Eksperimen ini menggunakan perangkat lunak Python 3.11 dengan pustaka pendukung seperti scikit-learn untuk membangun dan mengevaluasi model ML, serta Pandas dan Seaborn untuk pengolahan dan visualisasi data. Waktu pengambilan keputusan manusia dicatat menggunakan stopwatch manual, sedangkan waktu prediksi algoritma diukur melalui fungsi time dalam Python.

E-ISSN : 2827-8550
P- ISSN : 1978-5569



Gambar 2. *Diagram Alur Eksperimen: Keputusan Manusia vs Mesin*

2.5. Evaluasi

Objek dalam penelitian ini adalah *output* keputusan klasifikasi yang dihasilkan oleh dua pihak: kelompok manusia dan model machine learning. Fokus penelitian berada pada tiga metrik evaluasi utama, yaitu:

- Akurasi terhadap ground truth (data kebenaran)
- Konsistensi antar individu/model
- Rata-rata waktu pengambilan keputusan

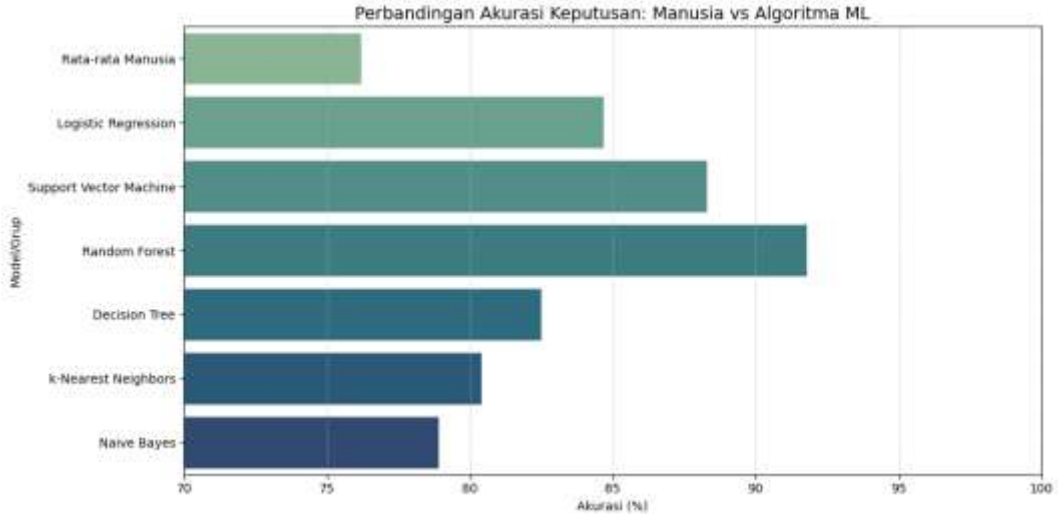
Lingkup penelitian dibatasi pada data berjenis tabular yang memiliki atribut numerik dan kategorikal. Jenis data tersebut dipilih karena representatif terhadap banyak kasus nyata seperti evaluasi risiko kredit, diagnosis kesehatan awal, dan klasifikasi kelayakan keputusan.

E-ISSN : 2827-8550
P- ISSN : 1978-5569

3. Hasil dan Pembahasan

3.1. *Hasil Penelitian*

Eksperimen menghasilkan data akurasi sebagai berikut:



Gambar 3. *Perbandingan Akurasi Keputusan: Manusia vs Algoritma ML*

Rata-rata akurasi dari keputusan manusia adalah 76.2%, dengan standar deviasi antar peserta sebesar ±5.1%. Sementara itu, model Random Forest menunjukkan akurasi tertinggi sebesar 91.8%, diikuti oleh SVM (88.3%) dan Logistic Regression (84.7%). Algoritma lain seperti Decision Tree dan k-Nearest Neighbors masing-masing mencatat akurasi sebesar 82.5% dan 80.4%, sedangkan Naive Bayes berada di angka 78.9%.

Hasil ini menunjukkan bahwa seluruh algoritma machine learning dalam eksperimen ini mampu melampaui rata-rata performa manusia dalam tugas klasifikasi.

3.2 Pembahasan

Konsistensi menjadi aspek penting dalam pengambilan keputusan. Keputusan manusia menunjukkan variabilitas antar individu, terutama pada kasus borderline. Kecepatan pengambilan keputusan juga menunjukkan kontras signifikan. Manusia membutuhkan rata-rata 18 detik untuk setiap kasus, sedangkan algoritma ML mampu menyelesaikan seluruh batch kasus dalam hitungan milidetik.

Temuan ini sejalan dengan penelitian sebelumnya oleh Esteva et al. (2017), yang menunjukkan keunggulan model deep learning dalam diagnosis kulit dibandingkan dokter manusia. Dalam konteks lain, ML juga telah digunakan secara luas dalam evaluasi kredit, deteksi penipuan, dan sistem rekomendasi dengan performa yang melampaui standar manual (Breiman, 2001; Ribeiro et al., 2016).

Namun demikian, penting untuk dicatat bahwa algoritma seperti SVM dan Random Forest masih memerlukan pelatihan dan validasi yang baik untuk menghindari overfitting. Selain itu, interpretabilitas model tetap menjadi tantangan dalam aplikasi kritis, yang menuntut transparansi dalam pengambilan keputusan.

4. Kesimpulan

Penelitian ini membuktikan algoritma machine learning mampu menghasilkan keputusan klasifikasi dengan akurasi dan konsistensi yang lebih tinggi dibandingkan manusia, khususnya dalam kondisi berbasis data terstruktur. Selisih akurasi yang signifikan (hingga 15%) menunjukkan potensi besar teknologi ini dalam mengurangi human error, terutama dalam domain dengan kebutuhan presisi tinggi.

E-ISSN : 2827-8550
P- ISSN : 1978-5569

Namun, penggunaan algoritma tetap harus mempertimbangkan konteks penggunaannya, interpretabilitas model, serta etika dalam pengambilan keputusan. Untuk implementasi lebih lanjut, kombinasi antara sistem otomatis dan pengawasan manusia tetap menjadi pendekatan ideal di banyak sektor.

Daftar Pustaka

Jurnal:

Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, 20, 273–297.

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint, arXiv:1702.08608.

Esteva, A., Kuprel, B., Novoa, R. A., et al. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115–118.

Holzinger, A., Biemann, C., Pattichis, C. S., & Kell, D. B. (2017). What do we need to build explainable AI systems for the medical domain? Review Journal of Biomedical Informatics, 71, 28–41.

Lipton, Z. C. (2018). The mythos of model interpretability. Communications of the ACM, 61(10), 36–43.

Shrestha, Y. R., Ben-Menahem, S. M., & Krogh, G. V. (2019). Organizational decision-making structures in the age of artificial intelligence. California Management Review, 61(4), 66–83.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185(4157), 1124–1131.

Varshney, K. R., & Alemzadeh, H. (2017). On the safety of machine learning: Cyber-physical systems, decision sciences, and data products. Big Data, 5(3), 246–255.

Buku:

Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus and Giroux.

Marcus, G. (2020). Rebooting AI: Building Artificial Intelligence We Can Trust. Vintage Books.

Reason, J. (1990). *Human Error*. Cambridge University Press.

Russell, S., & Norvig, P. (2016). Artificial Intelligence: A Modern Approach (3rd ed.). Pearson Education.

Prosiding seminar:

Amershi, S., Weld, D., Vorvoreanu, M., et al. (2019). Guidelines for human-AI interaction. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 1–13.

Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. Proceedings of the 2018 Conference on Fairness, Accountability and Transparency, 149–159.

Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 1–13.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD*.

Skripsi/tesis/disertasi:

Suhendra, R. (2020). Perbandingan Akurasi Model Machine Learning dalam Deteksi Penyakit Menggunakan Dataset Terbuka (Skripsi). Universitas Gadjah Mada.