

## Perbandingan Kinerja Algoritma Pembelajaran Mesin untuk Deteksi Dini Penyakit Alzheimer

Sidik\*

Program Studi Informatika, Fakultas Teknologi Informasi, Universitas Nusa Mandiri  
Jl.Margonda, No.545, Beji, Depok, Jawa Barat, Indonesia  
e-mail: [sdk.sidik1207@gmail.com](mailto:sdk.sidik1207@gmail.com)

(\* ) Corresponding Author

Artikel Info : Diterima : 12-08-2024 | Direvisi : 16-10-2024 | Disetujui : 23-10-2024

**Abstrak** Penyakit Alzheimer merupakan salah satu bentuk paling umum dari demensia yang ditandai dengan penurunan kemampuan berpikir, mengingat, dan berperilaku secara bertahap. Penyakit ini berkembang secara perlahan dan biasanya dimulai dengan hilangnya memori jangka pendek, diikuti oleh kesulitan dalam bahasa, disorientasi, dan perubahan kepribadian. Deteksi dini Alzheimer penting untuk memperlambat perkembangannya. Penelitian ini bertujuan untuk mengimplementasikan dan membandingkan kinerja algoritma klasifikasi (*Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, dan K-Nearest Neighbor*) dalam mendeteksi dini Alzheimer. Dataset dari Kaggle berisi 2.149 data dengan 35 fitur digunakan dalam penelitian ini. Proses KDD diterapkan dengan tahapan pra-pemrosesan data, penyeimbangan data menggunakan SMOTE, dan klasifikasi perbandingan algoritma. Eksperimen dilakukan dengan membagi data menjadi data latih dan data uji sebesar 70% dan 30%. Hasil penelitian menunjukkan bahwa algoritma Random Forest memiliki kinerja terbaik dengan akurasi sekitar 0,98, presisi 0,96, recall sebesar 1,00, F1-score 0,97, dan AUC sekitar 0,99. Algoritma ini terbukti lebih unggul dibandingkan algoritma lain dalam mendeteksi dini penyakit Alzheimer. Penelitian ini memberikan kontribusi dalam bentuk pendekatan KDD yang terstruktur, menunjukkan efektivitas SMOTE, dan memberikan dasar bagi penelitian lanjutan dalam bidang deteksi dini Alzheimer serta panduan praktis bagi pengembang sistem kesehatan.

Kata Kunci : penyakit *alzheimer*, KDD, data mining, SMOTE, model algoritma

**Abstracts** – *Alzheimer's disease is one of the most common forms of dementia characterized by a gradual decline in the ability to think, remember, and behave. The disease progresses slowly and usually begins with short-term memory loss, followed by language difficulties, disorientation, and personality changes. Early detection of Alzheimer's is important to slow its progression. This study aims to implement and compare the performance of classification algorithms (Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, and K-Nearest Neighbor) in early detection of Alzheimer's. The Kaggle dataset contains 2,149 data with 35 features used in this study. The KDD process is applied with the stages of data pre-processing, data balancing using SMOTE, and algorithm comparison classification. The experiment was carried out by dividing the data into training data and test data of 70% and 30%. The results showed that the Random Forest algorithm had the best performance with an accuracy of around 0.98, a precision of 0.96, a recall of 1.00, an F1-score of 0.97, and an AUC of around 0.99. This algorithm has been proven to be superior to other algorithms in early detection of Alzheimer's disease. This study contributes in the form of a structured KDD approach, demonstrating the effectiveness of SMOTE, and providing a basis for further research in the field of early detection of Alzheimer's disease as well as practical guidance for health system developers.*

**Keywords** : *alzheimer's disease, KDD, data mining, SMOTE, algorithm model*

### PENDAHULUAN

Penyakit Alzheimer merupakan salah satu bentuk paling umum dari demensia (Susanti et al., 2024) yang ditandai dengan penurunan kemampuan berpikir, mengingat, dan berperilaku secara bertahap (Kesehatan et al., n.d.). Penyakit ini berkembang secara perlahan dan biasanya dimulai dengan hilangnya memori jangka pendek, diikuti oleh kesulitan dalam bahasa, disorientasi, dan perubahan kepribadian (Kasprata & Harahap, 2023). Seiring waktu, individu yang terkena *Alzheimer* akan mengalami kesulitan dalam melakukan aktivitas sehari-hari, dan pada tahap lanjut, mereka memerlukan bantuan penuh dari orang lain.



Penyebab utama *Alzheimer* belum sepenuhnya dipahami, namun penyakit ini sering dikaitkan dengan akumulasi protein abnormal di otak, seperti plak beta-amiloid dan kusut neurofibrillary yang terdiri dari protein tau (Leuzy et al., 2022). Faktor risiko termasuk usia tua, riwayat keluarga, dan faktor genetik tertentu. Meskipun *Alzheimer* umumnya dialami oleh orang tua, ada juga kasus yang terjadi pada individu yang lebih muda, yang dikenal sebagai *Alzheimer* onset dini. Penyakit *Alzheimer* sementara waktu belum ditemukan obat yang paling tepat untuk menyembuhkannya (Wildah et al., 2020). Pengobatan yang tersedia hanya berfokus pada meredakan gejala dan memperlambat perkembangan penyakit *Alzheimer* (Rustiawati & Pondang, 2024). Oleh karena itu, deteksi dini menjadi sangat penting untuk mengelola penyakit ini lebih efektif dan memberikan perawatan yang lebih baik bagi penderita (Arfina, 2021). Upaya untuk menemukan metode deteksi dini, termasuk melalui teknik data mining dan algoritma pembelajaran mesin, semakin berkembang untuk meningkatkan akurasi dan keandalan diagnosis di tahap awal.

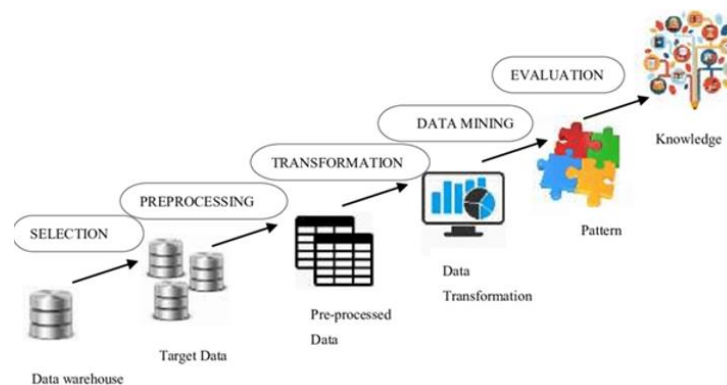
Permasalahan utama dalam penelitian ini adalah rendahnya akurasi deteksi dini *Alzheimer* yang mengandalkan metode konvensional. Metode-metode tradisional, seperti penilaian klinis dan tes kognitif, sering kali tidak cukup sensitif untuk mendeteksi perubahan kognitif awal yang mengarah ke *Alzheimer*. Akibatnya, banyak pasien yang baru terdiagnosis ketika penyakit sudah berada pada tahap lanjut, sehingga intervensi medis yang diberikan menjadi kurang efektif. Di sinilah teknologi *machine learning* dapat memainkan peran penting dengan menyediakan alat yang lebih canggih dan akurat untuk mendeteksi tanda-tanda awal *Alzheimer* berdasarkan data medis yang kompleks.

Dalam beberapa tahun terakhir, algoritma *Random Forest* telah muncul sebagai salah satu metode yang unggul dalam analisis data medis untuk prediksi dan deteksi penyakit, termasuk *Alzheimer*. Berdasarkan penelitian terbaru, algoritma ini menunjukkan kemampuan yang luar biasa dalam menangani data dengan dimensi yang tinggi dan menangani masalah overfitting yang umum terjadi pada model *machine learning* lainnya. Sebagai contoh, penelitian yang dilakukan oleh (Spurrier et al., 2022) menunjukkan bahwa *Random Forest* dapat mencapai akurasi yang sangat tinggi dalam mendeteksi *Alzheimer* dengan menggunakan data pencitraan otak dan biomarker lainnya. Studi lain oleh (Leuzy et al., 2022) juga menemukan bahwa algoritma ini mampu memprediksi perkembangan penyakit dari tahap awal dengan lebih baik dibandingkan algoritma lain seperti *Support Vector Machine* dan *Decision Tree*. Penelitian oleh (O'Donnell, 2023) menambahkan bahwa *Random Forest* efektif dalam menggabungkan berbagai jenis data, seperti data klinis dan genetik, untuk meningkatkan akurasi prediksi. Penelitian sebelumnya seperti oleh Spurrier et al. (2022) dan Leuzy et al. (2022) menunjukkan keberhasilan algoritma *Random Forest* dalam mendeteksi *Alzheimer* menggunakan data pencitraan otak dan biomarker. Namun, penelitian ini lebih fokus pada data yang kompleks dan bervariasi, termasuk data klinis dan genetik, yang belum sepenuhnya dieksplorasi dalam penelitian sebelumnya. Meskipun beberapa penelitian telah menggunakan algoritma *Random Forest*, penelitian ini menambahkan teknik *Synthetic Minority Over-sampling Technique* (SMOTE) untuk penyeimbangan data, yang tidak secara eksplisit dijelaskan dalam penelitian sebelumnya.

Penelitian ini bertujuan untuk mengimplementasikan dan membandingkan kinerja algoritma klasifikasi (*Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, dan K-Nearest Neighbor*) dalam mendeteksi dini *Alzheimer* (Fauzi et al., 2022) dalam deteksi dini penyakit *Alzheimer*. Dengan memanfaatkan kemampuan algoritma ini dalam menangani data yang kompleks dan bervariasi, diharapkan penelitian ini dapat memberikan kontribusi signifikan dalam pengembangan alat diagnosis yang lebih akurat dan dapat diandalkan. Penelitian ini memberikan kontribusi dalam bentuk pendekatan KDD yang terstruktur, menunjukkan efektivitas SMOTE, dan memberikan dasar bagi penelitian lanjutan dalam bidang deteksi dini *Alzheimer* serta panduan praktis bagi pengembang sistem kesehatan.

## **METODE PENELITIAN**

Dalam menyusun tahapan penelitian, penulis menggunakan metode penelitian *Knowledge Discovery in Database* (KDD) (Pramudya, 2022). Metode ini merupakan satu dari beberapa metode yang dapat digunakan dalam melakukan proses penambangan data (*data mining*) (Buani, 2024). *Knowledge Discovery in Database* (KDD) merupakan suatu proses analisis yang sistematis dengan tujuan memperoleh informasi baru dan akurat, serta menemukan pola dari data yang kompleks dan bermanfaat (Fakhri et al., 2021). Tujuan dari proses KDD digunakan untuk mengenali potensi data yang diperoleh dari database, yang kemudian akan diperiksa melalui pola tertentu, dianalisis, dan divisualisasikan untuk memudahkan pemahaman bagi pengguna.



Sumber: (Fahlevi, 2021)

Gambar 1. Tahapan pada KDD

Penjelasan tahapan penelitian menggunakan KDD (Fahlevi, 2021):

- a. **Pembersihan Data:** Proses menyaring data mentah untuk menghilangkan data yang tidak akurat, tidak lengkap, atau tidak relevan, sehingga hanya data yang berkualitas tinggi yang digunakan.
- b. **Integrasi Data:** Proses menggabungkan data dari berbagai sumber menjadi satu kesatuan yang koheren, menghilangkan duplikasi data dan memastikan konsistensi.
- c. **Seleksi Data:** Proses memilih data yang paling relevan dengan tujuan analisis, dengan fokus pada data yang akan memberikan jawaban atas pertanyaan penelitian.
- d. **Transformasi Data:** Proses mengubah data ke dalam format yang sesuai untuk analisis, seperti mengubah tipe data atau menormalkan data.
- e. **Penambangan Data:** Proses mengaplikasikan algoritma dan teknik statistik untuk menemukan pola, tren, atau hubungan yang tersembunyi dalam data yang telah disiapkan.
- f. **Evaluasi Pola:** Proses menilai pola-pola yang ditemukan untuk memastikan bahwa pola tersebut signifikan dan relevan dengan tujuan analisis.
- g. **Presentasi Pengetahuan:** Proses mengkomunikasikan hasil penambangan data dalam bentuk yang mudah dipahami, seperti visualisasi atau laporan, sehingga informasi yang diperoleh dapat digunakan untuk mengambil keputusan.

## HASIL DAN PEMBAHASAN

### 1. Data Original Alzheimer

Data yang digunakan pada penelitian ini diambil dari dataset publik yang ada pada *kaggle*. Jumlah data sebanyak 2149 dengan 35 fitur (Shakir, 2020). Dilakukan proses klasifikasi pada dataset dengan variabel tujuan (*dependen*) Diagnosis dengan target Yes (menderita) dan No (tidak menderita). Persebaran datanya dapat dilihat pada tabel 1.

Tabel 1. Data Alzheimer

No	Diagnosis	Bobot Nilai
1	Yes	52
2	No	146

Sumber: Hasil penelitian (2024)

Tabel 1 menampilkan distribusi data asli dari dataset *Alzheimer* dengan variabel Diagnosis yang menunjukkan jumlah responden yang terdeteksi akan menderita *Alzheimer* (Yes) sebanyak 52 dan yang tidak menderita *Alzheimer* (No) sebanyak 146.

### 2. Pra-processing

Pada tahap ini dilakukan penghapusan terhadap variabel-variabel yang tidak ada kaitan dan hubungan dengan prediksi deteksi awal penyakit *alzheimer*. Tahapan diawali dengan menampilkan semua variabel yang digunakan, kemudian dilanjutkan dengan menampilkan apakah terdapat data yang kosong (*null*). Kemudian dilakukan pengecekan data-data yang berada di *outlier* dan lakukan penghapusan terhadap data *outlier* tersebut. Tahap terakhir adalah lakukan penghapusan (*drop*) terhadap variabel *DoctorInCharge* karena tidak terkait secara

langsung dengan pendeteksian awal penyakit *alzheimer*. Proses *encoding* atau merubah tipe data pada suatu variabel tidak di lakukan karena variabel *DoctorInCharge* nantinya akan di hapus.

Tabel 2. Penghapusan variabel *DoctorInCharge*

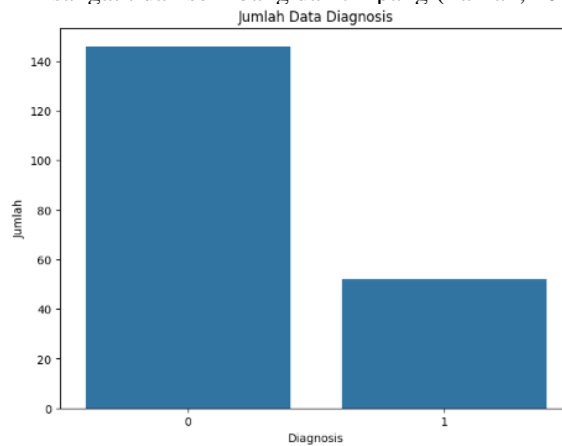
No	Variabel	Keterangan
1	<i>DoctorInCharge</i>	<i>Dropped</i>

Sumber: Hasil penelitian (2024)

Tabel 2 menjelaskan variabel yang dihapus karena tidak terkait langsung dengan deteksi awal penyakit *Alzheimer*. Variabel yang dihapus adalah *DoctorInCharge*. Penghapusan variabel *DoctorInCharge* dilakukan untuk memastikan bahwa model prediksi hanya menggunakan fitur-fitur yang relevan dan bermakna dalam konteks deteksi dini penyakit *Alzheimer*. Hal ini membantu dalam meningkatkan akurasi prediksi, efisiensi komputasi, dan interpretabilitas model, serta menghindari potensi bias dan kebisingan dalam data

3. *Handling imbalance data*

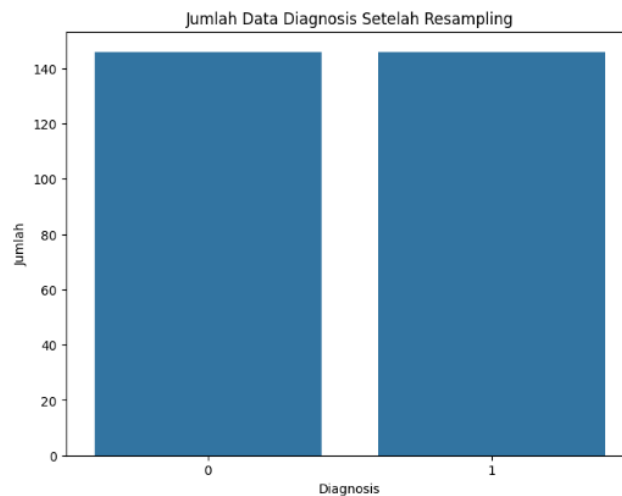
Setelah data sudah bersih, tahapan berikutnya melakukan proses penyeimbang data. *Imbalanced Dataset* merupakan keadaan *dataset* dalam *classification task* dengan kondisi pembagian dari label (biasa juga disebut kelas atau target) yang dimiliki sangat tidak seimbang dan timpang (Farhan, 2021).



Sumber: Hasil penelitian (2024)

Gambar 2. Hasil variabel *Diagnosis* dalam keadaan tidak seimbang

Pada gambar 2 terlihat data tidak seimbang untuk variabel *Diagnosis* nilai 0 sebesar 146 sedangkan nilai 1 sebesar 52. Ketidakseimbangan data dalam *machine learning* dapat mengakibatkan bias terhadap kelas mayoritas, penurunan kinerja pada kelas minoritas, metrik evaluasi yang menyesatkan, dan kesulitan dalam generalisasi.



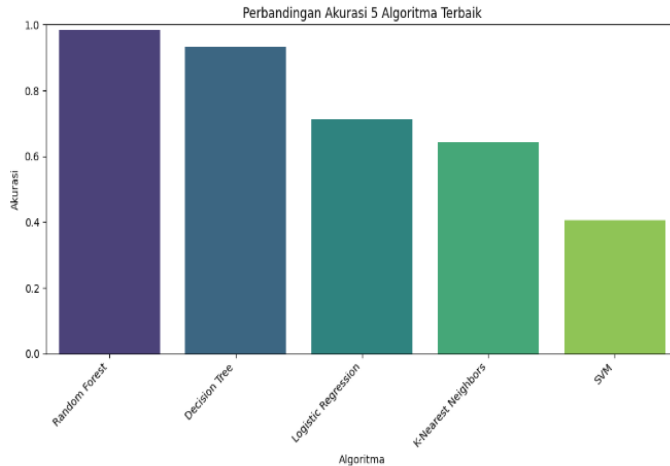
Sumber: Hasil penelitian (2024)

Gambar 3. Hasil variabel *Diagnosis* setelah dilakukan SMOTE

Pada gambar 3 menunjukkan keseimbangan data setelah dilakukan proses SMOTE, dengan nilai 0 dan 1 masing-masing sebesar 146. Proses penyeimbang data digunakan metode SMOTE (*Synthetic Minority Oversampling Technique*) yaitu sebuah metode yang menerapkan oversampling ketika jumlah data tidak mencukupi yaitu dengan cara menyeimbangkan dataset dengan meningkatkan ukuran sampel langka (Rumagit, 2019). Apabila dibandingkan dengan metode undersampling, metode SMOTE dapat menampilkan hasil lebih baik, karena tidak adanya pengurangan data (Kusuma & Sasongko, 2023).

4. Data Mining

Pada tahapan ini dilakukan implementasi atau penerapan dari 5 (lima) jenis algoritma kalsifikasi yang terdiri dari *Logistic Regression* (LR), *Decision Tree* (DT), *Random Forest* (RF), *Support Vector Machine* (SVM), dan *K-Nearest Neighbor* (KNN). Hasil dari pengujian klasifikasi kelima algoritma terhadap deteksi awal penyakit *alzheimer* terlihat pada gambar 4.



Sumber: Hasil penelitian (2024)

Gambar 4. Hasil grafik akurasi dari pengujian model algoritma untuk menguji deteksi awal penyakit *alzheimer*

Pada gambar 4 merupakan sebuah grafik yang menampilkan hasil akurasi dari pengujian kelima algoritma. Hasilnya menunjukkan bahwa *Random Forest* memiliki akurasi tertinggi (0,98), diikuti oleh *Decision Tree* (0,93), *Logistic Regression* (0,71), *K-Nearest Neighbor* (0,64), dan *Support Vector Machine* (0,40)..

5. Evaluation

Tahap evaluasi merupakan tahap berikutnya setelah tahapan penambangan data (*data mining*) dari penggunaan metode KDD. Tahap ini digunakan untuk menilai model algoritma mana yang tingkat akurasinya paling tinggi dalam upaya melakukan deteksi awal terhadap penyakit *alzheimer*. Evaluasi yang digunakan untuk mengukur tingkat akurasi model yaitu ROC (*Receiver Operating Characteristic*) yang sering disebut juga dengan kurva ROC. Metode evaluasi yang digunakan selain ROC, pada penelitian ini menerapkan *Confusion matrix* yaitu dengan melakukan pengecekan pada variable-variabel *accuracy*, *precision*, *recall*, dan *F1-score*.

- a) Penentuan nilai *Accuracy*, *Precision*, *Recall*, dan *F1-score* dapat dilihat dari persamaan berikut:

$$Accuracy = \frac{(TP+TN)}{(TP+FP+FN+TN)} \tag{1}$$

$$Precision = \frac{(TP)}{(TP+FP)} \tag{2}$$

$$Recall = \frac{(TP)}{(FP+TN)} \tag{3}$$

$$F1 - score = \frac{2 * (Precision * Recall)}{(Precision + Recall)} \tag{4}$$

Tabel 3. Hasil *Confusion matrix* kelima algoritma

No	Algoritma	Accuracy	Precision	Recall	F1-score
1	<i>Logistic Regression</i>	0.711864	0.6	0.875	0.711864
2	<i>Decision Tree</i>	0.932203	0.916667	0.916667	0.916667
3	<i>Random Forest</i>	0.983051	0.96	1	0.979592
4	<i>K-Nearest Neighbor</i>	0.40678	0.40678	1	0.578313

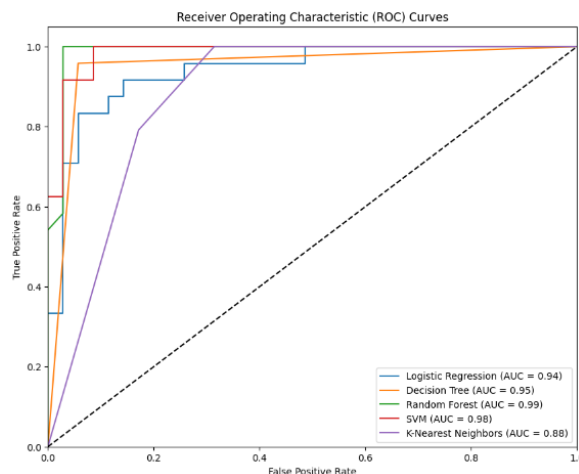
5	Support Vector Machine	0.644068	0.538462	0.875	0.666667
---	------------------------	----------	----------	-------	----------

Sumber: Hasil penelitian (2024)

Berdasarkan dari tabel 3 hasil analisis pengujian *confusion matrix* dapat disajikan sebagai berikut:

1. **Random Forest** memiliki kinerja terbaik di antara kelima algoritma dengan accuracy, precision, recall, dan F1-score yang paling tinggi. Hal ini menunjukkan bahwa Random Forest sangat efektif dalam mendeteksi penyakit Alzheimer dengan tingkat kesalahan yang sangat rendah.
  2. **Decision Tree** juga menunjukkan kinerja yang sangat baik dengan semua metrik evaluasi mendekati 0.92, yang menjadikannya algoritma yang andal untuk deteksi dini Alzheimer.
  3. **Logistic Regression** menunjukkan kinerja yang moderat dengan accuracy sekitar 71%, tetapi precision dan F1-score lebih rendah dibandingkan dengan Decision Tree dan Random Forest.
  4. **Support Vector Machine (SVM)** memiliki accuracy dan precision yang lebih rendah dibandingkan dengan Logistic Regression, tetapi recall yang tinggi menunjukkan bahwa SVM cukup baik dalam mendeteksi kasus positif namun sering menghasilkan false positives.
  5. **K-Nearest Neighbor (KNN)** menunjukkan kinerja yang paling rendah dengan accuracy dan precision di bawah 50%, meskipun recall-nya sempurna (1). Hal ini menunjukkan bahwa KNN mungkin terlalu sering memprediksi positif, yang mengurangi keefektifannya dalam prediksi.
- b) Penentuan evaluasi dengan ROC (*Receiver Operating Characteristic*).

ROC mengukur kemampuan model dalam membedakan antara kelas positif dan negatif. Evaluasi ini memberikan wawasan mendalam tentang kinerja model dan membantu peneliti memilih pendekatan yang paling efektif (Aprilia et al., 2021). ROC merupakan alat yang digunakan untuk mengukur performa klasifikasi dengan mengamati nilai AUC (Area Under Curve). Semakin tinggi nilai AUC, semakin baik dan bagus hasil klasifikasinya (Buani, 2024). Hasil nilai ROC dan AUC disajikan pada gambar 5.



Sumber: Hasil penelitian (2024)

Gambar 5. Hasil pengukuran ROC dan AUC dari kelima algoritma yang di modelkan

Model dengan AUC tertinggi adalah *Random Forest*, menunjukkan kemampuan yang baik dalam membedakan antara kelas positif dan negatif. Sebaliknya, *Logistic Regression* memiliki AUC terendah, menunjukkan potensi ruang untuk perbaikan. Berikut nilai AUC dari masing-masing algoritma: *Logistic Regression (LR)* mendapatkan hasil untuk AUC sebesar 0,94, *Decision Tree (DT)* mendapatkan hasil untuk AUC sebesar 0,95, *Random Forest (RF)* mendapatkan hasil untuk AUC sebesar 0,99, *Support Vector Machine (SVM)* mendapatkan hasil untuk AUC sebesar 0,98, dan *K-Nearest Neighbor (KNN)* mendapatkan hasil untuk AUC sebesar 0,88.

## KESIMPULAN

Penelitian ini telah mengimplementasikan dan membandingkan kinerja lima algoritma klasifikasi yaitu *Logistic Regression*, *Decision Tree*, *Random Forest*, *Support Vector Machine*, dan *K-Nearest Neighbor* dalam mendeteksi dini penyakit *Alzheimer*. Hasil dari penelitian menunjukkan bahwa algoritma *Random Forest* memiliki kinerja terbaik dibandingkan dengan algoritma lainnya.

Algoritma *Random Forest* mencapai akurasi sebesar 98%, presisi 96%, recall 100%, F1-score 97%, dan AUC 99%. Ini menunjukkan bahwa *Random Forest* mampu mengidentifikasi penderita *Alzheimer* dengan tingkat ketepatan yang sangat tinggi, serta meminimalisir kesalahan deteksi. Algoritma *Decision Tree* juga menunjukkan kinerja yang cukup baik dengan akurasi sebesar 93% dan AUC 95%, namun masih di bawah *Random Forest*.

Algoritma Logistic Regression, K-Nearest Neighbor, dan Support Vector Machine menunjukkan kinerja yang lebih rendah dengan masing-masing akurasi sebesar 71%, 64%, dan 40%.

Penelitian ini juga menekankan pentingnya penggunaan metode SMOTE (*Synthetic Minority Oversampling Technique*) untuk penyeimbangan data, yang terbukti efektif dalam meningkatkan kinerja model klasifikasi. Penggunaan pendekatan *Knowledge Discovery in Database* (KDD) yang terstruktur juga memberikan kontribusi signifikan dalam proses deteksi dini *Alzheimer*. Secara keseluruhan, penelitian ini menunjukkan bahwa algoritma Random Forest merupakan pilihan terbaik untuk deteksi dini penyakit *Alzheimer*, memberikan dasar bagi penelitian lanjutan dan panduan praktis bagi pengembang sistem kesehatan.

## REFERENSI

- Apriliah, W., Kurniawan, I., Baydhowi, M., Haryati, T., Informasi Kampus Kabupaten Karawang, S., Teknik dan Informatika, F., Bina Sarana Informatika, U., Banten No, J., & Karawang Barat, K. (2021). Prediksi Kemungkinan Diabetes pada Tahap Awal Menggunakan Algoritma Klasifikasi Random Forest. *Sistemasi: Jurnal Sistem Informasi*, 10(1), 163–171. <https://doi.org/10.32520/STMSI.V10I1.1129>
- Arfina, A. (2021). Pengaruh Edukasi Terhadap Pengetahuan Masyarakat Tentang Deteksi Dini Alzheimer Di Kelurahan Labuh Baru Pekanbaru. *Health Care: Jurnal Kesehatan*, 10(01), 256–261.
- Buani, D. C. P. (2024). Deteksi Dini Penyakit Diabetes dengan Menggunakan Algoritma Random Forest. *EVOLUSI: Jurnal Sains Dan Manajemen*, 12(1), 1–8. <https://doi.org/10.31294/evolusi.v12i1.21005>
- Fahlevi, A. (2021). *Proses Data Mining KDD*. School of Information Systems. <https://sis.binus.ac.id/2021/09/30/proses-data-mining-kdd/>
- Fakhri, D. A., Defit, S., & Sumijan. (2021). Optimalisasi Pelayanan Perpustakaan terhadap Minat Baca Menggunakan Metode K-Means Clustering. *Jurnal Informasi Dan Teknologi*, 3, 160–166. <https://doi.org/10.37034/jidt.v3i3.137>
- Farhan, Ru. (2021). *Mengenal Resampling dan Imbalanced Dataset*.
- Fauzi, A., Heri, A., & #2, Y. (2022). Optimasi Algoritma Klasifikasi Naive Bayes, Decision Tree, K – Nearest Neighbor, dan Random Forest menggunakan Algoritma Particle Swarm Optimization pada Diabetes Dataset. *JEPIN (Jurnal Edukasi Dan Penelitian Informatika)*, 8(3), 470–481. <https://doi.org/10.26418/JP.V8I3.56656>
- Kasprata, H. N., & Harahap, H. S. (2023). TINJAUAN PUSTAKA: DIAGNOSIS DAN TATALAKSANA DEMENSIA ALZHEIMER. *Jurnalmalahayati*, 10(3), 1672–1680.
- Kesehatan, J. I., Husada, S., & Husodo, B. (n.d.). *Konsumsi Kopi untuk Mencegah Penyakit Alzheimer*. <https://doi.org/10.35816/jiskh.v10i2.439>
- Leuzy, A., Smith, R., Cullen, N. C., Strandberg, O., Vogel, J. W., Binette, A. P., Borroni, E., Janelidze, S., Ohlsson, T., Jögi, J., Ossenkoppele, R., Palmqvist, S., Mattsson-Carlgrén, N., Klein, G., Stomrud, E., & Hansson, O. (2022). Biomarker-Based Prediction of Longitudinal Tau Positron Emission Tomography in Alzheimer Disease. *JAMA Neurology*, 79(2), 149–158. <https://doi.org/10.1001/jamaneurol.2021.4654>
- O'Donnell, H. (2023). A Review of Primary, Secondary, and Tertiary Prevention Strategies for Alzheimer's Disease. *Undergraduate Journal of Public Health*, 7(0). <https://doi.org/10.3998/ujph.3946>
- Pramudya, M. I. (2022). *PENERAPAN MARKET BASKET ANALYSIS MENGGUNAKAN PROSES KNOWLEDGE DISCOVERY IN DATABASE (KDD) SEBAGAI STRATEGI PENJUALAN PRODUK SWALAYAN DALAM PERSPEKTIF EKONOMI ISLAM (Studi Pada Putra Baru Swalayan Korpri)*.
- Rustiawati, C. A. B., & Pondang, J. (2024). Penggunaan Lecanemab Pada Alzheimer: Sebuah Harapan Baru. *Jurnal Sehat Indonesia (JUSINDO)*, 6(02), 479–494. <https://doi.org/10.59141/jsi.v6i02.96>
- Shakir, Y. H. (2020). *Dataset Alzheimer*.
- Spurrier, J., Nicholson, L. S., Fang, X. T., Stoner, A. J., Toyonaga, T., Holden, D., Siegert, T. R., Laird, W., Allnutt, M. A., Chiasseu, M., Brody, A. H., Takahashi, H., Nies, S. H., Pérez-Cañamás, A., Sadasivam, P., Lee, S., Li, S., Zhang, L., Huang, Y. H., ... Strittmatter, S. M. (2022). Reversal of synapse loss in Alzheimer mouse models by targeting mGluR5 to prevent synaptic tagging by C1Q. *Science Translational Medicine*, 14(647), 46–54. <https://doi.org/10.1126/scitranslmed.abi8593>
- Susanti, N., Hairini Siregar, N., Ramadhani, N., Sihite, R. N., Masyarakat, F. K., Islam, U., & Sumatera Utara, N. (2024). ALZHEIMER DAN DIMENSIA. *Jurnal Kesehatan Tambusai*, 5(2), 5736–5743. <https://doi.org/10.31004/JKT.V5I2.28471>
- Wildah, S. K., Agustiani, S., S, M. R. R., Gata, W., & Nawawi, H. M. (2020). Deteksi Penyakit Alzheimer Menggunakan Algoritma Naive Bayes Dan Correlation Based Feature Selection. *Jurnal Informatika*, 7(2), 166–173. <https://doi.org/10.31294/ji.v7i2.8226>