

**Jurnal Politeknik Caltex Riau**Terbit Online pada laman <https://jurnal.pcr.ac.id/index.php/jkt/>

| e- ISSN : 2460-5255 (Online) | p- ISSN : 2443-4159 (Print) |

Komparasi Algoritma Machine Learning Untuk Memprediksi Penyakit Alzheimer

Firman Akbar¹, Rahmaddeni^{*2}¹STMIK Amik Riau, Teknik Informatika, email: 2110031802067@sar.ac.id²STMIK Amik Riau, Teknik Informatika, email: rahmaddeni@sar.ac.id

[1] Abstrak

Penyakit Alzheimer adalah penyakit degeneratif yang menyerang banyak orang. Ini berkaitan dengan penurunan daya ingat, kesehatan mental, keterampilan berpikir dan kemampuan psikologis lainnya yang mempengaruhi individu dalam kegiatan sehari-hari, maka deteksi dini diperlukan. Salah satu metode pendeteksian penyakit alzheimer adalah memanfaatkan model algoritma machine learning dengan melatih model algoritma ini untuk mengidentifikasi seorang pasien pengidap Alzheimer berdasarkan atribut yang cocok. Tim penulis bertujuan untuk membandingkan model algoritma machine learning agar diketahui model dengan hasil yang lebih baik dalam prediksi penyakit Alzheimer. Model machine learning untuk penelitian ini dibangun dengan menggunakan algoritma ANN, LR, NB, RF, dan SVM. Tim penulis kemudian menguji dengan menggunakan 373 data pasien alzheimer yang diperoleh dari Kaggle Open Datasets dan menunjukkan bahwa model algoritma Logistic Regression mampu menghasilkan tingkat akurasi yang lebih baik yaitu sebesar 85,71%.

Kata kunci: Alzheimer, Artificial Neural Network, Logistic Regression, Machine Learning, Naïve Bayes, Random Forest, Support Vector Machine

[2] Abstract

Alzheimer's disease is a degenerative disease that affects many people. This is simultaneously characterized by a decrease in memory, mental health, thinking skills and other psychological abilities that affect individuals in their daily activities, hence early detection is required. One way to detect Alzheimer's is to use models of machine learning algorithms by training this model algorithms to identify a patient with Alzheimer's based on matching attributes. The authors' team aimed to compare models of machine learning algorithms to find the one that gives better results in prediction Alzheimer's disease. Machine learning models algorithms in this study were built using ANN, LR, NB, RF, and SVM. The author's team then tested his 373 Alzheimer's disease patient data from Kaggle Open Datasets and showed that the Logistic Regression algorithm model can achieve better with 85,71% accuracy rate.

Keywords: Alzheimer, Artificial Neural Network, Logistic Regression, Machine Learning, Naïve Bayes, Random Forest, Support Vector Machine

1. Pendahuluan

Penyakit alzheimer adalah penyebab paling utama dari demensia di seluruh dunia [1] dan jumlah kasus akan meningkat di masa depan karena populasi yang menua cenderung menekankan perlunya layanan kesehatan yang ada [2]. Dengan meningkatnya morbiditas dan mortalitas pada orang tua [3], pengobatan yang lebih baik dan identifikasi penyakit alzheimer yang tepat waktu sangat diperlukan. Dalam beberapa tahun terakhir telah adanya upaya pengembangan berupa, feritin cairan serebrospinal (CSF) sebagai biomarker [4], serta modalitas pencitraan canggih seperti amyloid dan positron-emission tomography (PET) [5].

Selanjutnya, pada zaman sekarang terapi penyembuhan penyakit Alzheimer telah memasuki bidang klinis meskipun kemanjurannya masih kontroversial. Terlepas dari kemajuan ini, banyak modalitas diagnostik dan terapi baru masih terbatas pada konteks penelitian [6]. Oleh karena itu, identifikasi otomatis dengan akurasi yang lebih baik sangat penting untuk mendeteksi dini penyakit Alzheimer [7].

Machine learning telah berkembang, andal, dan alat pendukung di bidang medis dalam beberapa tahun terakhir. Model algoritma *machine learning* juga dapat membantu mengotomatiskan prediksi penyakit Alzheimer dengan mengidentifikasi pola tersembunyi dari data menggunakan metode yang optimal, sehingga hasil prediksi dapat menjadi acuan [7].

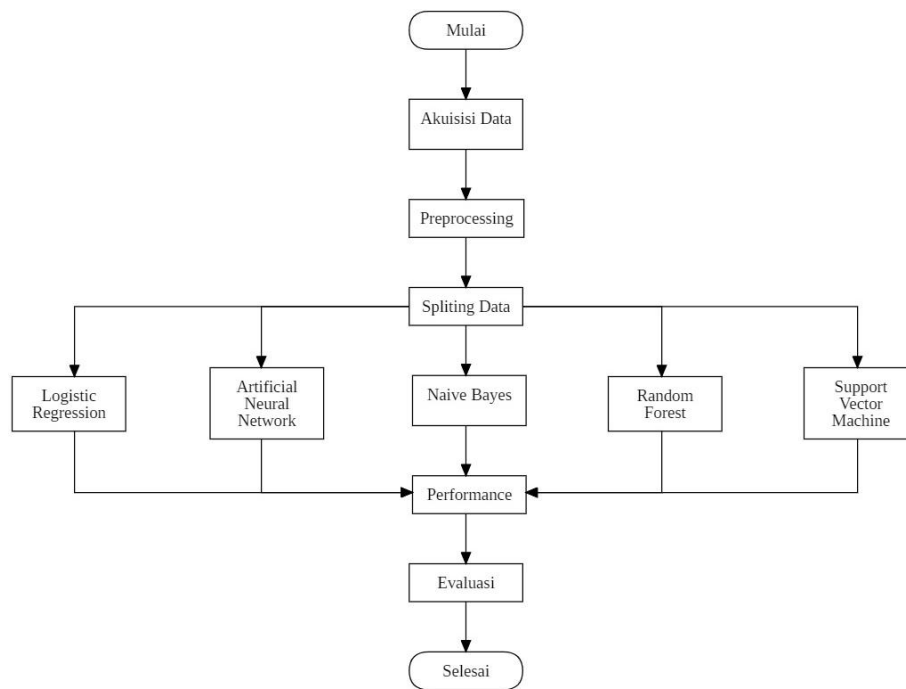
Ini adalah masalah terbesar di kota-kota kecil, negara-negara maju dan berkembang yang di mana para ahlinya langka [8]. Penelitian sebelumnya menggunakan dataset yang sama dengan rasio data latih dan uji 80:20, dengan menggunakan metode *Naïve Bayes* bersama *Correlation Based Feature Selection* sehingga memberikan hasil ketepatan yang bernilai 94,64% dan *AUC* bernilai 0,945% [9].

Dalam studi berikutnya, penelitian [10] yang menggunakan beberapa algoritma antara lain *Decision Tree*, *Logistic Regression*, *Random Forest*, dan *Support Vector Machine*. Studi ini menunjukkan bahwa *Support Vector Machine* mampu memberikan hasil dengan akurasi 92,00% dan *AUC* sebesar 0,919%, *Logistic Regression* mampu memberikan hasil dengan akurasi 74,70% dan *AUC* sebesar 0,746%, *Decision Tree* mampu memberikan hasil dengan akurasi 80,00% dan *AUC* sebesar 0,797%, sedangkan *Random Forest* mampu memberikan hasil dengan akurasi 81,30% dan *AUC* sebesar 0,767%.

Berdasarkan hal di atas, pada studi ini dilakukan pengaplikasian beberapa algoritma klasifikasi diantaranya ANN, LR, NB, RF, dan SVM terhadap 373 dataset penyakit alzheimer menggunakan perbandingan 70:30 antara data latih dan data uji yang diperoleh pada *Kaggle Open Datasets*.

2. Metode Penelitian

Penelitian ini dimulai dengan beberapa langkah terpisah, termasuk mengumpulkan data yang diperlukan untuk proses analisis. Mendistribusikan data latih dan data uji menggunakan rasio 70:30. Selanjutnya merancang model algoritma ANN, LR, NB, RF, dan SVM. Hingga pada langkah terakhir yaitu, melatih dan menganalisa performa model-model algoritma *machine learning* yang diuji penelitian ini. Adapun tahapan proses ini dapat terlihat secara visual pada berikut.



Gambar 1. Alur Penelitian

2.1 Akuisisi Data

Adapun dataset dalam studi ini adalah kumpulan data atau informasi pasien penyakit *alzheimer* yang diperoleh dari situs *Kaggle Open Datasets* (kaggle.com) sebanyak 373 data. Kumpulan data ini memiliki 8 fitur, yaitu: *Person's gender*, *Person's age*, *Education period*, *Socioeconomic status*, *Dementia diagnostic test score*, *Intracranial volume estimation*, *Normalization of brain volume*, dan *Brain volume expansion*. Berikut terlihat deskripsi kumpulan data yang digunakan.

Tabel 1. Deskripsi Kumpulan Data

No	Feature	Deskripsi
1.	<i>Gender</i>	Jenis Kelamin
2.	<i>Age</i>	Umur
3.	<i>EDUC</i>	Lama Belajar (Tahun)
4.	<i>SES</i>	Status Sosioekonomi
5.	<i>MMSE</i>	Skor Pemeriksaan Diagnosis Demenia
6.	<i>eTIV</i>	Perkiraan Volume Intrakranial
7.	<i>nWBV</i>	Normalisasi Volume Otak
8.	<i>ASF</i>	Ekspansi Volume Otak

2.2 Preprocessing

Proses ini merupakan tahapan awal sebelum melakukan pengujian model algoritma, dimana dataset yang akan digunakan akan diolah menjadi data bersih yang siap untuk diuji. *Preprocessing* bertujuan untuk menghilangkan noise dan menyeragamkan bentuk data agar sesuai dan dapat diuji dengan model algoritma yang digunakan. Proses *preprocessing* mencakup dua tahapan yaitu *cleaning data* dan *transform data*. *Cleaning data* merupakan langkah awal *preprocessing* yang bertujuan untuk menghilangkan data-data yang hilang (*missing value*) dan data yang tidak relevan atau tidak stabil [11], [12]. *Transform data* merupakan langkah terakhir *preprocessing* pada penelitian ini yang bertujuan untuk perubahan struktur fitur data yang digunakan agar menyesuaikan dengan kebutuhan model algoritma [13].

2.3 Splitting Data

Setelah tahapan *preprocessing*, tahap selanjutnya adalah mendistribusi kumpulan data ke bentuk data latih dan data uji. Di tahapan *splitting data*, *library scikit-learn* yang telah disediakan di bahasa pemrograman *Python* digunakan. Data latih ialah informasi yang akan dalam melatih model, sedangkan data uji ialah informasi yang tidak diperlukan untuk melatih model dan digunakan sebagai data atau informasi untuk mengevaluasi akurat atau tidaknya dalam melatih model [14].

2.4 Logistic Regression

Hubungan antara tipe respons dan variabel dimodelkan menggunakan *logistic regression*. Ada kombinasi tertentu dari garis linier dari variabel independen dengan probabilitas log-peluang pada situasi tertentu. Dibandingkan menggunakannya untuk regresi, algoritma *logistic regression* lebih cocok untuk masalah klasifikasi. Dalam literatur, *logistic regression* sering disebut dengan regresi *logit*, *classifier* entropi maksimum (*MaxEnt*), atau *classifier* log-linier. Kemungkinan yang menunjukkan kemungkinan dari uji tunggal dalam *logit* menerapkan fungsi *logistic*. Algoritma *logistic regression* mencakup *logistic regression biner*, *one-vs-rest*, atau *multinomial logistic regression* dengan l_1 , l_2 atau regulasi *elastic-net* [15].

2.5 Artificial Neural Network

Artificial Neural Network adalah bagian dari *Artificial Intelligence* yang mampu belajar dari data tanpa waktu yang lama [16]. Kemudian *Artificial Neural Network* juga dapat digunakan untuk mendeteksi ketidakstabilan debit yang tidak terkait dengan bentuk model [17]. *Artificial neural network* adalah algoritma *machine learning* yang memiliki sifat identik dengan jaringan saraf biologis manusia. Setiap jaringan pada *artificial neural network* adalah gabungan pada sebagian neuron. Jaringan tersebut mencakup dari satu lapisan masukan (*input layer*), satu lapisan keluaran (*output layer*), dan ada prospek satu atau lebih lapisan yang juga dikenal sebagai lapisan tersembunyi (*hidden layer*). Setiap lapisan mencakup dari beberapa neuron, dan sel-sel ini berkomunikasi terhadap sel lain di lapisan terdekatnya [18]. *Artificial Neural Network* memiliki keuntungan bahwa metode prediksi dapat belajar dari data latih dengan algoritma pembelajaran dan pelatihan layaknya otak manusia [19].

2.6 Naive Bayes

Algoritma klasifikasi yang dikenal sebagai *Naïve Bayes* adalah salah satu dari algoritma prediksi dan klasifikasi dimana *naïve bayes* memiliki kelebihan yang mencakup efektifitas (menemukan hasil yang valid) dan efisiensi (komputasi *input* sangat cepat). *Naïve Bayes* dilandaskan dengan dugaan simplifikasi bahwa nilai atribut secara tentatif saling bebas jika ada *output*. Menurut

definisi ini, *output* merupakan produk dari probabilitas individu. Salah satu kelebihan metode *Naïve Bayes* adalah menggunakan sejumlah data pelatihan (*Training Data*) untuk mengestimasi parameter yang digunakan prosedur analisis statistik. *Naïve Bayes* sering bekerja jauh meningkat umumnya dalam situasi di kehidupan nyata yang berbelit-belit ketimbang yang diharapkan [20]. Teorema *bayes* dapat diterapkan dengan persamaan dibawah ini.

$$P(H|X) = \frac{(P(X|H) \times P(H))}{(P(X))} \quad (1)$$

2.7 Random Forest

Random Forest adalah metode yang mampu menghasilkan keakuratan yang akurat karena menggunakan metode analisis sederhana untuk memilih *node*. Metode ini digunakan untuk membuat *root node*, *internal node*, dan *leaf node* dengan menggunakan atribut dan informasi yang sama terlepas dari ketentuan mana yang digunakan. *Root node* merupakan simpul terkecil, atau digunakan sebagai akar dari pohon keputusan. *Internal node* merupakan simpul percabangan, dengan minimal terdapat dua *output* dan satu *input*. Sedangkan *leaf node* atau terminal *node* merupakan simpul akhir yang terdapat satu *input* dan tidak ada *output*. Pohon keputusan didefinisikan sebagai proses komputasi nilai *entropy* sebagai perolehan ketidakmurnian atribut dan nilai *information gain* [21].

2.8 Support Vector Machine

Support Vector Machine adalah metode klasifikasi populer yang dapat menangani regresi *linear* dan *non-linear*. *Support Vector Machine* sering digunakan untuk membuat prediksi, seperti ramalan jangka panjang [22]. *Support Vector Machine* mengalami perkembangan yang sangat pesat [22], *Support Vector Machine* mampu menyelesaikan klasifikasi dan regresi dengan *linear* ataupun *non-linear kernel hyperplane* yang menjadikannya algoritma *machine learning* paling efisien untuk klasifikasi [23]. Selain kemampuan-kemampuan yang sudah dibahas, metode *Support Vector Machine* terdapat kelemahan dalam memilih parameter *Support Vector Machine* yang paling sesuai, ketepatan klasifikasi ataupun regresi metode ini ditentukan dengan kumpulan parameter yang optimal.

2.9 Evaluasi

Untuk melihat, mengevaluasi, serta menganalisis kinerja model algoritma *machine learning* tim penulis menggunakan acuan *cross validation*. *Cross validation* atau dapat disebut estimasi rotasi yang merupakan suatu teknik konfirmasi model yang digunakan dalam menentukan dengan cara menerapkan hasil analisis statistik terhadap sejumlah besar data bebas. Teknik ini umumnya diaplikasikan untuk menghasilkan model prediktif dan memverifikasi ketepatan beberapa model saat digunakan. *K-fold cross validation* ialah satu-satunya metode estimasi rotasi yang mengubah data menjadi k bagian set data yang identik [24].

Persamaa berikut digunakan untuk mengukur kinerja prediksi tiap-tiap model algoritma yang digunakan.

$$accuracy = \frac{\sum \text{klasifikasi benar}}{\sum \text{data uji}} \times 100\% \quad (2)$$

3. Hasil dan Pembahasan

Kumpulan data yang akan dipergunakan dalam studi ini adalah data terbuka yang bersumber dari *Kaggle Open Datasets*. Dataset tersebut merupakan data pasien *alzheimer* yang berjumlah 373 data.

Tabel 2. Kumpulan Data yang digunakan

No.	Group	M/F	Age	EDUC	SES	MMSE	CDR	eTIV	nWBV	ASF
1.	Nondemented	M	87	14	2.0	27.0	0.0	1987	0.696	0.883
2.	Nondemented	M	88	14	2.0	30.0	0.0	2004	0.681	0.876
3.	Demented	M	75	12	NaN	23.0	0.5	1678	0.736	1.046
4.	Demented	M	76	12	NaN	28.0	0.5	1738	0.713	1.010
...
371.	Nondemented	F	61	13	2.0	30.0	0.0	1319	0.701	1.331
372.	Nondemented	F	63	13	2.0	30.0	0.0	1327	0.796	1.323
373.	Nondemented	F	65	13	2.0	30.0	0.0	1333	0.801	1.317

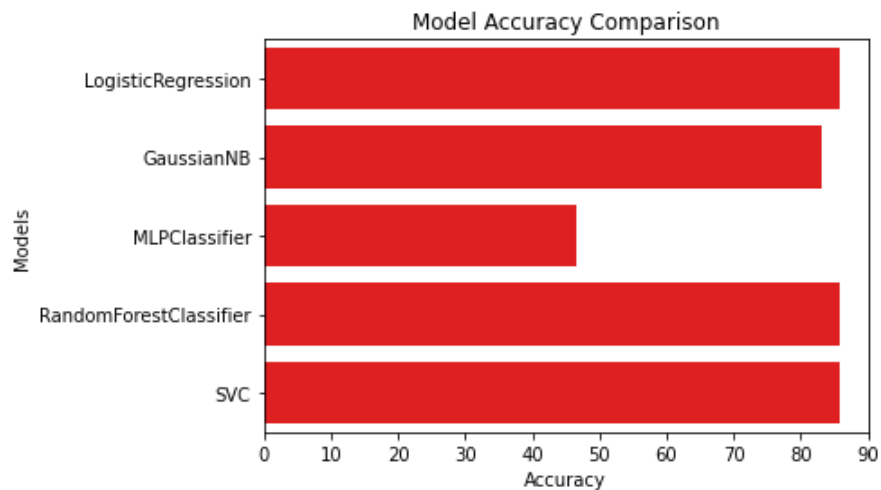
Sebelum data digunakan untuk melatih model-model algoritma *machine learning*, perlu dilakukannya *preprocessing* data terlebih dahulu, agar data dapat dipastikan sesuai dengan kebutuhan model algoritma. Proses ini terdiri dari menghilangkan data berbobot kosong atau *null* dan transformasi data.

Tabel 3. Kumpulan Data setelah Preprocessing

No.	Group	M/F	Age	EDUC	SES	MMSE	CDR	eTIV	nWBV	ASF
1.	2	1	87	14	2.0	27.0	0.0	1987	0.696	0.883
2.	2	1	88	14	2.0	30.0	0.0	2004	0.681	0.876
3.	1	1	75	12	2.5	23.0	0.5	1678	0.736	1.046
4.	1	1	76	12	2.5	28.0	0.5	1738	0.713	1.010
...
371.	2	0	61	13	2.0	30.0	0.0	1319	0.701	1.331
372.	2	0	63	13	2.0	30.0	0.0	1327	0.796	1.323
373.	2	0	65	13	2.0	30.0	0.0	1333	0.801	1.317

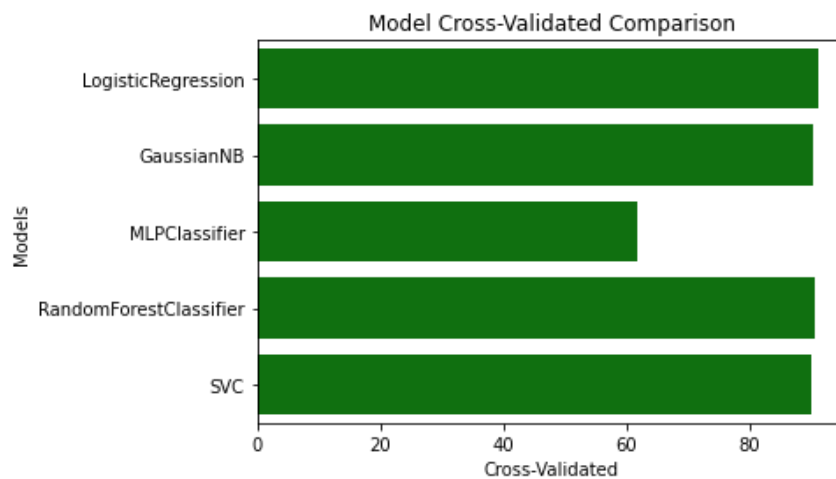
Setelah data melewati *preprocessing* ke dalam bentuk yang tepat. Selanjutnya adalah melakukan pembagian data menggunakan rasio 70 menjadi data latih dan 30 data uji, hingga dilakukannya implementasi algoritma ANN, LR, NB, RF, dan SVM.

Berikut terlihat hasil perbandingan akurasi terhadap model-model algoritma *machine learning* yang diujikan sebagai bentuk diagram batang berikut.



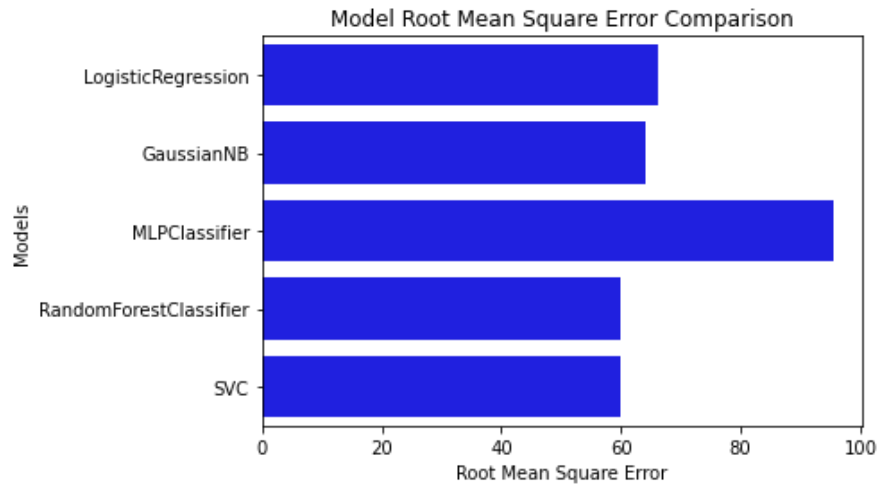
Gambar 2. Perbandingan Akurasi

Berikut terlihat hasil evaluasi terhadap model-model algoritma *machine learning* yang diujikan sebagai bentuk diagram batang berikut.



Gambar 3. Perbandingan Evaluasi Prediksi

Berikut terlihat hasil perbandingan *RMSE* terhadap model-model algoritma *machine learning* yang diujikan sebagai bentuk diagram batang berikut.



Gambar 4. Perbandingan Nilai RMSE

Dari pengujian yang telah dilakukan dan juga visualisasi yang dihasilkan dapat dirangkum dalam tabel perbandingan berikut.

Tabel 2. Hasil Perbandingan Model Algoritma Machine Learning

No	Model	Accuracy	Cross-Validated	Root Mean Square Error
1.	<i>Logistic Regression</i>	85,71%	91,20%	66,14%
2.	<i>Naïve Bayes</i>	83,04%	90,46%	64,09%
3.	<i>Artificial Neural Network</i>	46,43%	61,71%	95,43%
4.	<i>Random Forest</i>	85,71%	90,81%	59,76%
5.	<i>Support Vector Machine</i>	85,71%	90,09%	59,76%

4. Kesimpulan

Berlandaskan hasil penguraian dan riset yang sudah dilaksanakan tim peneliti untuk memprediksi penyakit *alzheimer* menggunakan model-model algoritma *machine learning* dari 373 data pasien *alzheimer* yang diperoleh dari *Kaggle Open Datasets* dapat ditarik kesimpulan. Terbukti bahwa model algoritma *Logistic Regression* memiliki performa lebih baik dari seluruh model algoritma yang diujikan pada penelitian ini dengan hasil *accuracy* 85,71%, *cross validated* 91,20%, dan *root mean square error* 66,14%.

Daftar Pustaka

- [1] L. Brand, K. Nichols, H. Wang, H. Huang, dan L. Shen, "Predicting Longitudinal Outcomes of Alzheimer's Disease via a Tensor-Based Joint Classification and Regression Model the Creative Commons Attribution Non-Commercial (CC BY-NC) 4.0 License. HHS Public Access," 2020.
- [2] K. M. Mehta dan G. W. Yeo, "Systematic review of dementia prevalence and incidence in United States race/ethnic populations," *Alzheimer's & Dementia*, vol. 13, no. 1, hlm. 72–83, 2017, doi: <https://doi.org/10.1016/j.jalz.2016.06.2360>.
- [3] N. Petrucciani *dkk.*, "Pancreatectomy combined with multivisceral resection for pancreatic malignancies: is it justified? Results of a systematic review," *HPB*, vol. 20, no. 1, hlm. 3–10, 2018, doi: <https://doi.org/10.1016/j.hpb.2017.08.002>.
- [4] S. Janelidze *dkk.*, "Plasma P-tau181 in Alzheimer's disease: relationship to other biomarkers, differential diagnosis, neuropathology and longitudinal progression to Alzheimer's dementia," *Nat Med*, vol. 26, no. 3, hlm. 379–386, 2020, doi: [10.1038/s41591-020-0755-1](https://doi.org/10.1038/s41591-020-0755-1).
- [5] N. Mattsson *dkk.*, "Predicting diagnosis and cognition with 18F-AV-1451 tau PET and structural MRI in Alzheimer's disease," *Alzheimer's & Dementia*, vol. 15, no. 4, hlm. 570–580, 2019, doi: <https://doi.org/10.1016/j.jalz.2018.12.001>.
- [6] C. Song *dkk.*, "Immunotherapy for Alzheimer's disease: targeting β -amyloid and beyond," *Transl Neurodegener*, vol. 11, no. 1, hlm. 18, 2022, doi: [10.1186/s40035-022-00292-3](https://doi.org/10.1186/s40035-022-00292-3).
- [7] J. J. Khanam dan S. Y. Foo, "A comparison of machine learning algorithms for diabetes prediction," *ICT Express*, vol. 7, no. 4, hlm. 432–439, 2021.
- [8] M. Aucoin *dkk.*, "The effect of Echinacea spp. on the prevention or treatment of COVID-19 and other respiratory tract infections in humans: A rapid review," *Adv Integr Med*, vol. 7, no. 4, hlm. 203–217, 2020.
- [9] S. Khotimatul Wildah, S. Agustiani, M. S. Ranga Ramadhan, W. Gata, H. Mahmud Nawawi, dan S. Nusa Mandiri, "Deteksi Penyakit Alzheimer Menggunakan Algoritma Naïve Bayes dan Correlation Based Feature Selection," *JURNAL INFORMATIKA*, vol. 7, no. 2, hlm. 166–173, 2020, [Daring]. Available: <http://ejournal.bsi.ac.id/ejurnal/index.php/ji/article/view/8226/0>.
- [10] M. Bari Antor *dkk.*, "A Comparative Analysis of Machine Learning Algorithms to Predict Alzheimer's Disease," *J Healthc Eng*, vol. 2021, hlm. 9917919, 2021, doi: [10.1155/2021/9917919](https://doi.org/10.1155/2021/9917919).
- [11] R. Rahmaddeni, M. K. Anam, Y. Irawan, S. Susanti, dan M. Jamaris, "Comparison of Support Vector Machine and XGBSVM in Analyzing Public Opinion on Covid-19 Vaccination," *ILKOM Jurnal Ilmiah*, vol. 14, no. 1, 2022.
- [12] R. R. Rerung, "Penerapan Data Mining dengan Memanfaatkan Metode Association Rule untuk Promosi Produk," *Jurnal Teknologi Rekayasa*, vol. 3, no. 1, hlm. 89–98, 2018, doi: [10.31544/jtera.v3.i1.2018.89-98](https://doi.org/10.31544/jtera.v3.i1.2018.89-98).
- [13] H. S. Obaid, S. A. Dheyab, dan S. S. Sabry, "The Impact of Data Pre-Processing Techniques and Dimensionality Reduction on the Accuracy of Machine Learning," dalam *2019*

9th Annual Information Technology, Electromechanical Engineering and Microelectronics Conference (IEMECON), 2019, hlm. 279–283. doi: 10.1109/IEMECONX.2019.8877011.

[14] F. Akbar, H. W. Saputra, A. K. Maulaya, M. F. Hidayat, dan R. Rahmaddeni, “Implementasi Algoritma Decision Tree C4. 5 dan Support Vector Regression untuk Prediksi Penyakit Stroke: Implementation of Decision Tree Algorithm C4. 5 and Support Vector Regression for Stroke Disease Prediction,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 2, no. 2, hlm. 61–67, 2022.

[15] J. Hao dan T. K. Ho, “Machine Learning Made Easy: A Review of Scikit-learn Package in Python Programming Language,” *Journal of Educational and Behavioral Statistics*, vol. 44, no. 3, hlm. 348–361, Feb 2019, doi: 10.3102/1076998619832248.

[16] H. Jayadianti, T. A. Cahyadi, N. A. Amri, dan M. F. Pitayandanu, “Metode Komparasi Artificial Neural Network pada Prediksi Curah Hujan-Literature Review,” *Jurnal Tekno Insentif*, vol. 14, no. 2, hlm. 47–53, 2020.

[17] L. Khanady, “PREDIKSI HARGA SAHAM DENGAN MENGGUNAKAN JST (JARINGAN SYARAF TIRUAN),” *JURNAL ILMIAH INFORMATIKA*, vol. 7, no. 01, hlm. 1–4, Mar 2019, doi: 10.33884/jif.v7i01.793.

[18] G. A. V. Pai, “Fundamentals of Neural Networks,” *NEURAL NETWORKS, FUZZY SYSTEMS AND EVOLUTIONARY ALGORITHMS: SYNTHESIS AND APPLICATIONS*, hlm. 11, 2017.

[19] P. Matondang, S. Saifullah, dan J. T. Hardinata, “Penerapan Algoritma Backproagation Untuk Memprediksi Tingkat Kerawanan Banjir di Wilayah Kabupaten Mandailing Natal,” *TIN: Terapan Informatika Nusantara*, vol. 1, no. 11, hlm. 582–586, 2021.

[20] M. Dennis, R. Rahmaddeni, F. Zoromi, dan M. K. Anam, “Penerapan Algoritma Naïve Bayes Untuk Pengelompokkan Predikat Peserta Uji Kemahiran Berbahasa Indonesia,” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 2, hlm. 1183–1190, 2022.

[21] V. W. Siburian dan I. E. Mulyana, “Prediksi Harga Ponsel Menggunakan Metode Random Forest,” dalam *Annual Research Seminar (ARS)*, 2019, vol. 4, no. 1, hlm. 144–147.

[22] E. H. Houssein, A. Hammad, dan A. A. Ali, “Human emotion recognition from EEG-based brain–computer interface using machine learning: a comprehensive review,” *Neural Comput Appl*, vol. 34, no. 15, hlm. 12527–12557, 2022, doi: 10.1007/s00521-022-07292-4.

[23] D. Chandola, A. Mehta, S. Singh, V. A. Tikkiwal, dan H. Agrawal, “Forecasting Directional Movement of Stock Prices using Deep Learning,” *Annals of Data Science*, 2022, doi: 10.1007/s40745-022-00432-6.

[24] F. Tempola, M. Muhammad, dan A. Khairan, “Perbandingan Klasifikasi Antara KNN dan Naive Bayes pada Penentuan Status Gunung Berapi dengan K-Fold Cross Validation,” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 5, no. 5, hlm. 577–584, 2018.