

Komparasi Metode *Deep Learning*, *Naïve Bayes* dan *Random Forest* untuk Prediksi Penyakit Jantung

Ivana Alhabib^{1,*}, Ahmad Faqih¹, Fatihanursari Dikananda¹

¹ Teknik Informatika; STMIK IKMI Cirebon; Alamat, Jl. Perjuangan No.10BKaryamulya, Kec. Kesambi, Kota Cirebon, Jawa Barat 45131, 0231-490480; e-mail: info.ikmicirebon@gmail.com.

* Korespondensi : e-mail: thohaalhabib@gmail.com

Diterima: 31 Agustus 2022; Review: 12 September 2022; Disetujui: 20 September 2022

Cara sitasi: Alhabib I, Faqih A, Dikananda F. Penerapan Komparasi Metode *Deep Learning*, *Naïve Bayes* Dan *Random Forest* Untuk Prediksi Penyakit Jantung. *Informatics for Educators and Professionals*. 6 (2) : 176 - 185.

Abstrak : Jantung adalah organ yang mempunyai peranan penting dalam kelangsungan hidup manusia karena fungsinya mendistribusikan darah dari paru-paru ke setiap kantong di tubuh yang faktanya bahwa darah ini mengandung banyak oksigen, jantung merupakan organ penting dalam kehidupan manusia karena membantu mengatur metabolisme di tubulus manusia.. Ada banyak aktivitas dalam tubuh manusia yang tidak dapat diprediksi dalam bentuk umum. Serangan jantung adalah salah satunya, dan itu adalah aktivitas yang sangat serius dalam tubuh manusia yang menyebabkan kematian manusia. Meskipun tidak terlalu terlihat dalam kondisi normal, itu dilakukan secara tiba-tiba, ini adalah salah satu kejadian yang sangat tidak terduga dalam tubuh manusia. Dengan kemajuan teknologi beberapa algoritma penambangan data dikembangkan untuk memprediksi serangan jantung. Selanjutnya algoritma penambangan data yang berbeda dengan pendekatan machine learning mampu memprediksi terjadinya serangan jantung dalam tubuh manusia. Ini adalah salah satu tugas diagnosis yang khas, tetapi harus dicapai secara akurat dan efisien dengan bantuan pembelajaran mesin. penelitian ini adalah upaya untuk memodelkan dan memecahkan masalah prediksi serangan jantung. Algoritma mesin yang berbeda seperti *Deep Learning*, *Naives Bayes* dan *Random Forest* diambil di sini untuk membentuk model dalam penelitian ini. pendekatan pembelajaran mesin adalah pendekatan yang baik untuk memprediksi terjadinya serangan jantung. Dataset diambil dari laman Kaggle dengan judul heart attack *analysis* dan prediction dataset. Akurasi tertinggi yang dapat dicapai dengan menggunakan algoritma deep learning, yang menghasilkan akurasi sebesar 83,49%.

Kata kunci: *heart attack*, *prediction*, *deep learning*, *random forest*, *naïve bayes*

Abstract : *The heart is an organ that has an important role in the living blood of humans because of its function of distributing the lungs to every body in the body. The fact that this bag contains a lot of oxygen, the heart is an important organ in human life because it helps regulate the human tubules. There are many activities in the human body. unpredictable human body in general form. Heart attack is one of them, and it is a very serious activity in the human body that causes human death. Although it didn't look too under normal conditions, it was done suddenly, this was one of the most unexpected occurrences in the human body. With the advancement of technology several data mining algorithms have been developed to predict heart attacks. Furthermore, different data mining with a machine learning approach is able to predict the occurrence of heart attacks in the human body. This is one of the typical diagnostic tasks, but it must be accomplished accurately and efficiently with the help of machine learning. This study is an attempt to model and solve the problem of predicting a heart attack. Different machine algorithms like Deep Learning, Naives Bayes and Random Forest are taken here to form the*

model in this research. Machine learning approach is a good approach to predict the occurrence of heart attacks. The dataset is taken from the Kaggle page with the title heart attack analysis and prediction dataset. The highest accuracy that can be achieved using yahoo deep learning, which results in an accuracy of 83.49%.

Keywords: heart attack, prediction, deep learning, random forest, naïve bayes

1. Pendahuluan

Organ utama, atau jantung, harus berfungsi dengan benar dan teratur. Jika gangguan tersebut terjadi, maka akibat yang fatal, yaitu timbulnya serangan jantung, akan terjadi. Serangan jantung termasuk 10 penyakit dengan risiko tinggi infeksi kematian. Karena kurangnya kesadaran atau pemahaman tentang fungsi jantung, gejala penyakit dan gangguan sering tidak dilaporkan oleh masyarakat umum. Ini juga akibat biaya pemeriksaan kesehatan yang mahal dan kurangnya waktu luang karena komitmen kerja. Ada empat gejala umum penyakit jantung atau kardiovaskulitis, termasuk nyeri dada (terutama dada sebelah kiri), ritme atau detak jantung yang tidak teratogenik, detak jantung yang sangat cepat dan kadang-kadang menyebabkan hasil positif palsu untuk denyut nadi sama sekali, dan sesak napas. Cara mencegah penyakit jantung maka ke empat gejala tersebut tidak boleh diabaikan dan pasien harus segera melakukan pemeriksaan ke dokter atau ahli jantung terdekat.[1]

Serangan jantung atau dalam medis bernama *Myocardial Infarction* atau *infark miokard* adalah gangguan jantung yang sangat serius. Dalam pendeteksian ini menggunakan komplikasi- komplikasi yang diderita oleh pasien.[2] Penderita penyakit *cardiovascular* (CVD) adalah gangguan pada jantung atau penyakit jantung koroner di seluruh dunia terus mengalami peningkatan dan menjadi penyakit yang paling mematikan. [3]. Hal ini disebabkan oleh faktor-faktor antara lain stres, tekanan darah tinggi, gula darah, dan lain-lain. Tujuan dari penelitian ini adalah untuk mengidentifikasi penyakit jantung menggunakan algoritma *machine learning* (ML) sebagai tindakan pencegahan sistem informasi desktop. Model pembelajaran mesin hybrid dapat meningkatkan tingkat performa teknik ML tertentu yang dikombinasikan dengan teknik ML lainnya. Akurasi yang diperoleh dari pembelajaran mesin model *hybrid* menggunakan metode *Random Forest* dan *Logistic Regression* adalah sekitar 84,48%, meningkat sekitar 1,32%. [4] Statistik survei baru-baru ini melaporkan bahwa angka kematian meningkat karena obesitas, kolesterol, tekanan darah tinggi dan penggunaan tembakau di antara orang-orang [5]

Berdasarkan pengetahuan mengenai penyakit jantung dan gejalanya maka tentu dengan penerapan teknologi informasi akan semakin mempermudah masyarakat untuk mengidentifikasi gejala yang dialami. Perkembangan teknologi yang semakin pesat sangat membantu proses mendeteksi gejala- gejala awal penyakit jantung, salah satunya perkembangan teknologi seperti kecerdasan buatan. Data *mining* adalah sebuah ilmu yang dapat menggali informasi baru dari sebuah data dan mengubahnya menjadi model baru dan banyak digunakan untuk memprediksi penyakit jantung algoritma klasifikasi *Naïve Bayes*, *Random Forest*, *Decision tree*, dan *Support Vector Machine*. [6]

Menurut penelitian yang dilakukan peneliti Devina Larassati dengan topik "Sistem Prediksi Penyakit Jantung Koroner Menggunakan Metode *Naïve Bayes*" mengatakan bahwa Penyebab dari penyakit jantung koroner yaitu penyumbatan pembuluh darah koroner, penyakit ini sangat diperhatikan oleh seluruh kalangan masyarakat dikarenakan pengaruh yang disebabkan oleh penyakit tersebut. Penelitian ini memiliki tujuan untuk membuat prediksi yang akan membantu para dokter untuk melakukan diagnose dengan tepat dan akurat sehingga penyakit jantung koroner dapat ditangani lebih awal. Salah satu algoritma klasifikasi data mining yang digunakan pada penelitian ini adalah algoritma *Naïve Bayes Classifier*. Algoritma ini diterapkan dengan tujuan untuk menghitung probabilitas kemungkinan seseorang pasien berdasarkan data rekam medis pasien.[7]. Data mining adalah cara pencarian informasi dari data dalam jumlah besar untuk keperluan berbagai aplikasi. Beberapa teknik dalam data mining dapat digunakan untuk asosiasi, klasifikasi, *clustering*, prediksi, dan pemodelan sekuensial. Machine learning digunakan dalam ilmu kedokteran untuk membantu tim medis mengetahui kondisi pasien penyakit jantung [8]

Stroke merupakan suatu penyakit dengan kondisi yang mengancam dan menjadi penyebab kematian, kelumpuhan, dan kecacatan lainnya di seluruh dunia. Jumlah penderita stroke setiap tahun semakin meningkat. Stroke dapat dicegah dengan menjalankan hidup sehat, makan makanan padat gizi, dan aktivitas fisik. [9] Untuk melakukan ini, model yang

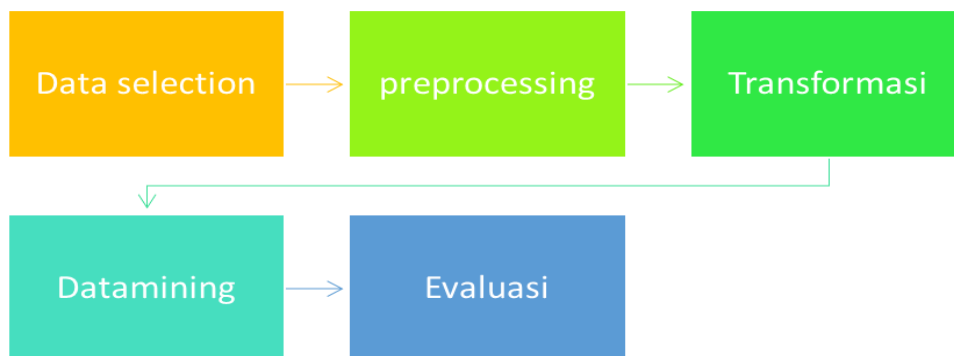
efektif untuk memprediksi stroke harus dibuat dengan menggunakan parameter dari faktor gaya hidup, faktor yang dapat disesuaikan, seperti faktor risiko medis, dan faktor yang tidak dapat disesuaikan. Prediksi mungkin tidak selalu memberikan hasil atau prediksi yang akurat; melainkan, mereka berfungsi sebagai panduan untuk memahami dan memperoleh prediksi yang mendekati apa yang bisa terjadi.[7] Penelitian ini Perlu dilakukan klasifikasi dengan pendekatan algoritma *deep learning*, *Naïve Bayes* dan *random forest* untuk memberikan informasi diagnose penyakit jantung dengan tepat dan akurat dan Perlu dilakukan komparasi antara algoritma *deep learning*, *Naïve Bayes* dan *random forest* untuk mengetahui perbedaan dari ketiga algoritma tersebut

Akar masalah dari penelitian ini memandang perlu untuk menerapkan komparasi metode *deep learning*, *naïve bayes* dan *random forest* untuk memprediksi penyakit jantung terhadap data sebanyak 303 data *public* [Heart Attack Analysis & Prediction Dataset | Kaggle](#). yang akan dijadikan bahan penelitian.

2. Metode Penelitian

Pada Penelitian ini penulis menggunakan metode penelitian kuantitatif. Data Kuantitatif adalah data *public* [Heart Attack Analysis & Prediction Dataset | Kaggle](#) yang akan di analisis menggunakan pendekatan metode *deep learning*, *naive bayes* dan *random forest*. Jenis data yang digunakan dalam penelitian ini bersifat rasio yang merupakan data yang berbentuk angka sebenarnya.[10]

Tahap penelitian adalah proses yang memperoleh, mendapatkan, menerima, atau mendiskusikan informasi tentang sesuatu atau memecahkan masalah yang telah diidentifikasi. Hal ini dilakukan secara ilmiah, sistematis, dan logis. Tahapan Penelitian pada Penelitian ini adalah sebagai berikut:



Sumber : Hasil Penelitian (2022)

Gambar 1. Tahapan Penelitian

3. Hasil dan Pembahasan Hasil

Data yang digunakan saat ini akan diseleksi dengan cara mencocokkan data dengan topik atau judul penelitian yang akan diterbitkan, dalam hal ini data yang berasal dari data *public* [Heart Attack Analysis & Prediction Dataset | Kaggle](#) sudah sesuai dengan format data yang terdiri dari 13 atribut *prediktor* yaitu *age*, *sek*, *Cp*, *Trestbps*, *Chol*, *Fbs*, *Restecg*, *Thalach*, *Exang*, *Oldpeak*, *Slope*, *Ca*, *Thal* dan *atribu target 1* yaitu *prediksi*. Dibawah ini adalah data set terdiri dari 303 *record*

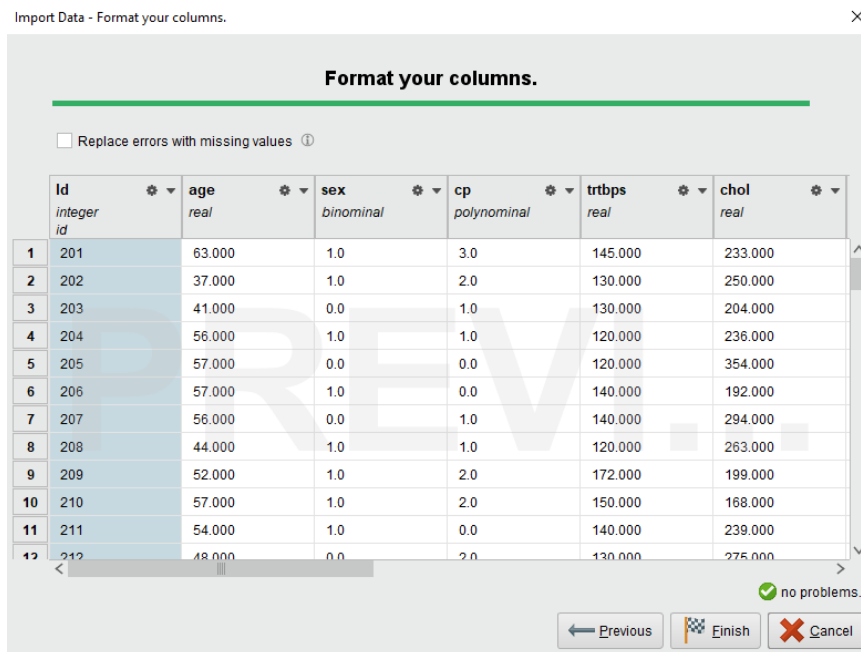
Preprocessing data pada penelitian ini menggunakan Operator *replace missing value*. Operator ini menggantikan nilai yang hilang dengan nilai minimum, maksimum, atau rata-rata dari Atribut yang sesuai. Nol juga dapat digunakan untuk melawan mata uang bernilai rendah. Setiap nilai pengisian memiliki potensi untuk dilihat sebagai pengganti nilai yang di luar jangkauan. Ada operator yang menggantikan nilai yang hilang, seperti yang terlihat pada gambar 2.



Sumber : Hasil Penelitian (2022)

Gambar 2. Operator replace missing value

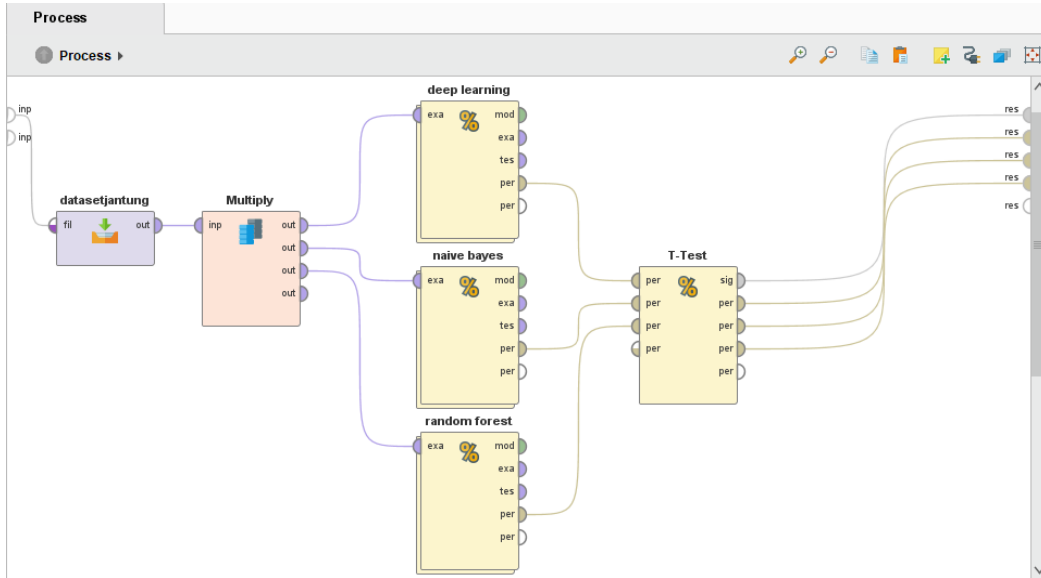
Pada proses data transformasi Data akan ditransformasikan ke dalam bentuk yang sesuai untuk proses data mining. Karena sifat teoritis dari metodologi penelitian dan penggunaan program data mining RapidMiner, maka data yang diproses sebelumnya akan diubah sehingga akan kompatibel dengan algoritma yang dipilih penelitian yaitu, algoritma algoritma *deep learning*, *naïve bayes* dan *random forest*, pada penelitian ini operator *replace missing value*.



Sumber : Hasil Penelitian (2022)

Gambar 3. Hasil operator *replace missing value*

Pada penelitian ini menggunakan teknik data mining yang sesuai. Kemudian data tersebut akan distandarisasi melalui proses data mining sehingga aman dan dapat digunakan dengan menggunakan RapidMiner dengan metode klasifikasi algoritma *deep learning*, *naïve bayes* dan *random forest*. Kemudian data *penyakit jantung* di import masukan operator *multiply*, *replace missing value*, algoritma *deep learning*, *Naïve Bayes* dan *random forest*, *T-Test*, *Apply Model* dan operator *Performance*. Semua operator dihubungkan sehingga akan membentuk model seperti pada Gambar 2 Model Proses Algoritma *deep learning*, *Naïve Bayes* dan *random forest*.



Sumber : Hasil Penelitian (2022)

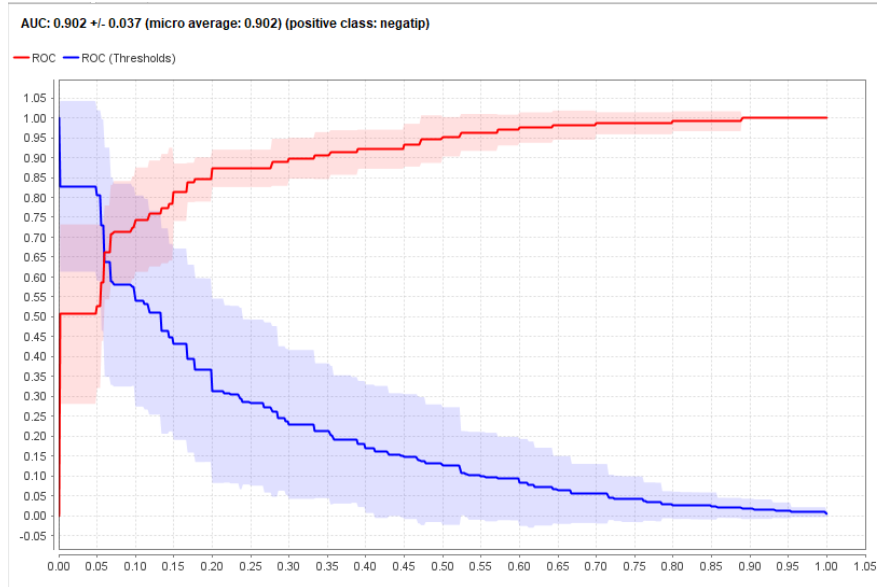
Gambar 4. Model Proses Pengujian Algoritma *deep learning*, *Naïve Bayes* dan *random forest*

Kurva ROC adalah tool dua dimensi yang digunakan untuk menilai kinerja klasifikasi yang menggunakan dua class keputusan, masing-masing objek diarahkan ke salah satu elemen dari himpunan pasangan, positif atau negatif. Pada kurva ROC, TP rate diplot pada sumbu Y dan FP rate diplot pada sumbu X. Untuk klasifikasi data mining, nilai AUC dapat dibagi menjadi beberapa kelompok, yaitu:

Tabel 1 *Performance Keakurasian AUC*

Performance	Klasifikasi
0.90-1.00	Excellent classification
0.80-0.90	Good Classification
0.70-0.80	Fair Classification
0.60-0.70	Poor Classification
0.50-0.60	Failure

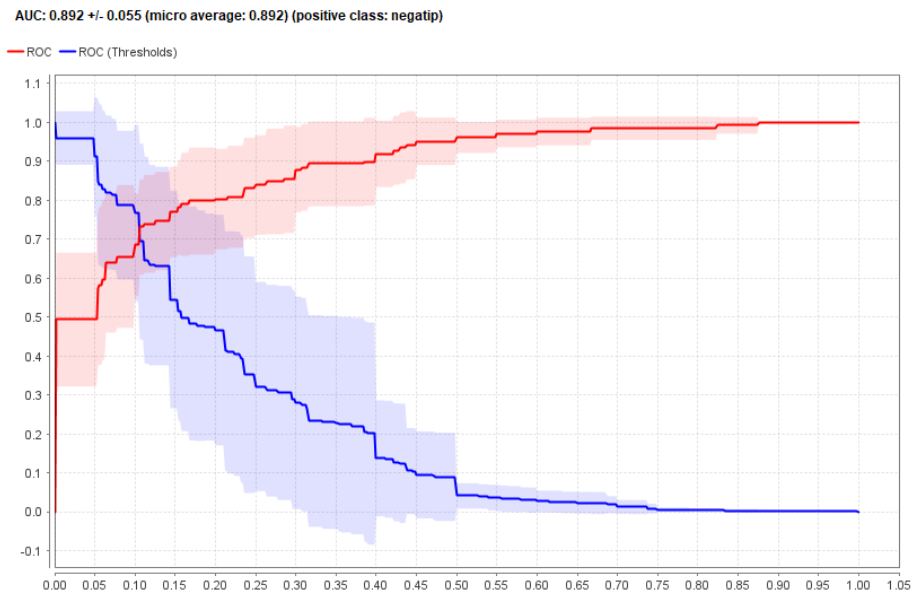
Sumber : Hasil Penelitian (2022)



Sumber : Hasil Penelitian (2022)

Gambar 5. Nilai AUC dengan algoritma deep learning

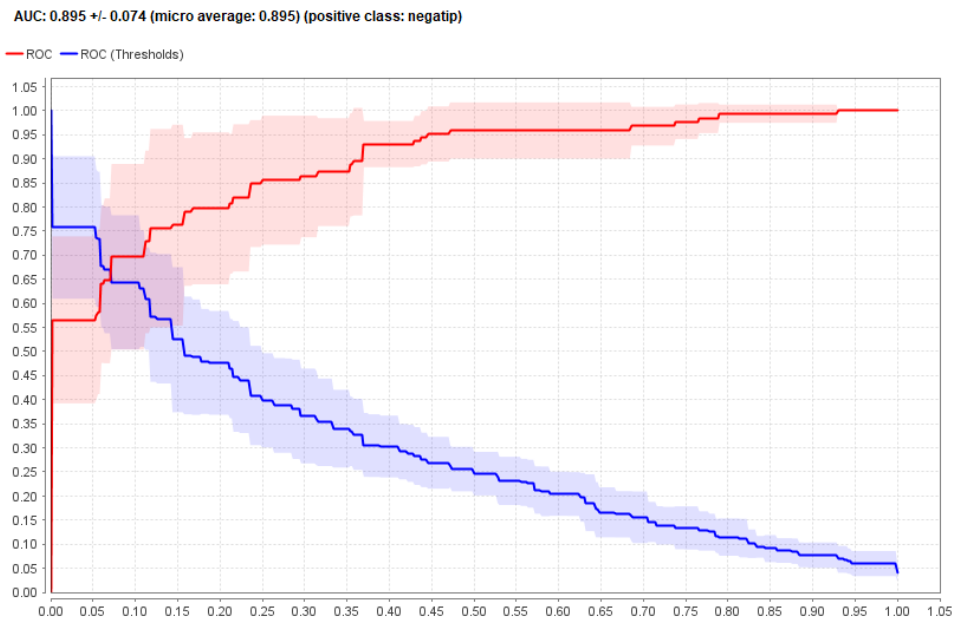
Dari pengujian dengan model algoritma *deep learning*, didapatkan grafik ROC seperti Gambar 4 dengan nilai AUC (Area Under Curve) sebesar 0,902 dengan diagnosa hasilnya *Excellent classification*.



Sumber : Hasil Penelitian (2022)

Gambar 6. Nilai AUC dengan algoritma Naïve Bayes

Dari pengujian dengan model algoritma *naïve bayes*, didapatkan grafik ROC seperti Gambar 5 dengan nilai AUC (Area Under Curve) sebesar 0,892 dengan diagnosa hasilnya *Good Classification*.



Sumber : Hasil Penelitian (2022)

Gambar 7. Nilai AUC dengan algoritma Random Forest

Dari pengujian dengan model algoritma *random forest*, didapatkan grafik ROC seperti Gambar 6 dengan nilai AUC (*Area Under Curve*) sebesar 0,895 dengan diagnosa hasilnya *Good Classification*.

Dalam percobaan penelitian ini menggunakan dataset yang terdiri dari 13 atribut, Metode yang diuji adalah algoritma klasifikasi deep learning, Naïve bayes dan random forest, Hasil eksperimen yang disajikan adalah accuracy, precision, recall, dan AUC.

Tabel 2 Tabel hasil eksperimen algoritma linier regresi

Test	Criteria	Deep Learning	Naïve Bayes	Random Forest
1	Accuracy	79.54%	81.88%	80.18%
	Precision	79.61%	81.48%	80.05%
	Recall	74.73%	78.90%	76.15%
	AUC	0.893	0.889	0.898
2	Accuracy	77.91%	81.88%	80.18%
	Precision	76.82%	81.48%	80.05%
	Recall	76.81%	78.90%	76.15%
	AUC	0.883	0.889	0.898
3	Accuracy	78.26%	81.88%	80.18%
	Precision	74.67%	81.48%	80.05%
	Recall	81.04%	78.90%	76.15%
	AUC	0.899	0.889	0.898
4	Accuracy	82.53%	81.88%	80.18%
	Precision	81.41%	81.48%	80.05%
	Recall	80.38%	78.90%	76.15%
	AUC	0.896	0.889	0.898
5	Accuracy	83.49%	81.84%	80.18%
	Precision	82.32%	82.50%	78.94%
	Recall	81.76%	76.29%	74.81%
	AUC	0.902	0.892	0.895

Sumber : Hasil Penelitian (2022)

Hasil *performancevektor* algoritma *deep learning*, *naïve bayes* dan *random forest* sebagai berikut :

Tabel 3 Tabel perbandingan *performancevektor* algoritma *deep learning*, *naïve bayes* dan *random forest*

No	Algoritma	PerformanceVektor
1	<i>Deep learning</i>	83,49%
2	<i>Naïve Bayes</i>	81,84%
3	<i>Random Forest</i>	80,18%

Sumber : Hasil Penelitian (2022)

Berdasarkan hasil komparasi algoritma *deep learning*, *naïve bayes* dan *random forest* dalam memprediksi penyakit jantung dapat dilihat pada table 3 tabel perbandingan *performancevektor algoritma deep learning, naïve bayes dan random forest* bahwa algoritma *deep learning* menghasilkan nilai akurasi tertinggi dengan nilai akurasi 83,49%.

Akurasi dalam klasifikasi menentukan persentase ketepatan record data yang di klasifikasikan secara benar. Nilai *accuracy* dari model Algoritma *deep learning* untuk prediksi diagnose penyakit jantung berdasarkan penerapan komparasi metode *deep learning, Naive bayes* dan *random forest* dalam penelitian ini dari hasil pengujian menghasilkan nilai akurasi terbesar adalah algoritma *Deep Learning* dengan nilai akurasi sebesar 83.49% precision: 82.32% recall: 81.76%, AUC : 0.902 +/- 0.037 (micro average: 0.902) (positive class: negative)

Pembahasan

Algoritma Deep Learning, Naive Bayes, dan Random Forest akan digunakan sebagai tiga algoritma untuk pengujian model perbandingan pada percobaan ini. Model terbaik akan dipilih dari masing-masing model berdasarkan hasil percobaan. Pengukuran terhadap model algoritmik diperlukan untuk menentukan model terbaik. Masing-masing algoritma memiliki tingkat akurasi rata-rata sebesar 83,49% untuk model Deep Learning, 81,84% untuk model Naive Bayes, dan 80,48% untuk model Random Forest. untuk evaluasi model menggunakan Area Under Curve (AUC), hasilnya menunjukkan bahwa Deep Learning mencapai skor sekitar 0,902, Naive Bayes mencapai skor sekitar 0,892, dan model Random Forest mencapai skor sekitar 0,895.

accuracy: 83.49% +/- 4.43% (micro average: 83.50%)

	true positif	true negatif	class precision
pred. positif	141	26	84.43%
pred. negatif	24	112	82.35%
class recall	85.45%	81.16%	

Sumber : Hasil Penelitian (2022)

Gambar 8. Hasil algoritma Deep Learning

accuracy: 81.84% +/- 4.96% (micro average: 81.85%)

	true positif	true negatif	class precision
pred. positif	142	32	81.61%
pred. negatif	23	106	82.17%
class recall	86.06%	76.81%	

Sumber : Hasil Penelitian (2022)

Gambar 9. Hasil algoritma Naïve Bayes

accuracy: 80.48% +/- 8.61% (micro average: 80.53%)

	true positif	true negatif	class precision
pred. positif	139	33	80.81%
pred. negatif	26	105	80.15%
class recall	84.24%	76.09%	

Sumber : Hasil Penelitian (2022)

Gambar 10. Hasil algoritma Random Forest

Kesimpulan

Penerapan kopmarasi metode algoritma *deep learning*, *naïve bayes* dan *random forest* dalam memprediksi diagnose penyakit jantung menghasilkan nilai akurasi *performancevector* terbaik algoritma *Deep Learning* sebesar 83,49%. Hasil klasifikasi berdasarkan hasil pengujian didapat grafik ROC seperti pada gambar 3 dengan nilai AUC (Area Under Curve) sebesar 0.902 dengan diagnosa hasilnya *Excellent classification*. Sehingga dari penerapan ketiga algoritma tersebut terhadap data diagnose penyakit jantung hasil *performance* terbaik adalah algoritma *Deep Learning*.

Referensi

- [1] J. J. Pangaribuan *et al.*, "mendeteksi penyakit jantung menggunakan *machine learning* dengan algoritma *logistic regression*" vol. 6, no. 2, 2021.
- [2] Yudistira, Novanto Putra, Aldi Fianda "Algoritma Decision Tree Dan Smote Untuk Klasifikasi Serangan Jantung Miokarditis Yang Imbalance" vol. 2, no. 2, pp. 112–122, 2021.
- [3] W. Nugraha, "Prediksi penyakit jantung cardiovascular menggunakan model algoritma klasifikasi," no. February, 2022.
- [4] S. Ath *et al.*, " *Hybrid machine learning* model untuk memprediksi penyakit jantung dengan metode *logistic regression* dan *random forest*," vol. 8, no. 1, pp. 40–46, 2022.
- [5] M. Swathy and K. Saruladha, "A comparative study of classification and prediction of Cardio-Vascular Diseases (CVD) using Machine Learning and Deep Learning techniques," *ICT Express*, vol. 8, no. 1, pp. 109–116, 2022, doi: 10.1016/j.icte.2021.08.021.
- [6] D. A. Firdlous, "Komparasi Algoritma Klasifikasi Data Mining untuk Memprediksi Penyakit Jantung," vol. 16, no. 1, 2022.
- [7] D. Larassati, A. Zaidiah, and S. Afrizal, "Sistem prediksi penyakit jantung koroner menggunakan metode *naïve bayes*," vol. 07, pp. 533–546, 2022.
- [8] Mulyawan, A. Bahtiar, G. Dwilestari, F. M. Basysyar, and N. Suarna, "Data mining techniques with machine learning algorithm to predict patients of heart disease," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1088, no. 1, p. 012035, 2021, doi: 10.1088/1757-899x/1088/1/012035.
- [9] R. A. Zuama, S. Rahmatullah, and Y. Yuliani, "Analisis Performa Algoritma Machine Learning pada Prediksi Penyakit Cerebrovascular Accidents," vol. 6, pp. 531–534, 2022, doi: 10.30865/mib.v6i1.3488.
- [10] I. Mahendra, "Implementasi kebijakan pendataan pemilih dalam pemilihan umum kepala daerah kota Malang 2013 (*Policy Implementation of Voters Data Collection In Mayor*

- Election Of Malang City,*” vol. 8, pp. 28–36, 2018.
- [11] R. Annisa *et al.*, “Analisis Komparasi Algoritma Klasifikasi Data Mining Untuk Prediksi Penderita Penyakit Jantung,” vol. 3, no. 1, 2019.