

Perbandingan Algoritma *Machine Learning* pada Klasifikasi Penyakit Jantung

Musriatun Napiah^{1*}, Sujiliani Heristian²

¹Ilmu Komputer, Universitas Bina Sarana Informatika
e-mail: musriatun.mph@bsi.ac.id

²Teknologi Komputer, Universitas Bina Sarana Informatika
e-mail: sujiliani.she@bsi.ac.id

Diterima	Direvisi	Disetujui
01-04-2024	06-05-2024	06-05-2024

Abstrak - Jantung adalah organ utama yang memompa darah keseluruh tubuh bergerak melalui sistem peredaran darah. Menurut WHO, penyakit jantung koroner (PJK) telah menjadi masalah kesehatan yang meningkat pesat, menyebabkan 6,7 juta kematian pada tahun 2017. Banyak alternatif atau cara yang digunakan untuk mencegah dan mendeteksi penyakit jantung, kekurangan pengetahuan sering kali membuat penderita terlambat memeriksa diri ke dokter. Tujuan dari penelitian ini adalah untuk mendeteksi penyakit jantung sejak dini akibat gangguan kardiovaskular sehingga dapat memungkinkan untuk pencegahan yang lebih efektif dan pengelolaan yang baik terhadap kondisi kardiovaskular, dengan menggunakan dataset dari <http://archive.ics.uci.edu/ml/> sebanyak 1026 pasien penyakit jantung. Metode yang digunakan pada penelitian ini adalah *Machine Learning* dengan algoritma *logistic regression*, *naive bayes*, dan *k-nearest neighbour* (KNN) . Hasil yang didapatkan dari penelitian ini yang tertinggi adalah dengan menggunakan metode *k-nearest neighbour* (KNN) yaitu akurasi sebesar 91%, sedangkan dengan algoritma *logistic regression* akurasinya sebesar 87%, dan *naive bayes* akurasinya sebesar 83%.

Kata Kunci: Jantung, *logistic regression*, *naive bayes*, dan *k-nearest neighbour*

Abstract - The heart is the main organ that pumps blood throughout the body through blood vessels. WHO notes that coronary heart disease (CHD) is a health problem that is increasing rapidly, causing 6.7 million deaths in 2017. There are many alternatives or methods used to prevent and detect heart disease, lack of knowledge often makes it too late for sufferers to check themselves. doctor. The aim of this research is to detect heart disease early due to cardiovascular disorders so that it can allow for more effective prevention and good management of cardiovascular conditions, using a dataset from <http://archive.ics.uci.edu/ml/> of 1026 heart disease patients. The method used in this research is machine learning with *logistic regression*, *naive Bayes*, and *k-nearest neighbor* (KNN) algorithms. The highest results obtained from this research were using the *k-nearest neighbor* (KNN) method, namely an accuracy of 91%, while with the *logistic regression* algorithm the accuracy was 87%, and *naive Bayes* the accuracy was 83%.

Keyword : Heart, *Logistic Regression*, *Naive Bayes*, and *K-Nearst Neighbor*

PENDAHULUAN

Jantung adalah salah satu organ yang memegang peranan penting dalam sistem peredaran darah manusia. Penyakit jantung merupakan salah satu penyakit yang paling mematikan di dunia, dan istilah untuk semua jenis gangguan yang mempengaruhi jantung. Penyakit jantung tidak sama dengan penyakit kardiovaskular mengacu pada gangguan pembuluh darah dan jantung, sedangkan penyakit jantung mengacu hanya pada bagian hati (Annisa et al., 2019).

Di Indonesia, penyakit jantung merupakan kondisi yang paling umum terjadi pada wanita dewasa, sesuai dengan informasi tentang kasus penyakit di negara ini melalui Data Indonesia menurut provinsinya, penyakit jantung merupakan masalah serius di Indonesia, dengan jumlah kasus yang tinggi dan angka kematian yang meningkat. Fakta bahwa biaya klaim terbesar dalam program Jaminan Kesehatan Nasional (JKN) disebabkan oleh penyakit jantung menunjukkan dampak yang signifikan dari kondisi ini terhadap sistem kesehatan dan ekonomi negara. Angka 15,5 juta kasus penyakit

jantung di Indonesia pada tahun 2022 menunjukkan besarnya beban penyakit ini terhadap populasi (DataIndonesia.id, 2023). Dari informasi tersebut, terlihat bahwa banyak individu yang belum menganggap serius faktor penyebab penyakit ini. Setelah menjalani pemeriksaan kesehatan, dokter menemukan bahwa penyakit ini telah mencapai tingkat stadium yang lanjut (Journal et al., 2019).

Kurangnya akses informasi/media dalam mencari tahu tentang penyakit jantung sehingga menyebabkan keterlambatan untuk pemeriksaan awal ke dokter, merupakan penyebab banyaknya angka kematian semakin tinggi terhadap penyakit ini. Dengan demikian diagnosa sejak dini sangat penting untuk dilakukan (Utomo, 2020), Sehingga diperlukan sistem klasifikasi untuk mendeteksi sejak dini tentang penyakit jantung, agar dapat memberikan informasi tentang penyakit jantung yang di derita oleh seseorang.

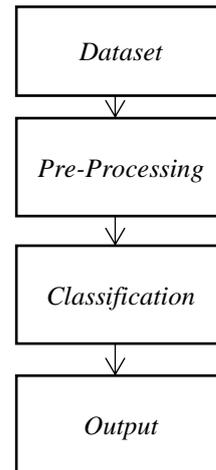
Machine Learning (ML) merupakan salah satu penerapan dari Kecerdasan Buatan (AI) yang berfokus pada pengembangan sistem yang mampu belajar secara mandiri tanpa perlu diprogram berulang kali. ML memerlukan data (biasanya disebut sebagai data pelatihan) sebagai bagian dari proses pembelajaran sebelum menghasilkan output. Menggunakan algoritma *Machine Learning* untuk mendiagnosis penyakit kanker diharapkan dapat menghasilkan prediksi yang akurat dan memungkinkan peningkatan dalam deteksi dini serta perawatan yang lebih efektif bagi pasien (Informatika & Informasi, 2020).

Naïve Bayes merupakan salah satu metode yang dapat digunakan untuk mengklasifikasikan data. Bayesian classification merupakan pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class (Damuri et al., 2021). KNN merupakan algoritma yang efektif untuk data yang besar, tahan terhadap data pelatihan yang noise, dan memiliki performa yang baik (Putry et al., 2022), sedangkan *Logistic Regression* merupakan klasifikasi linier yang menangani masalah klasifikasi multi kelas dan telah terbukti menghasilkan klasifikasi yang powerful (Prasetyo et al., 2021).

Penelitian ini akan membandingkan algoritma Naive Bayes, Regresi Logistik, dan KNN dalam klasifikasi penyakit jantung. Diharapkan penelitian ini akan memberikan acuan dan referensi serta memajukan pengetahuan untuk penelitian mendatang.

METODE PENELITIAN

Penelitian yang sudah dilakukan merupakan hasil eksperimen yang bertujuan untuk membandingkan hasil dari algoritma yang digunakan pada data penyakit jantung. Metode yang diusulkan sebagai berikut pada gambar.



Sumber : peneliti (2024)

Gambar 1. Metode Penelitian

1. Dataset

Dataset merupakan data mentah berupa tabel yang akan kita olah dengan menggunakan algoritma yang diusulkan. Penelitian ini menggunakan data pasien penyakit jantung diambil dari sumber yang terdapat di <http://archive.ics.uci.edu/ml/>, dengan total 1026 rekaman dan 76 atribut.

2. Pre – Processing

Pre-processing adalah proses pengolahan data atau citra asli sebelum data atau citra tersebut diolah oleh algoritma yang diusulkan pada metode penelitian (Cnn et al., 2020). Preprocessing data adalah tahap di mana data yang akan digunakan untuk pelatihan disiapkan terlebih dahulu. Dalam penelitian ini, preprocessing data dilakukan dengan proses untuk mengatasi data yang kosong atau hilang, yang melibatkan metode seperti mencari rata-rata atribut untuk kelas yang serupa (Rizki et al., 2020).

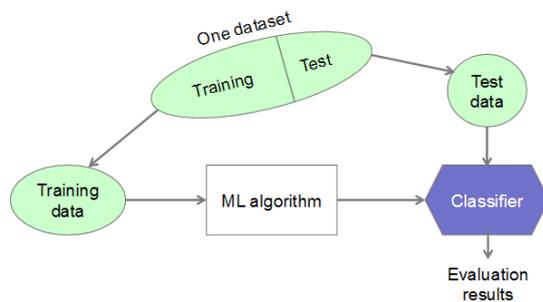
Langkah-langkah pada tahap *pre-processing* meliputi pembagian data menjadi dua bagian, sebagai berikut:

1. Train Data

Data training merupakan himpunan data yang telah diberi label atau kelas tertentu, yang digunakan oleh mesin untuk mengenali fitur-fiturnya serta memahami pola yang ada. Dengan memahami pola ini, mesin dapat membuat model atau pola data yang digunakan untuk tujuan tertentu, seperti klasifikasi atau prediksi (Musu et al., 2021). Data pelatihan (train data) sebanyak 20% untuk mengevaluasi algoritma.

2. Testing Data

Data testing adalah kumpulan data yang juga memiliki label atau kelas, yang digunakan untuk menguji keakuratan pola atau model yang telah dibuat dalam mengklasifikasikan data. Saat melakukan pengujian model, atribut label dari data testing disembunyikan selama proses klasifikasi, dan kemudian digunakan untuk membandingkan hasil klasifikasi dengan label sebenarnya. Hal ini memungkinkan untuk menilai seberapa baik model tersebut dalam melakukan klasifikasi dengan akurasi yang sesuai dengan data sebenarnya (Musu et al., 2021). Data pengujian (testing data) sebanyak 80% untuk menentukan metode algoritma yang paling sesuai dalam penelitian.

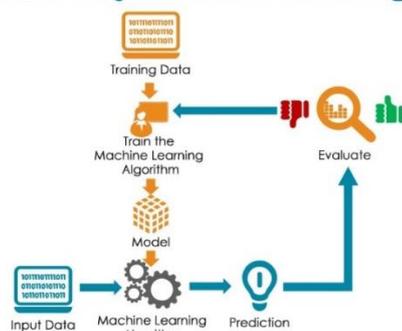


Sumber : (Caballe, 2019)
Gambar 2. Data Train dan Data Testing

3. Machine Learning

Machine learning adalah penerapan komputer dan algoritma matematika yang menggunakan pembelajaran dari data untuk menghasilkan prediksi pada periode yang akan datang (Roihan et al., 2020).

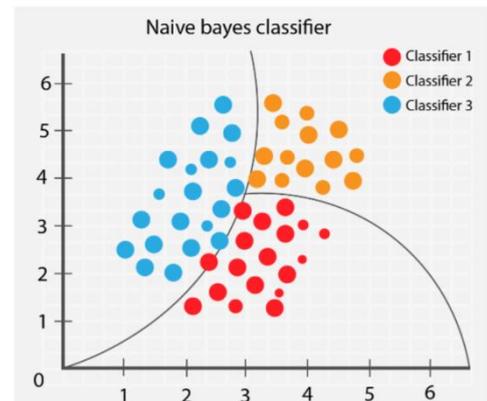
Machine Learning berhubungan erat dengan komputasi variabel, fokusnya adalah pada membuat prediksi menggunakan komputer. Studi tentang optimasi matematika menyediakan metode, teori, dan aplikasi di berbagai bidang pembelajaran mesin. Data mining adalah disiplin studi dalam pembelajaran mesin, yang menitikberatkan pada eksplorasi dan analisis data melalui pembelajaran tanpa pengawasan (Pratama, 2020).



Sumber : (dqlab.id, 2021)
Gambar 3. Cara Kerja Machine Learning

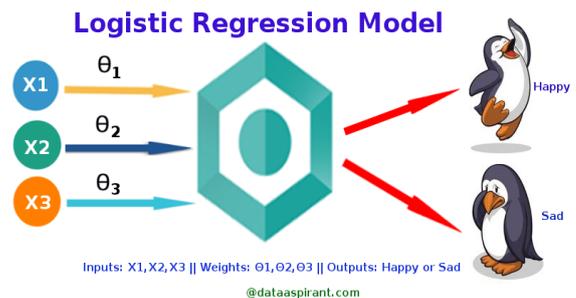
Klasifikasi adalah proses yang didasarkan pada kelas dan *variable dependen* (Prasetio & Ripandi, 2019). Klasifikasi adalah metode pembelajaran yang disupervisi (*supervised learning*) yang memerlukan data pelatihan yang sudah diberi label untuk menghasilkan aturan yang membagi atau mengklasifikasikan data uji ke dalam kelompok atau kelas yang telah ditentukan sebelumnya (Bimo et al., 2020). Model klasifikasi yang populer adalah *Decision Tree*, *Naive Bayes*, *Neural Network*, *Genetic Algorithm*, dan *Support Vector Machine* (Prasetio & Ripandi, 2019).

Naive Bayes adalah metode klasifikasi yang berasal dari *teorema Bayes*. Metode ini mengandalkan Probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris bernama Thomas Bayes, bertujuan untuk meramalkan kemungkinan di masa mendatang berdasarkan pengalaman yang terjadi di masa sebelumnya. Karena sifatnya yang mengasumsikan independensi antara fitur-fitur, metode ini disebut sebagai *Naive Bayes* (Watratn et al., 2020).



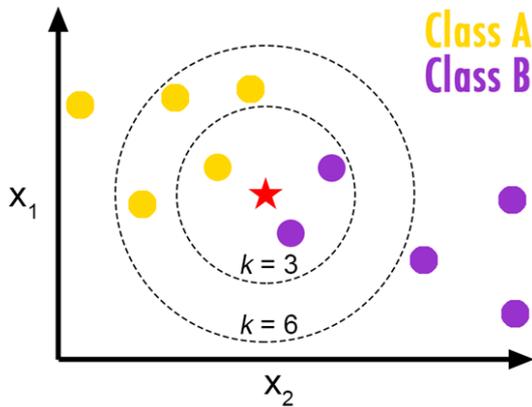
Sumber : (analytics vidhya, 2022)
Gambar 4. Algoritma Naive Bayes

Logistic Regression merupakan bagian dari metode data mining yang digunakan untuk menganalisis data yang menjelaskan hubungan antara satu variabel *respon (dependen)* dengan satu atau lebih variabel *prediktor* (Reviatika et al., 2021).



Sumber : (Mining, 2020)
Gambar 5. Model Logistic Regression

Metode KNN adalah algoritma yang dipakai dalam mengklasifikasikan data dengan jarak yang dekat, dan merupakan teknik *lazy learning* yang populer dalam kategori pembelajaran berbasis instansi (Puspita & Widodo, 2021).



Sumber : (suprianto, 2020)

Gambar 6. Model KNN

HASIL DAN PEMBAHASAN

Dari penelitian yang sudah dilakukan berikut adalah hasil yang didapatkan :

1. Dataset

Dataset yang diterapkan dalam penelitian ini adalah Ada 1025 data pasien dengan penyakit jantung, di mana 80% atau 820 data telah dipilih sebagai data pelatihan, sementara 20% sisanya digunakan sebagai data uji. Atribut-atribut yang dimasukkan meliputi usia (dalam tahun), jenis kelamin, cp, trestbps (tekanan darah istirahat pada saat masuk rumah sakit), chol (kolesterol serum), fbs (gula darah puasa > 120 mg/dl), restecg, dan thalach (denyut jantung maksimum yang dicapai), dan *target* (*diagnosis of heart disease*). Dataset dapat dilihat pada gambar 7.

age	sex	cp	trestbps	chol	Fbs	restecg	thalach	target
52	1	0	125	212	0	1	168	0
53	1	0	140	203	1	0	155	0
34	0	1	118	210	0	1	192	1
51	0	2	140	308	0	0	142	1
58	1	2	140	211	1	0	165	1
60	1	2	140	185	0	0	155	0
67	0	0	106	223	0	1	142	1

Sumber : (UC irvine Machine Learning Repository, 1988)

Gambar 7. Dataset Pasien Penyakit Jantung

Dari gambar di atas dapat dilihat bahwa target dari dataset yang sudah dikumpulkan berdasarkan atribut yang sudah dijelaskan adalah

terdiagnosis penyakit jantung dengan nilai 1, dan tidak terdiagnosis penyakit jantung dengan nilai 0.

2. Arsitektur Model Machine Learning

Terdapat tiga model *machine learning* yang digunakan dalam penelitian ini, yakni *Naïve Bayes*, *Regresi Logistik*, dan KNN .

Pada tahap awal menggunakan algoritma *Naïve Bayes* dengan menggunakan aplikasi *Python*, dimulai dengan melakukan *import* pada *numpy*, *matplotlib*, *pandas*, dll.

1. Metode Naïve Bayes

```
[ ] print(classifier.score(X_test2, y_test2))
0.8341463414634146

[ ] y_true = y_test2
y_pred2 = predicted
print(accuracy_score(y_true, y_pred2))
0.8341463414634146

[ ] print(confusion_matrix(y_test2, y_pred2))
print(classification_report(y_test2, y_pred2))

[[76 21]
 [13 95]]
```

	precision	recall	f1-score	support
0	0.85	0.78	0.82	97
1	0.82	0.88	0.85	108
accuracy			0.83	205
macro avg	0.84	0.83	0.83	205
weighted avg	0.84	0.83	0.83	205

Sumber : Peneliti (2024)

Gambar 8. Metode Naïve Bayes

Dari pengujian data pasien penyakit jantung menggunakan metode *Naïve Bayes*, diperoleh tingkat akurasi sebesar 83% dari total 1025 data.

2. Metode Logistic Regression

```
[ ] from sklearn.naive_bayes import GaussianNB
nb = GaussianNB()
nb.fit(x_train.T, y_train.T)

acc = nb.score(x_test.T, y_test.T)*100
accuracies['Naive Bayes'] = acc
print("Accuracy of Naive Bayes: {:.2f}%".format(acc))

Accuracy of Naive Bayes: 86.89%
```

Sumber : Peneliti (2023)

Gambar 9. Metode Logistic Regression

Dari pengujian yang sudah dilakukan pada data pasien penyakit jantung dengan menggunakan metode *Logistic Regression* diperoleh hasil akurasi sebesar 87% dari 1025 data.

3. Metode KNN (*k-nearest neighbour*)

```
[ ] knn.score(X_test3,y_test3)
0.9073170731707317

[ ] y_pred3 = knn.predict(X_test3)
print(confusion_matrix(y_test3, y_pred3))
print(classification_report(y_test3, y_pred3))
```

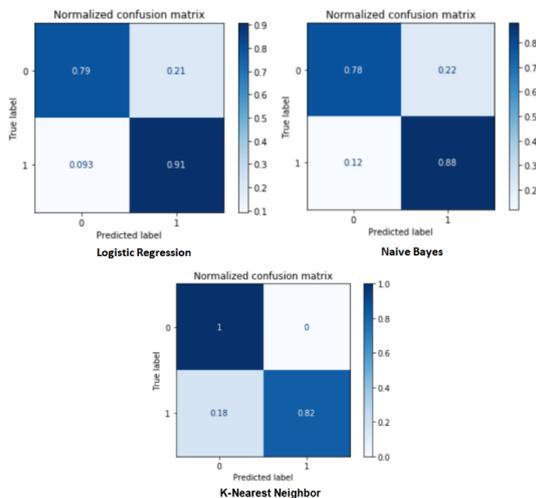
	precision	recall	f1-score	support
0	0.84	1.00	0.91	97
1	1.00	0.82	0.90	108
accuracy			0.91	205
macro avg	0.92	0.91	0.91	205
weighted avg	0.92	0.91	0.91	205

Sumber : Peneliti (2024)

Gambar 10. Metode KNN

Dari pengujian yang sudah dilakukan pada data pasien penyakit jantung dengan menggunakan metode *K-NN (k-nearest neighbour)* diperoleh hasil akurasi sebesar 91% dari 1025 data.

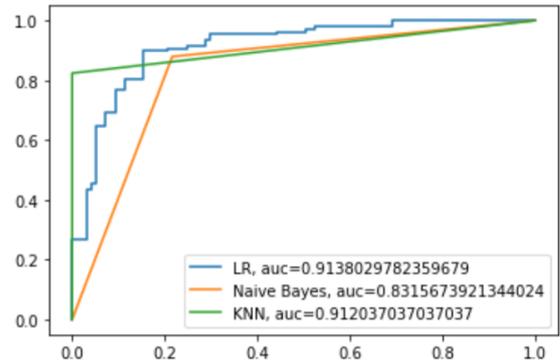
Dari matriks kebingungan yang dihasilkan, terbukti bahwa algoritma KNN memiliki akurasi tertinggi dalam hasil klasifikasi, mencapai 0,90731 dengan nilai K yang sesuai dengan jumlah target, yaitu 2. Algoritma Logistic Regression mendapatkan nilai akurasi sebesar 0,84878, sementara naive bayes mencatat tingkat akurasi terendah, yaitu 0,83414.



Sumber : Peneliti (2024)

Gambar 11. Matrix Model Algoritma

Gambar 11 Tampilan hasil confusion matrix yang telah dinormalisasi untuk setiap algoritma dipaparkan. Evaluasi dimulai dengan menggunakan Area Under Curve (AUC) sebagai metrik untuk mengukur akurasi dari ketiga algoritma yang berbeda. Logistic Regression, KNN, dan Naive Bayes, yang memiliki tingkat yang berbeda. Algoritma Logistic Regression dan KNN menunjukkan peningkatan akurasi, sedangkan Naive Bayes tidak mengalami peningkatan.



Sumber : Peneliti (2024)

Gambar 12. Nilai AUC dari ketiga algoritma

Pada Gambar 12 Nilai AUC dari ketiga algoritma ditampilkan, dengan rentang nilai antara 0 dan 1. Sebuah nilai 1 menunjukkan bahwa model memiliki kemampuan sempurna dalam membedakan kelas, sedangkan nilai 0,5 menunjukkan kinerja yang setara dengan pengacakan kelas.

Dibawah ini merupakan hasil perbandingan dari ketiga model algoritma yang dimanfaatkan. Tabel 1 mencerminkan kinerja yang dihasilkan oleh ketiga algoritma, di mana algoritma K-Nearest Neighbor menunjukkan akurasi tertinggi, mencapai 91%, sementara metode Naive Bayes mencatat akurasi terendah, yaitu 83%.

Tabel 1. Hasil Perbandingan Kinerja Algoritma

Parameter Kinerja	Logistic Regression	K-Nearest Neighbor	Naive Bayes
Accuracy	0.85	0.91	0.83
Precision	0.83	1	0.82
Recall	0.91	0.82	0.88
AUC	0.91	0.91	0.83

Sumber: Peneliti (2024)

KESIMPULAN

Kesimpulan yang diperoleh dari penelitian yang telah dilakukan dengan menggunakan data pasien penyakit jantung dari data *public* adalah algoritma yang tertinggi dengan menggunakan metode *k-nearest neighbour (KNN)* yaitu akurasi sebesar 91%, sedangkan dengan algoritma *logistic regression* akurasinya sebesar 85%, dan *naive bayes* akurasinya sebesar 83%.

Oleh karena itu, kesimpulan yang dapat diambil adalah bahwa algoritma *k-nearest neighbour (KNN)* menghasilkan akurasi tertinggi, mencapai 91%, diikuti oleh Logistic Regression dan Naive Bayes.

REFERENSI

- analytics vidhya. (2022). *No Title*.
- Annisa, R., Studi, P., Informasi, S., Kampus, A., Pontianak, K., Teknologi, F., Bina, U., Informatika, S., Barat, K., & Forest, R. (2019). *ANALISIS KOMPARASI ALGORITMA KLASIFIKASI DATA MINING*. 3(1).
- Bimo, P., Setio, N., Retno, D., Saputro, S., & Winarno, B. (2020). *Klasifikasi dengan Pohon Keputusan Berbasis Algoritme*. 3, 64–71.
- Caballe, S. (2019). *No Title*. Web Page.
- Cnn, D. M., Arsal, M., Agus, B., & Anggraini, D. (2020). *Jurnal Nasional Teknologi dan Sistem Informasi Face Recognition Untuk Akses Pegawai Bank Menggunakan Deep Learning*. 01, 55–63.
- Damuri, A., Riyanto, U., Rusdianto, H., & Aminudin, M. (2021). *Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako*. 8(6), 219–225. <https://doi.org/10.30865/jurikom.v8i6.3655>
- DataIndonesia.id. (2023). *No Title*. <https://dataindonesia.id/kesehatan/detail/kematan-akibat-penyakit-jantung-di-indonesia-terus-meningkat>
- dqlab.id. (2021). *No Title*.
- Informatika, J., & Informasi, S. (2020). *INFORMASI (Jurnal Informatika dan Sistem Informasi) Volume 12 No.1 / Mei/ 2020*. 12(1), 67–80.
- Journal, C., Bianto, M. A., Informatika, M. T., & Yogyakarta, U. A. (2019). *Perancangan Sistem Klasifikasi Penyakit Jantung Menggunakan Naïve Bayes*. 6(1).
- Mining, D. (2020). *No Title*.
- Musu, W., Ibrahim, A., Studi, P., Informatika, T., Makassar, U. D., Studi, P., Informatika, M., & Makassar, U. D. (2021). Pengaruh Komposisi Data Training dan Testing terhadap Akurasi Algoritma C4 . 5. *PROSIDING SEMINAR ILMIAH SISTEM INFORMASI DAN TEKNOLOGI INFORMASI*, X(1), 186–195.
- Prasetyo, R. T., & Ripandi, E. (2019). *Optimasi Klasifikasi Jenis Hutan Menggunakan Deep Learning Berbasis Optimize Selection*. 6(1), 100–106.
- Prasetyo, R., Nawawi, I., & Fauzi, A. (2021). *Komparasi Algoritma Logistic Regression dan Random Forest pada Prediksi Cacat Software*. 06(Siringoringo 2017), 275–281.
- Pratama, R. R. (2020). *Analisis Model Machine Learning Terhadap Pengenalan Aktifitas Manusia*. 19(2), 302–311.
- Puspita, R., & Widodo, A. (2021). *Perbandingan Metode KNN , Decision Tree , dan Naïve Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS*. 5(4), 646–654.
- Putry, N. M., Sari, B. N., Kom, M., Informatika, T., & Karawang, U. S. (2022). *KOMPARASI ALGORITMA KNN DAN NAÏVE BAYES UNTUK KLASIFIKASI DIAGNOSIS PENYAKIT DIABETES MELITUS*. 10(1).
- Reviantika, F., Informatika, J. T., Malang, U. M., Azhar, Y., Informatika, J. T., Malang, U. M., Marthasari, G. I., Informatika, J. T., & Malang, U. M. (2021). *Analisis Klasifikasi SMS Spam Menggunakan Logistic Regression*. 04(03), 155–160.
- Rizki, M., Basuki, S., & Azhar, Y. (2020). *Implementasi Deep Learning Menggunakan Arsitektur Long Short Term Memory Untuk Prediksi Curah Hujan Kota Malang*. 2(3), 331–338.
- Roihan, A., Sunarya, P. A., & Rafika, A. S. (2020). *Pemanfaatan Machine Learning dalam Berbagai Bidang : Review paper*. 5(April), 75–82.
- suprianto, dodit. (2020). *No Title*.
- UC irvine Machine Learning Repostory. (1988). *No Title*.
- Utomo, D. P. (2020). *Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung*. 4(April), 437–444. <https://doi.org/10.30865/mib.v4i2.2080>
- Watratan, A. F., B, A. P., Moeis, D., Informasi, S., & Makassar, S. P. (2020). *JOURNAL OF APPLIED COMPUTER SCIENCE AND TECHNOLOGY (JACOST) Implementasi Algoritma Naïve Bayes Untuk Memprediksi Tingkat Penyebaran Covid-19 Di Indonesia*. 1(1), 7–14.

USULAN
PENELITIAN MANDIRI



**Perbandingan Algoritma *Machine Learning* pada Klasifikasi
Penyakit Jantung**

PENGUSUL

Musriatun Napiah, M.Kom (0318059601)

Sujiliani Heristian, M.Kom (0330109501)

UNIVERSITAS BINA SARANA INFORMATIKA

JANUARI 2024

HALAMAN PENGESAHAN

Judul Penelitian : Perbandingan Algoritma *Machine Learning* pada Klasifikasi Penyakit Jantung

Pengusul

Nama Lengkap : Musriatun Napiah, M.Kom

NIDN : 0318059601

Jabatan Fungsional : Asisten Ahli KUM : 150

Program Studi : Ilmu Komputer

Nomor HP : 085697337523

Alamat surel (e-mail) : musriatun.mph@bsi.ac.id

Anggota Pengusul (1)

Nama Lengkap : Sujiliani Heristian, M.Kom

NIDN : 0330109501

Jabatan Fungsional : Asisten Ahli KUM : 150

Program Studi : Teknologi Komputer

Biaya yang diusulkan : 4.800.000

Jakarta, 10 Januari 2024

Menyetujui,
Ketua LPPM

Pengusul

(Dr. Taufiq Baidawi, M.Kom)

NIP. 200304891

(Musriatun Napiah, M.Kom)

NIP. 202210126

Mengetahui,

Rektor

(Prof. Dr. Ir. Mochamad Wahyudi, M.Kom, MM, M.Pd, IPU, ASEAN Eng)

NIP. 199810339

DAFTAR ISI

	Halaman
Halaman Sampul.....	i
Halaman Pengesahan	ii
Daftar Isi	iii
Ringkasan.....	iv
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	2
1.3. Ruang Lingkup.....	2
1.4. Tujuan Penelitian.....	2
BAB II METODE PENELITIAN.....	3
2.1. Pengumpulan Dataset.....	3
2.2. Pre-processing	3
2.3. Machine Learning	4
BAB III BIAYA PENELITIAN.....	7
BAB IV JADWAL PENELITIAN	8
BAB V LUARAN DAN TARGET CAPAIAN	9
DAFTAR PUSTAKA	10
LAMPIRAN	12

RINGKASAN

Jantung adalah organ utama yang memompa darah keseluruh tubuh bergerak melalui sistem peredaran darah. Menurut WHO, penyakit jantung koroner (PJK) telah menjadi masalah kesehatan yang meningkat pesat, menyebabkan 6,7 juta kematian pada tahun 2017. Banyak alternatif atau cara yang digunakan untuk mencegah dan mendeteksi penyakit jantung, kekurangan pengetahuan sering kali membuat penderita terlambat memeriksa diri ke dokter. Tujuan dari penelitian ini adalah untuk mendeteksi penyakit jantung sejak dini akibat gangguan kardiovaskular sehingga dapat memungkinkan untuk pencegahan yang lebih efektif dan pengelolaan yang baik terhadap kondisi kardiovaskular, dengan menggunakan dataset dari <http://archive.ics.uci.edu/ml/> sebanyak 1026 pasien penyakit jantung. Metode yang digunakan pada penelitian ini adalah *Machine Learning* dengan algoritma *logistic regression*, *naive bayes*, dan *k-nearest neighbour* (KNN) . Hasil yang didapatkan dari penelitian ini yang tertinggi adalah dengan menggunakan metode *k-nearest neighbour* (KNN) yaitu akurasi sebesar 91%, sedangkan dengan algoritma *logistic regression* akurasinya sebesar 87%, dan *naive bayes* akurasinya sebesar 83%.

BAB I

PENDAHULUAN

1.1. Latar Belakang

Jantung adalah salah satu organ yang memegang peranan penting dalam sistem peredaran darah manusia. Penyakit jantung merupakan salah satu penyakit yang paling mematikan di dunia, dan istilah untuk semua jenis gangguan yang mempengaruhi jantung. Penyakit jantung tidak sama dengan penyakit kardiovaskular mengacu pada gangguan pembuluh darah dan jantung, sedangkan penyakit jantung mengacu hanya pada bagian hati (1).

Di Indonesia, penyakit jantung merupakan kondisi yang paling umum terjadi pada wanita dewasa, sesuai dengan informasi tentang kasus penyakit di negara ini melalui Data Indonesia menurut provinsinya, penyakit jantung merupakan masalah serius di Indonesia, dengan jumlah kasus yang tinggi dan angka kematian yang meningkat. Fakta bahwa biaya klaim terbesar dalam program Jaminan Kesehatan Nasional (JKN) disebabkan oleh penyakit jantung menunjukkan dampak yang signifikan dari kondisi ini terhadap sistem kesehatan dan ekonomi negara. Angka 15,5 juta kasus penyakit jantung di Indonesia pada tahun 2022 menunjukkan besarnya beban penyakit ini terhadap populasi (2). Dari informasi tersebut, terlihat bahwa banyak individu yang belum menganggap serius faktor penyebab penyakit ini. Setelah menjalani pemeriksaan kesehatan, dokter menemukan bahwa penyakit ini telah mencapai tingkat stadium yang lanjut (3).

Kurangnya akses informasi/media dalam mencari tahu tentang penyakit jantung sehingga menyebabkan keterlambatan untuk pemeriksaan awal ke dokter, merupakan penyebab banyaknya angka kematian semakin tinggi terhadap penyakit ini. Dengan demikian diagnosa sejak dini sangat penting untuk dilakukan (4), Sehingga diperlukan sistem klasifikasi untuk mendeteksi sejak dini tentang penyakit jantung, agar dapat memberikan informasi tentang penyakit jantung yang di derita oleh seseorang.

Machine Learning (ML) merupakan salah satu penerapan dari Kecerdasan Buatan (AI) yang berfokus pada pengembangan sistem yang mampu belajar secara mandiri tanpa perlu diprogram berulang kali. ML memerlukan data (biasanya disebut

sebagai data pelatihan) sebagai bagian dari proses pembelajaran sebelum menghasilkan output. Menggunakan algoritma *Machine Learning* untuk mendiagnosis penyakit kanker diharapkan dapat menghasilkan prediksi yang akurat dan memungkinkan peningkatan dalam deteksi dini serta perawatan yang lebih efektif bagi pasien (5).

Naïve Bayes merupakan salah satu metode yang dapat digunakan untuk mengklasifikasikan data. Bayesian classification merupakan pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class (6). KNN merupakan algoritma yang efektif untuk data yang besar, tahan terhadap data pelatihan yang noise, dan memiliki performa yang baik (7), sedangkan *Logistic Regression* merupakan klasifikasi linier yang menangani masalah klasifikasi multi kelas dan telah terbukti menghasilkan klasifikasi yang powerful (8).

Penelitian ini akan membandingkan algoritma Naive Bayes, Regresi Logistik, dan KNN dalam klasifikasi penyakit jantung. Diharapkan penelitian ini akan memberikan acuan dan referensi serta memajukan pengetahuan untuk penelitian mendatang.

1.2. Rumusan Masalah

Berdasarkan uraian latar belakang diatas, maka rumusan masalah dalam penelitian ini sebagai berikut:

1. Bagaimana menganalisis perbandingan klasifikasi citra genus panthera dengan pendekatan *deep learning* ?
2. Bagaimana hasil akurasi yang didapatkan dalam model yang berbasis *MobileNet* ?
3. Hasil Akurasi klasifikasi citra genus panthera dapat memberikan kontribusi dalam pengembangan teknologi klasifikasi citra.

1.3. Ruang Lingkup

Dalam penulisan proposal ini penulis membahas tentang analisa perbandingan klasifikasi penyakit jantung. Pendekatan dalam klasifikasi yang dilakukan menggunakan metode *machine learning*.

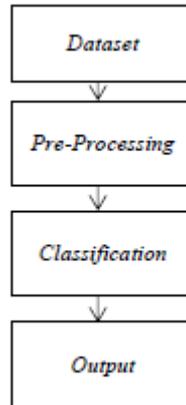
1.4. Tujuan Penelitian

Sejalan dengan perumusan masalah di atas, maka tujuan penelitian ini adalah untuk mendeteksi penyakit jantung sejak dini akibat gangguan kardiovaskular sehingga dapat memungkinkan untuk pencegahan yang lebih efektif dan pengelolaan yang baik terhadap kondisi kardiovaskular.

BAB II

METODE PENELITIAN

Pada Gambar 1 berikut adalah alur metode yang digunakan dalam penelitian perbandingan algoritma *machine learning* pada klasifikasi Penyakit Jantung.



Gambar 1. Alur Metode Penelitian

2.1. Pengumpulan Dataset

Dataset merupakan data mentah berupa tabel yang akan kita olah dengan menggunakan algoritma yang diusulkan. Penelitian ini menggunakan data pasien penyakit jantung diambil dari sumber yang terdapat di <http://archive.ics.uci.edu/ml/> , dengan total 1026 rekaman dan 76 atribut.

2.2. Pre-Processing

Pre-processing adalah proses pengolahan data atau citra asli sebelum data atau citra tersebut diolah oleh algoritma yang diusulkan pada metode penelitian (9). Preprocessing data adalah tahap di mana data yang akan digunakan untuk pelatihan disiapkan terlebih dahulu. Dalam penelitian ini, preprocessing data dilakukan dengan proses untuk mengatasi data yang kosong atau hilang, yang melibatkan metode seperti mencari rata-rata atribut untuk kelas yang serupa (10).

Langkah-langkah pada tahap *pre-processing* meliputi pembagian data menjadi dua bagian, sebagai berikut:

1. Train Data

Data training merupakan himpunan data yang telah diberi label atau kelas tertentu, yang digunakan oleh mesin untuk mengenali fitur-fiturnya serta memahami pola yang ada. Dengan memahami pola ini, mesin dapat membuat model atau pola data yang digunakan untuk tujuan tertentu, seperti klasifikasi atau prediksi (11). Data pelatihan (train data) sebanyak 20% untuk mengevaluasi algoritma.

2. Testing Data

Data testing adalah kumpulan data yang juga memiliki label atau kelas, yang digunakan untuk menguji keakuratan pola atau model yang telah dibuat dalam mengklasifikasikan data. Saat melakukan pengujian model, atribut label dari data testing disembunyikan selama proses klasifikasi, dan kemudian digunakan untuk membandingkan hasil klasifikasi dengan label sebenarnya. Hal ini memungkinkan untuk menilai seberapa baik model tersebut dalam melakukan klasifikasi dengan akurasi yang sesuai dengan data sebenarnya (11). Data pengujian (testing data) sebanyak 80% untuk menentukan metode algoritma yang paling sesuai dalam penelitian.

2.3. Machine Learning

Klasifikasi *Machine learning* adalah penerapan komputer dan algoritma matematika yang menggunakan pembelajaran dari data untuk menghasilkan prediksi pada periode yang akan datang (12). *Machine Learning* berhubungan erat dengan komputasi variabel, fokusnya adalah pada membuat prediksi menggunakan komputer. Studi tentang optimasi matematika menyediakan metode, teori, dan aplikasi di berbagai bidang pembelajaran mesin. Data mining adalah disiplin studi dalam pembelajaran mesin, yang menitikberatkan pada eksplorasi dan analisis data melalui pembelajaran tanpa pengawasan (13).

a) *Naïve Bayes*

Naive Bayes adalah metode klasifikasi yang berasal dari *teorema Bayes*. Metode ini mengandalkan Probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris bernama Thomas Bayes, bertujuan untuk meramalkan kemungkinan di masa mendatang berdasarkan pengalaman yang terjadi di

masa sebelumnya. Karena sifatnya yang mengasumsikan independensi antara fitur-fitur, metode ini disebut sebagai *Naive Bayes* (14).

b) *Logistic Regression*

Logistic Regression merupakan bagian dari metode data mining yang digunakan untuk menganalisis data yang menjelaskan hubungan antara satu variabel *respons (dependen)* dengan satu atau lebih variabel predictor (15).

c) *KNN*

Metode *KNN* adalah algoritma yang dipakai dalam mengklasifikasikan data dengan jarak yang dekat, dan merupakan teknik *lazy learning* yang populer dalam kategori pembelajaran berbasis instansi (16).

BAB III

BIAYA PENELITIAN

Tabel 1 Justifikasi Anggaran Penelitian

Alat dan Bahan				
No	Item Bahan	Volume	Harga Satuan (Rp)	Total (Rp)
1	Buku	10	Rp. 25.000	Rp. 250.000
2	Modem	-	-	Rp. 250.000
Total Alat dan Bahan (Rp)				Rp. 500.000
Bahan Pelatihan				
No	Item Pelatihan	Volume	Harga Satuan (Rp)	Total (Rp)
1	ATK	-	-	Rp. 500.000
2	Voucher Internet	-	-	Rp. 500.000
3	Pulsa	-	-	Rp. 300.000
4	Tinta Printer	4	Rp. 250.000	Rp. 1.000.000
Total Bahan Pelatihan (Rp)				Rp. 2.300.000
Perjalanan dan Konsumsi				
No	Item Barang	Volume	Harga Satuan (Rp)	Total (Rp)
1	Transportasi Perjalanan	-	-	Rp.750.000
2.	Konsumsi	-	-	Rp.750.000
Total Perjalanan dan Konsumsi (Rp)				Rp. 1.500.000
Lain-Lain				
No	Item Barang	Volume	Harga Satuan (Rp)	Total (Rp)
1	Publikasi Ilmiah	-	Rp. 500.000	Rp. 500.000
Total Lain-Lain (Rp)				Rp. 500.000
Total Anggaran Yang Diperlukan (Rp)				Rp. 4.800.000

BAB IV

JADWAL PENELITIAN

Tabel 2. Jadwal Penelitian

No	Nama Kegiatan	Bulan							
		1	2	3	4	5	6	7	8
1	Analisa Data	■	■						
2	Persiapan Data		■	■					
3	Pengambilan Data			■	■				
4	Pengembangan Data				■	■			
5	Pelatihan Data					■	■		
6	Pengujian Data						■	■	
7	Pembuatan laporan						■		
8	Publikasi laporan							■	■

BAB V

LUARAN DAN TARGET CAPAIAN

Dalam penelitian ini luaran dan target capaian penelitian ini berupa publikasi jurnal nasional terakreditasi S4 di Jurnal Information Technology (INFORTECH) Berikut linknya : <https://ejournal.bsi.ac.id/ejurnal/index.php/infortech/article/view/21888> .

DAFTAR PUSTAKA

1. Annisa R. Analisis Komparasi Algoritma Klasifikasi Data Mining Untuk Prediksi Penderita Penyakit Jantung. *Jurnal Teknik Informatika Kaputama (JTIK)*. 2019;3(1):22–8.
2. Mustajab R. DataIndonesia [Internet]. <https://dataindonesia.id/>. 2023. Tersedia pada: <https://dataindonesia.id/kesehatan/detail/kematian-akibat-penyakit-jantung-di-indonesia-terus-meningkat>
3. Journal C, Bianto MA, Informatika MT, Yogyakarta UA. Perancangan Sistem Klasifikasi Penyakit Jantung Menggunakan Naïve Bayes. 2019;6(1).
4. Utomo DP. Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung. 2020;4(April):437–44.
5. Informatika J, Informasi S. INFORMASI (Jurnal Informatika dan Sistem Informasi) Volume 12 No.1 / Mei/ 2020. 2020;12(1):67–80.
6. Damuri A, Riyanto U, Rusdianto H, Aminudin M. Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako. 2021;8(6):219–25.
7. Putry NM, Sari BN, Kom M, Informatika T, Karawang US. KOMPARASI ALGORITMA KNN DAN NAÏVE BAYES UNTUK KLASIFIKASI DIAGNOSIS PENYAKIT DIABETES MELITUS. 2022;10(1).
8. Prasetyo R, Nawawi I, Fauzi A. Komparasi Algoritma Logistic Regression dan Random Forest pada Prediksi Cacat Software. 2021;06(Siringoringo 2017):275–81.
9. Cnn DM, Arsal M, Agus B, Anggraini D. *Jurnal Nasional Teknologi dan Sistem Informasi Face Recognition Untuk Akses Pegawai Bank Menggunakan Deep Learning*. 2020;01:55–63.
10. Rizki M, Basuki S, Azhar Y. Implementasi Deep Learning Menggunakan Arsitektur Long Short Term Memory Untuk Prediksi Curah Hujan Kota Malang. 2020;2(3):331–8.

11. Musu W, Ibrahim A, Studi P, Informatika T, Makassar UD, Studi P, et al. Pengaruh Komposisi Data Training dan Testing terhadap Akurasi Algoritma C4 . 5. PROSIDING SEMINAR ILMIAH SISTEM INFORMASI DAN TEKNOLOGI INFORMASI. 2021;X(1):186–95.
12. Roihan A, Sunarya PA, Rafika AS. Pemanfaatan Machine Learning dalam Berbagai Bidang : Review paper. 2020;5(April):75–82.
13. Pratama RR. Analisis Model Machine Learning Terhadap Pengenalan Aktifitas Manusia. 2020;19(2):302–11.
14. Watratan AF, B AP, Moeis D, Informasi S, Makassar SP. JOURNAL OF APPLIED COMPUTER SCIENCE AND TECHNOLOGY (JACOST) Implementasi Algoritma Naive Bayes Untuk Memprediksi Tingkat Penyebaran Covid-19 Di Indonesia. 2020;1(1):7–14.
15. Reviantika F, Informatika JT, Malang UM, Azhar Y, Informatika JT, Malang UM, et al. Analisis Klasifikasi SMS Spam Menggunakan Logistic Regression. 2021;04(03):155–60.
16. Puspita R, Widodo A. Perbandingan Metode KNN , Decision Tree , dan Naïve Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS. 2021;5(4):646–54.

LAMPIRAN

BIODATA KETUA PENGUSUL DAN ANGGOTA

1. Identitas Diri

- a. Nama Lengkap dan Gelar : Musriatun Napiah, M.Kom
- b. NIDN : 0318059601
- c. Jabatan Fungsional : Asisten Ahli
- d. Program Studi : Ilmu Komputer
- e. Perguruan Tinggi : Universitas Bina Sarana Informatika
- f. Jangka Waktu Penelitian : 8 bulan

2. Riwayat Pendidikan

	S-1	S-2
Nama Penguruan Tinggi	Universitas Nusa Mandiri	Universitas Nusa Mandiri
Tahun Masuk - Lulus	2015-2019	2020-2022

Semua data yang saya isikan dan tercantum dalam biodata ini adalah benar dan dapat dipertanggungjawabkan secara hukum. Apabila di kemudian hari ternyata dijumpai ketidaksesuaian dengan kenyataan, saya sanggup menerima sanksi.

Demikian biodata ini saya buat dengan sebenarnya untuk memenuhi salah satu persyaratan dalam pengajuan Penelitian Mandiri.

Jakarta, 10 Januari 2024

Pengusul



Musriatun Napiah, M.Kom

NIP. 202210126

1. Identitas Diri

- a. Nama Lengkap dan Gelar : Sujiliani Heristian, M.Kom
- b. NIDN : 0330109501
- c. Jabatan Fungsional : Asisten Ahli
- d. Program Studi : Teknologi Komputer
- e. Perguruan Tinggi : Universitas Bina Sarana Informatika
- f. Jangka Waktu Penelitian : 8 bulan

2. Riwayat Pendidikan

	S-1	S-2
Nama Perguruan Tinggi	Universitas Nusa Mandiri	Universitas Nusa Mandiri
Tahun Lulus	2016-2017	2017-2019

Semua data yang saya isikan dan tercantum dalam biodata ini adalah benar dan dapat dipertanggungjawabkan secara hukum. Apabila di kemudian hari ternyata dijumpai ketidaksesuaian dengan kenyataan, saya sanggup menerima sanksi.

Demikian biodata ini saya buat dengan sebenarnya untuk memenuhi salah satu persyaratan dalam pengajuan Penelitian Mandiri.

Jakarta, 10 Januari 2024

Pengusul



(Sujiliani Heristian, M.Kom)

NIP. 201911385

LAPORAN
PENELITIAN MANDIRI



**Perbandingan Algoritma Machine Learning pada Klasifikasi
Penyakit Jantung**

PENELITI

Musriatun Napiah, M.Kom (0318059601)

Sujiliani Heristian, M.Kom (0330109501)

UNIVERSITAS BINA SARANA INFORMATIKA

AGUSTUS 2024

HALAMAN PENGESAHAN

Judul Penelitian : Perbandingan Algoritma Machine Learning pada Klasifikasi Penyakit Jantung

Peneliti

Nama Lengkap : Musriatun Napiah, M.Kom
NIDN : 0318059601
Jabatan Fungsional : Asisten Ahli KUM : 150
Program Studi : Ilmu Komputer
Nomor HP : 085697337523
Alamat surel (e-mail) : musriatun.mph@bsi.ac.id

Anggota Peneliti (1)

Nama Lengkap : Sujiliani Heristian, M.Kom
NIDN : 0330109501
Jabatan Fungsional : Asisten Ahli KUM : 150
Program Studi : Teknologi Komputer
Biaya yang diusulkan : 4.800.000

Jakarta, 10 Agustus 2024

Menyetujui,
Ketua LPPM



(Dr. Taufiq Baidawi, M.Kom)

NIP. 200304891

Peneliti



(Musriatun Napiah, M.Kom)

NIP. 202210126

Mengetahui,

Rektor



(Prof. Dr. Ir. Mochamad Wahyudi, M.Kom, MM, M.Pd, IPU, ASEAN Eng)

NIP. 199810339

DAFTAR ISI

	Halaman
Halaman Sampul.....	i
Halaman Pengesahan	ii
Daftar Isi	iii
Ringkasan.....	iv
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	2
1.3. Ruang Lingkup.....	2
1.4. Tujuan Penelitian.....	2
BAB II TUJUAN DAN MANFAAT PENELITIAN.....	3
2.1. Tujuan Penelitian.....	3
2.2. Manfaat	3
BAB III METODE PENELITIAN	4
3.1. Pengumpulan Dataset.....	4
3.2. Pra-Pemrosesan Dataset	4
3.3. Klasifikasi Dataset	5
3.4. Pengembangan Model.....	6
BAB IV HASIL DAN LUARAN YANG DICAPAI	8
4.1. Hasil	8
4.2. Luaran Yang Dicapai.....	12
BAB V REALISASI BIAYA PENELITIAN	13
BAB VI KESIMPULAN DAN SARAN	14
6.1. Kesimpulan	14
6.2. Saran.....	14
DAFTAR PUSTAKA	15
LAMPIRAN	18

RINGKASAN

Jantung adalah organ utama yang memompa darah keseluruh tubuh bergerak melalui sistem peredaran darah. Menurut WHO, penyakit jantung koroner (PJK) telah menjadi masalah kesehatan yang meningkat pesat, menyebabkan 6,7 juta kematian pada tahun 2017. Banyak alternatif atau cara yang digunakan untuk mencegah dan mendeteksi penyakit jantung, kekurangan pengetahuan sering kali membuat penderita terlambat memeriksa diri ke dokter. Tujuan dari penelitian ini adalah untuk mendeteksi penyakit jantung sejak dini akibat gangguan kardiovaskular sehingga dapat memungkinkan untuk pencegahan yang lebih efektif dan pengelolaan yang baik terhadap kondisi kardiovaskular, dengan menggunakan dataset dari <http://archive.ics.uci.edu/ml/> sebanyak 1026 pasien penyakit jantung. Metode yang digunakan pada penelitian ini adalah Machine Learning dengan algoritma logistic regression, naive bayes, dan k-nearest neighbour (KNN) . Hasil yang didapatkan dari penelitian ini yang tertinggi adalah dengan menggunakan metode k-nearest neighbour (KNN) yaitu akurasi sebesar 91%, sedangkan dengan algoritma logistic regression akurasinya sebesar 87%, dan naive bayes akurasinya sebesar 83%.

BAB I

PENDAHULUAN

1.1. Latar Belakang

Jantung adalah salah satu organ yang memegang peranan penting dalam sistem peredaran darah manusia. Penyakit jantung merupakan salah satu penyakit yang paling mematikan di dunia, dan istilah untuk semua jenis gangguan yang mempengaruhi jantung. Penyakit jantung tidak sama dengan penyakit kardiovaskular mengacu pada gangguan pembuluh darah dan jantung, sedangkan penyakit jantung mengacu hanya pada bagian hati (1).

Di Indonesia, penyakit jantung merupakan kondisi yang paling umum terjadi pada wanita dewasa, sesuai dengan informasi tentang kasus penyakit di negara ini melalui Data Indonesia menurut provinsinya, penyakit jantung merupakan masalah serius di Indonesia, dengan jumlah kasus yang tinggi dan angka kematian yang meningkat. Fakta bahwa biaya klaim terbesar dalam program Jaminan Kesehatan Nasional (JKN) disebabkan oleh penyakit jantung menunjukkan dampak yang signifikan dari kondisi ini terhadap sistem kesehatan dan ekonomi negara. Angka 15,5 juta kasus penyakit jantung di Indonesia pada tahun 2022 menunjukkan besarnya beban penyakit ini terhadap populasi (2). Dari informasi tersebut, terlihat bahwa banyak individu yang belum menganggap serius faktor penyebab penyakit ini. Setelah menjalani pemeriksaan kesehatan, dokter menemukan bahwa penyakit ini telah mencapai tingkat stadium yang lanjut (3).

Kurangnya akses informasi/media dalam mencari tahu tentang penyakit jantung sehingga menyebabkan keterlambatan untuk pemeriksaan awal ke dokter, merupakan penyebab banyaknya angka kematian semakin tinggi terhadap penyakit ini. Dengan demikian diagnosa sejak dini sangat penting untuk dilakukan (4), Sehingga diperlukan sistem klasifikasi untuk mendeteksi sejak dini tentang penyakit jantung, agar dapat memberkan informasi tentang penyakit jantung yang di derita oleh seseorang.

Machine Learning (ML) merupakan salah satu penerapan dari Kecerdasan Buatan (AI) yang berfokus pada pengembangan sistem yang mampu belajar secara

mandiri tanpa perlu diprogram berulang kali. ML memerlukan data (biasanya disebut sebagai data pelatihan) sebagai bagian dari proses pembelajaran sebelum menghasilkan output. Menggunakan algoritma Machine Learning untuk mendiagnosis penyakit kanker diharapkan dapat menghasilkan prediksi yang akurat dan memungkinkan peningkatan dalam deteksi dini serta perawatan yang lebih efektif bagi pasien (5).

Naïve Bayes merupakan salah satu metode yang dapat digunakan untuk mengklasifikasikan data. Bayesian classification merupakan pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class (6). KNN merupakan algoritma yang efektif untuk data yang besar, tahan terhadap data pelatihan yang noise, dan memiliki performa yang baik (7), sedangkan Logistic Regression merupakan klasifikasi linier yang menangani masalah klasifikasi multi kelas dan telah terbukti menghasilkan klasifikasi yang powerful (8).

Penelitian ini akan membandingkan algoritma Naive Bayes, Regresi Logistik, dan KNN dalam klasifikasi penyakit jantung. Diharapkan penelitian ini akan memberikan acuan dan referensi serta memajukan pengetahuan untuk penelitian mendatang.

1.2. Rumusan Masalah

Berdasarkan uraian latar belakang diatas, maka rumusan masalah dalam penelitian ini sebagai berikut:

1. Bagaimana perbandingan akurasi, presisi, recall, dan nilai F1 dari algoritma-algoritma tersebut dalam mendeteksi penyakit jantung?
2. Faktor apa saja yang mempengaruhi kinerja masing-masing algoritma dalam klasifikasi penyakit jantung?
3. Algoritma manakah yang menunjukkan potensi terbaik untuk implementasi dalam sistem pendukung keputusan klinis?

1.3. Ruang Lingkup

Dalam penulisan proposal ini penulis membahas tentang analisa perbandingan klasifikasi penyakit jantung. Pendekatan dalam klasifikasi yang dilakukan menggunakan metode machine learning

1.4. Tujuan Penelitian

Sejalan dengan perumusan masalah di atas, maka tujuan penelitian ini adalah untuk mendeteksi penyakit jantung sejak dini akibat gangguan kardiovaskular sehingga dapat memungkinkan untuk pencegahan yang lebih efektif dan pengelolaan yang baik terhadap kondisi kardiovaskular.

BAB II

TUJUAN DAN MANFAAT PENELITIAN

2.1 Tujuan Penelitian

Sejalan dengan perumusan masalah di atas, maka tujuan penelitian ini adalah untuk apat memberikan kontribusi penting dalam pengembangan teknologi perbandingan mendeteksi penyakit jantung sejak dini.

2.2 Manfaat Penelitian

Penelitian “Perbandingan Algoritma Machine Learning pada Klasifikasi Penyakit Jantung” memiliki beberapa manfaat yang signifikan dalam berbagai konteks. Berikut adalah beberapa manfaat utamanya:

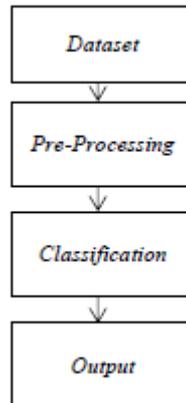
1. Deteksi Dini dan Diagnosa Akurat
2. Pendidikan dan Kesadaran Pasien
3. Identifikasi Faktor Risiko
4. Prediksi Komplikasi
5. Optimalisasi Sumber Daya Medis
6. Peningkatan Keputusan Klinis

Secara keseluruhan, penelitian ini dapat memberikan kontribusi yang luas tidak hanya dalam bidang teknologi dan ilmu komputer tetapi juga dalam konservasi, penelitian biologi, efisiensi operasional, pendidikan, dan kesadaran publik mengenai pentingnya mendeteksi penyakit jantung sejak dini.

BAB III

METODE PENELITIAN

Pada Gambar 1 berikut adalah alur metode yang digunakan dalam penelitian perbandingan algoritma machine learning pada klasifikasi Penyakit Jantung.



Gambar 1. Alur Metode Penelitian

3.1. Pengumpulan Dataset

Dataset merupakan data mentah berupa tabel yang akan kita olah dengan menggunakan algoritma yang diusulkan. Penelitian ini menggunakan data pasien penyakit jantung diambil dari sumber yang terdapat di <http://archive.ics.uci.edu/ml/> , dengan total 1026 rekaman dan 76 atribut.

3.2. Pra-Pemrosesan Dataset

Pre-processing adalah proses pengolahan data atau citra asli sebelum data atau citra tersebut diolah oleh algoritma yang diusulkan pada metode penelitian (9). Preprocessing data adalah tahap di mana data yang akan digunakan untuk pelatihan disiapkan terlebih dahulu. Dalam penelitian ini, preprocessing data dilakukan dengan proses untuk mengatasi data yang kosong atau hilang, yang melibatkan metode seperti mencari rata-rata atribut untuk kelas yang serupa (10).

Langkah-langkah pada tahap pre-processing meliputi pembagian data menjadi dua bagian, sebagai berikut:

1. Train Data

Data training merupakan himpunan data yang telah diberi label atau kelas tertentu, yang digunakan oleh mesin untuk mengenali fitur-fiturnya serta memahami pola yang ada. Dengan memahami pola ini, mesin dapat membuat model atau pola data yang digunakan untuk tujuan tertentu, seperti klasifikasi atau prediksi (11). Data pelatihan (train data) sebanyak 20% untuk mengevaluasi algoritma.

2. Testing Data

Data testing adalah kumpulan data yang juga memiliki label atau kelas, yang digunakan untuk menguji keakuratan pola atau model yang telah dibuat dalam mengklasifikasikan data. Saat melakukan pengujian model, atribut label dari data testing disembunyikan selama proses klasifikasi, dan kemudian digunakan untuk membandingkan hasil klasifikasi dengan label sebenarnya. Hal ini memungkinkan untuk menilai seberapa baik model tersebut dalam melakukan klasifikasi dengan akurasi yang sesuai dengan data sebenarnya (11). Data pengujian (testing data) sebanyak 80% untuk menentukan metode algoritma yang paling sesuai dalam penelitian.

3.3. Machine Learning

Klasifikasi Machine learning adalah penerapan komputer dan algoritma matematika yang menggunakan pembelajaran dari data untuk menghasilkan prediksi pada periode yang akan datang (12). Machine Learning berhubungan erat dengan komputasi variabel, fokusnya adalah pada membuat prediksi menggunakan komputer. Studi tentang optimasi matematika menyediakan metode, teori, dan aplikasi di berbagai bidang pembelajaran mesin. Data mining adalah disiplin studi dalam pembelajaran mesin, yang menitikberatkan pada eksplorasi dan analisis data melalui pembelajaran tanpa pengawasan (13).

a) Naïve Bayes

Naive Bayes adalah metode klasifikasi yang berasal dari teorema Bayes. Metode ini mengandalkan Probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris bernama Thomas Bayes, bertujuan untuk meramalkan kemungkinan di masa mendatang

berdasarkan pengalaman yang terjadi di masa sebelumnya. Karena sifatnya yang mengasumsikan independensi antara fitur-fitur, metode ini disebut sebagai Naive Bayes (14).

b) Logistic Regression

Logistic Regression merupakan bagian dari metode data mining yang digunakan untuk menganalisis data yang menjelaskan hubungan antara satu variabel respons (dependen) dengan satu atau lebih variabel predictor (15).

c) KNN

Metode KNN adalah algoritma yang dipakai dalam mengklasifikasikan data dengan jarak yang dekat, dan merupakan teknik lazy learning yang populer dalam kategori pembelajaran berbasis instansi (16).

3.4. Pengembangan Model

Dalam penelitian ini evaluasi model dilakukan dengan menggunakan 2 alat ukur yaitu:

a) Confusion matrix

Confusion matrix merupakan metrik yang digunakan untuk menggambarkan kinerja dan mengevaluasi keefektifan suatu model, lalu menggambarkan perbandingan antara hasil prediksi dan keadaan aktual dari data yang dianalisis oleh algoritma *machine learning*, terutama pada model klasifikasi (17). Melalui *confusion matrix*, kita dapat menghitung nilai-nilai seperti *accuracy*, *precision*, *recall*, dan *F1-Score*.

b) Kurva ROC dan AUC

Kurva Receiver Operating Characteristic (ROC) menampilkan informasi kinerja algoritma klasifikasi dalam bentuk grafik dapat digunakan atau *Precision-Recall Curve*. Kurva ROC dibuat berdasarkan nilai yang telah didapatkan dari perhitungan dengan *confusion matrix*, yaitu antara *False Positive Rate* dengan *True Positive Rate*. Untuk membandingkan nilai kinerja masing-masing algoritma dapat dilakukan dengan membandingkan luas di bawah kurva atau *Area Under Curve (AUC)* (18).

BAB IV

HASIL DAN LUARAN YANG DICAPAI

6.1. Hasil

Dataset yang diterapkan dalam penelitian ini adalah Ada 1025 data pasien dengan penyakit jantung, di mana 80% atau 820 data telah dipilih sebagai data pelatihan, sementara 20% sisanya digunakan sebagai data uji. Atribut-atribut yang dimasukkan meliputi usia (dalam tahun), jenis kelamin, cp, trestbps (tekanan darah istirahat pada saat masuk rumah sakit), chol (kolesterol serum), fbs (gula darah puasa > 120 mg/dl), restecg, dan thalach (denyut jantung maksimum yang dicapai), dan target (diagnosis of heart disease). Dataset dapat dilihat pada tabel 1.

Tabel 1. Dataset Pasien Penyakit Jantung

age	sex	trestbps	chol	Fbs	restecg	thalach	target
52	1	125	212	0	1	168	0
53	1	140	203	1	0	155	0
34	0	118	210	0	1	192	1
51	0	140	308	0	0	142	1
58	1	140	211	1	0	165	1
60	1	140	185	0	0	155	0
67	0	106	223	0	1	142	1

Dari tabel di atas dapat dilihat bahwa target dari dataset yang sudah dikumpulkan berdasarkan atribut yang sudah dijelaskan adalah terdiagnosis penyakit jantung dengan nilai 1, dan tidak terdiagnosis penyakit jantung dengan nilai 0.

```
[ ] print(classifier.score(X_test2, y_test2))
0.8341463414634146

[ ] y_true = y_test2
y_pred2 = predicted

print(accuracy_score(y_true, y_pred2))
0.8341463414634146

[ ] print(confusion_matrix(y_test2, y_pred2))
print(classification_report(y_test2, y_pred2))
```

	[[76 21]				
	[13 95]]				
		precision	recall	f1-score	support
0		0.85	0.78	0.82	97
1		0.82	0.88	0.85	108
				0.83	205
accuracy				0.83	205
macro avg		0.84	0.83	0.83	205
weighted avg		0.84	0.83	0.83	205

Gambar 1. Metode Naïve Bayes

Dari pengujian data pasien penyakit jantung menggunakan metode Naïve Bayes, diperoleh tingkat akurasi sebesar 83% dari total 1025 data.

```
[ ] from sklearn.naive_bayes import GaussianNB
nb = GaussianNB()
nb.fit(x_train.T, y_train.T)

acc = nb.score(x_test.T, y_test.T)*100
accuracies['Naive Bayes'] = acc
print("Accuracy of Naive Bayes: {:.2f}%".format(acc))

Accuracy of Naive Bayes: 86.89%
```

Gambar 2. Metode Logistic Regression

Dari pengujian yang sudah dilakukan pada data pasien penyakit jantung dengan menggunakan metode Logistic Regression diperoleh hasil akurasi sebesar 87% dari 1025 data.

```
[ ] knn.score(X_test3,y_test3)

0.9073170731707317

[ ] y_pred3 = knn.predict(X_test3)
print(confusion_matrix(y_test3, y_pred3))
print(classification_report(y_test3, y_pred3))

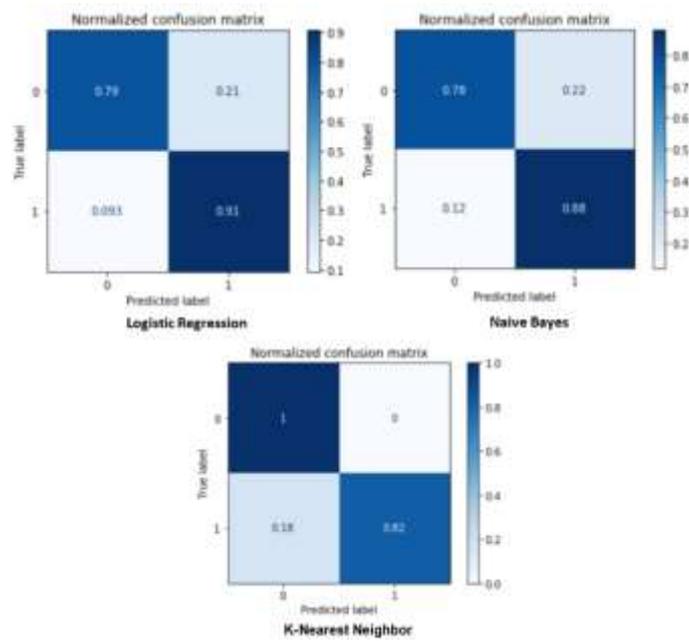
[[97  0]
 [19 89]]
```

	precision	recall	f1-score	support
0	0.84	1.00	0.91	97
1	1.00	0.82	0.90	108
accuracy			0.91	205
macro avg	0.92	0.91	0.91	205
weighted avg	0.92	0.91	0.91	205

Gambar 3. Metode KNN

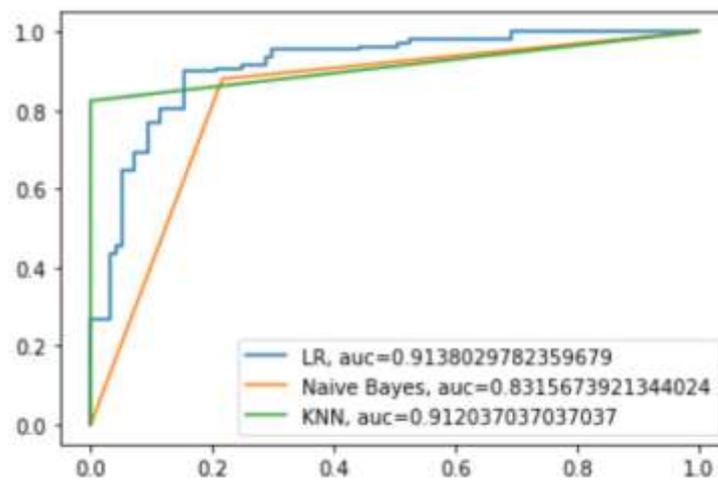
Dari pengujian yang sudah dilakukan pada data pasien penyakit jantung dengan menggunakan metode K-NN (k-nearest neighbour) diperoleh hasil akurasi sebesar 91% dari 1025 data.

Dari matriks kebingungan yang dihasilkan, terbukti bahwa algoritma KNN memiliki akurasi tertinggi dalam hasil klasifikasi, mencapai 0,90731 dengan nilai K yang sesuai dengan jumlah target, yaitu 2. Algoritma Logistic Regression mendapatkan nilai akurasi sebesar 0,84878, sementara naive bayes mencatat tingkat akurasi terendah, yaitu 0,83414.



Gambar 4. Matrix Model Algoritma

Gambar 4 Tampilan hasil confusion matrix yang telah dinormalisasi untuk setiap algoritma dipaparkan. Evaluasi dimulai dengan menggunakan Area Under Curve (AUC) sebagai metrik untuk mengukur akurasi dari ketiga algoritma yang berbeda. Logistic Regression, KNN, dan Naïve Bayes, yang memiliki tingkat yang berbeda. Algoritma Logistic Regression dan KNN menunjukkan peningkatan akurasi, sedangkan Naïve Bayes tidak mengalami peningkatan.



Gambar 5. Nilai AUC dari ketiga algoritma

Tabel 2. Pembagian Data Citra

No.	Nama Citra	Jumlah
1.	Data Training	5.168 Citra
2.	Data Validation	646 Citra
3.	Data testing	646 Citra

Hasil klasifikasi data citra *Genus Panthera* adalah penentuan atau pengelompokan citra-citra berdasarkan jenis spesies kucing besar dalam *Genus Panthera*, seperti *Jaguar*, *Leopard*, *Lion*, *Liones*, *Tiger*, *Snow Tiger* dan menggunakan algoritma atau model yang telah dilatih yaitu *MobileNet*, dengan tujuan untuk mengidentifikasi spesies dengan akurasi yang tinggi.

Type / Stride	Filter Shape	Input Size
Conv / s2	3 x 3 x 3 x 32	224 x 224 x 3
Conv dw / s1	3 x 3 x 32 dw	112 x 112 x 32
Conv / s1	1 x 1 x 32 x 64	112 x 112 x 32
Conv dw / s2	3 x 3 x 64 dw	112 x 112 x 64
Conv / s1	1 x 1 x 64 x 128	56 x 56 x 64
Conv dw / s1	3 x 3 x 128 dw	56 x 56 x 128
Conv / s1	1 x 1 x 128 x 128	56 x 56 x 128
Conv dw / s2	3 x 3 x 128 dw	56 x 56 x 128
Conv / s1	1 x 1 x 128 x 256	28 x 28 x 128
Conv dw / s1	3 x 3 x 256 dw	28 x 28 x 256
Conv / s1	1 x 1 x 256 x 256	28 x 28 x 256
Conv dw / s2	3 x 3 x 256 dw	28 x 28 x 256
Conv / s1	1 x 1 x 256 x 512	14 x 14 x 256
5 x Conv dw / s1	3 x 3 x 512 dw	14 x 14 x 512
Conv / s1	1 x 1 x 512 x 512	14 x 14 x 512
Conv dw / s2	3 x 3 x 512 dw	14 x 14 x 512
Conv / s1	1 x 1 x 512 x 1024	7 x 7 x 512
Conv dw / s2	3 x 3 x 1024 dw	7 x 7 x 1024
Conv / s1	1 x 1 x 1024 x 1024	7 x 7 x 1024
Avg Pool / s1	Pool 7 x 7	7 x 7 x 1024
FC / s1	1024 x 1000	1 x 1 x 1024
Softmax / s1	Classifier	1 x 1 x 100

Gambar 2. Arsitektur *MobileNetV1*

Pada gambar 2. merupakan model arsitektur *MobileNetV1* yang diimplementasikan, memiliki lapisan *Depthwise Convolution* yang merupakan lapisan kedalaman dengan jumlah langkah.

Input	Operator	f	c	n	s
224 ² x 3	Conv2d	-	32	1	2
112 ² x 32	bottleneck	1	16	1	1
112 ² x 16	bottleneck	6	24	2	2
56 ² x 24	bottleneck	6	32	3	2
28 ² x 32	bottleneck	6	64	4	2
14 ² x 64	bottleneck	6	96	3	1
14 ² x 96	bottleneck	6	160	3	2
7 ² x 160	bottleneck	6	320	1	1
7 ² x 320	Conv2d 1 x 1	-	1280	1	1
7 ² x 1280	Avgpool 7 x 7	-	-	1	-
1 x 1 x 1280	Conv2d 1 x 1	-	k	-	-

Gambar 3. Arsitektur *MobileNetV2*

Dalam Gambar 3 arsitektur model *MobileNetV2* yang diimplementasikan, dapat diketahui secara keseluruhan kumpulan blok konvolusional yang sangat sederhana dan umum.

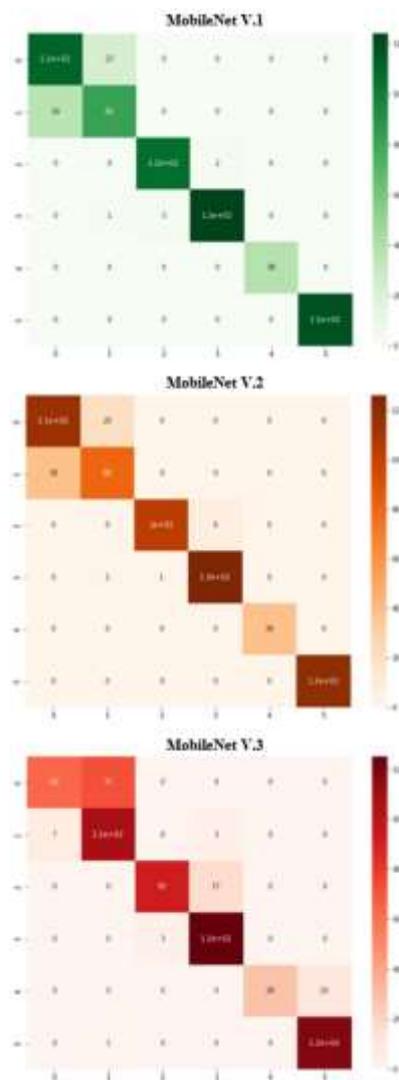
- a) t berarti tingkat ekspansi saluran yang menggunakan faktor 6 dan bukan 4.
- b) c mewakili jumlah saluran masukan.
- c) n seberapa sering blok tersebut diulang.
- d) s memberi tahu kita apakah pengulangan pertama sebuah blok menggunakan langkah 2 untuk proses *downsampling*

Dan arsitektur *MobileNetV3* yang diimplementasikan mendefinisikan model *small* dapat dilihat pada Gambar 4, berikut penjelasannya.

Input	Operator	exp size	#out	SE	NL	s
224 ² x 3	Conv2d, 3x3	-	16	-	HS	2
112 ² x 16	bneck, 3x3	16	16	✓	RE	2
56 ² x 16	bneck, 3x3	72	24	-	RE	2
28 ² x 24	bneck, 3x3	88	24	-	RE	1
28 ² x 24	bneck, 5x5	96	40	✓	HS	2
14 ² x 40	bneck, 5x5	240	40	✓	HS	1
14 ² x 40	bneck, 5x5	240	40	✓	HS	1
14 ² x 40	bneck, 5x5	120	48	✓	HS	1
14 ² x 48	bneck, 5x5	144	48	✓	HS	1
14 ² x 48	bneck, 5x5	288	96	✓	HS	2
7 ² x 96	bneck, 5x5	576	96	✓	HS	1
7 ² x 96	bneck, 5x5	576	96	✓	HS	1
7 ² x 96	Conv2d, 1x1	-	576	✓	HS	1
7 ² x 576	Pool, 7x7	-	-	-	-	1
1 ² x 576	Conv2d, 1x1, NBN	-	1024	-	HS	1
1 ² x 1024	Conv2d, 1x1, NBN	-	k	-	-	1

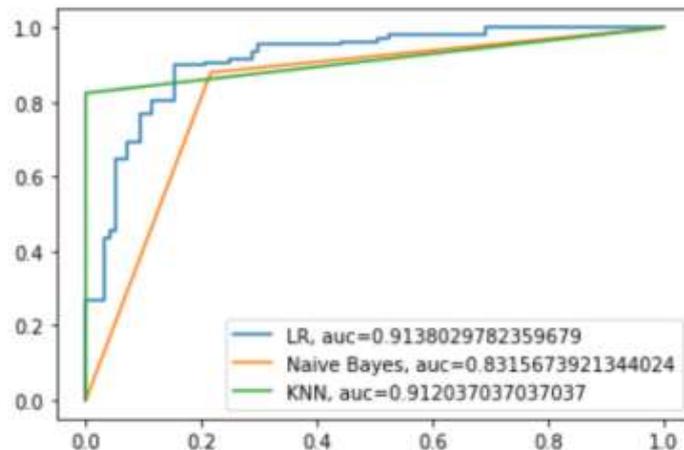
Gambar. 4 Arsitektur *MobileNetV3-small*

Hasil dari evaluasi klasifikasi citra berdasarkan *confusion matrix* yang kemudian dianalisis untuk menentukan keberhasilan model dalam melakukan klasifikasi citra *Genus Panthera*. Berdasarkan *confusion matrix* dari tiap model yang dapat terlihat pada Gambar 5.



Gambar. 5 Confusion Matrix

Gambar 5 Tampilan hasil confusion matrix yang telah dinormalisasi untuk setiap algoritma dipaparkan. Evaluasi dimulai dengan menggunakan Area Under Curve (AUC) sebagai metrik untuk mengukur akurasi dari ketiga algoritma yang berbeda. Logistic Regression, KNN, dan Naïve Bayes, yang memiliki tingkat yang berbeda. Algoritma Logistic Regression dan KNN menunjukkan peningkatan akurasi, sedangkan Naïve Bayes tidak mengalami peningkatan.



Gambar 6. Nilai AUC dari ketiga algoritma

Pada Gambar 6 Nilai AUC dari ketiga algoritma ditampilkan, dengan rentang nilai antara 0 dan 1. Sebuah nilai 1 menunjukkan bahwa model memiliki kemampuan sempurna dalam membedakan kelas, sedangkan nilai 0,5 menunjukkan kinerja yang setara dengan pengacakan kelas.

Tabel 2. Hasil Evaluasi Setiap Model

Paramater Kinerja	Logistic Regression	K-Nearest Neighbor	Naïve Bayes
<i>Accuracy</i>	0.85	0.91	0.83
<i>Precision</i>	0.83	1	0.82
<i>Recall</i>	0.91	0.82	0.88
<i>AUC</i>	0.91	0.91	0.83

Tabel diatas merupakan hasil perbandingan dari ketiga model algoritma yang dimanfaatkan. Tabel 2 mencerminkan kinerja yang dihasilkan oleh ketiga algoritma, di mana algoritma K-Nearest Neighbor menunjukkan akurasi tertinggi, mencapai 91%, sementara metode Naïve Bayes mencatat akurasi terendah, yaitu 83%.

6.2. Luaran Yang Dicapai

Dalam penelitian ini luaran dan target capaian penelitian ini berupa publikasi jurnal nasional terakreditasi S4 di Jurnal Infortech, e-ISSN 2715-8160. Vol.6, No. 1, Juni 2024,

Hal. 46-51. Berikut linknya :

<https://ejournal.bsi.ac.id/ejurnal/index.php/infortech/article/view/21888/6466>

BAB V

REALISASI BIAYA PENELITIAN

Tabel 1 Justifikasi Anggaran Penelitian

Alat dan Bahan				
No	Item Bahan	Volume	Harga Satuan (Rp)	Total (Rp)
1	Buku	10	Rp. 25.000	Rp. 250.000
2	Modem	-	-	Rp. 250.000
Total Alat dan Bahan (Rp)				Rp. 500.000
Bahan Pelatihan				
No	Item Pelatihan	Volume	Harga Satuan (Rp)	Total (Rp)
1	ATK	-	-	Rp. 500.000
2	Voucher Internet	-	-	Rp. 500.000
3	Pulsa	-	-	Rp. 300.000
4	Tinta Printer	4	Rp. 250.000	Rp. 1.000.000
Total Bahan Pelatihan (Rp)				Rp. 2.300.000
Perjalanan dan Konsumsi				
No	Item Barang	Volume	Harga Satuan (Rp)	Total (Rp)
1	Transportasi Perjalanan	-	-	Rp.750.000
2.	Konsumsi	-	-	Rp.750.000
Total Perjalanan dan Konsumsi (Rp)				Rp. 1.500.000
Lain-Lain				
No	Item Barang	Volume	Harga Satuan (Rp)	Total (Rp)
1	Publikasi Ilmiah	-	Rp. 500.000	Rp. 500.000
Total Lain-Lain (Rp)				Rp. 500.000
Total Keseluruhan (Rp)				Rp. 4.800.000

BAB VI

KESIMPULAN DAN SARAN

6.1. Kesimpulan

Kesimpulan yang diperoleh dari penelitian yang telah dilakukan dengan menggunakan data pasien penyakit jantung dari data public adalah algoritma yang tertinggi dengan menggunakan metode k-nearest neighbour (KNN) yaitu akurasi sebesar 91%, sedangkan dengan algoritma logistic regression akurasinya sebesar 85%, dan naive bayes akurasinya sebesar 83%.

Oleh karena itu, kesimpulan yang dapat diambil adalah bahwa algoritma k-nearest neighbour (KNN) menghasilkan akurasi tertinggi, mencapai 91%, diikuti oleh Logistic Regression dan Naive Bayes.

6.2. Saran

Pada penelitian selanjutnya diharapkan dapat memilih algoritma machine learning harus disesuaikan dengan jenis dan kompleksitas data yang tersedia. Algoritma yang lebih kompleks harus dipertimbangkan ketika akurasi yang tinggi menjadi prioritas dan mempertimbangkan penggunaan teknik ensemble seperti bagging atau boosting untuk meningkatkan performa model secara keseluruhan. Kombinasi beberapa algoritma dapat menghasilkan model yang lebih kuat dan robust.

DAFTAR PUSTAKA

1. Annisa R. Analisis Komparasi Algoritma Klasifikasi Data Mining Untuk Prediksi Penderita Penyakit Jantung. *Jurnal Teknik Informatika Kaputama (JTIK)*. 2019;3(1):22–8.
2. Mustajab R. DataIndonesia [Internet]. <https://dataindonesia.id/>. 2023. Tersedia pada: <https://dataindonesia.id/kesehatan/detail/kematian-akibat-penyakit-jantung-di-indonesia-terus-meningkat>
3. Journal C, Bianto MA, Informatika MT, Yogyakarta UA. Perancangan Sistem Klasifikasi Penyakit Jantung Menggunakan Naïve Bayes. 2019;6(1).
4. Utomo DP. Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung. 2020;4(April):437–44.
5. Informatika J, Informasi S. INFORMASI (Jurnal Informatika dan Sistem Informasi) Volume 12 No.1 / Mei/ 2020. 2020;12(1):67–80.
6. Damuri A, Riyanto U, Rusdianto H, Aminudin M. Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako. 2021;8(6):219–25.
7. Putry NM, Sari BN, Kom M, Informatika T, Karawang US. KOMPARASI ALGORITMA KNN DAN NAÏVE BAYES UNTUK KLASIFIKASI DIAGNOSIS PENYAKIT DIABETES MELITUS. 2022;10(1).
8. Prasetyo R, Nawawi I, Fauzi A. Komparasi Algoritma Logistic Regression dan Random Forest pada Prediksi Cacat Software. 2021;06(Siringoringo 2017):275–81.
9. Cnn DM, Arsal M, Agus B, Anggraini D. *Jurnal Nasional Teknologi dan Sistem Informasi Face Recognition Untuk Akses Pegawai Bank Menggunakan Deep Learning*. 2020;01:55–63.
10. Rizki M, Basuki S, Azhar Y. Implementasi Deep Learning Menggunakan

- Arsitektur Long Short Term Memory Untuk Prediksi Curah Hujan Kota Malang. 2020;2(3):331–8.
11. Musu W, Ibrahim A, Studi P, Informatika T, Makassar UD, Studi P, et al. Pengaruh Komposisi Data Training dan Testing terhadap Akurasi Algoritma C4 . 5. PROSIDING SEMINAR ILMIAH SISTEM INFORMASI DAN TEKNOLOGI INFORMASI. 2021;X(1):186–95.
 12. Roihan A, Sunarya PA, Rafika AS. Pemanfaatan Machine Learning dalam Berbagai Bidang : Review paper. 2020;5(April):75–82.
 13. Pratama RR. Analisis Model Machine Learning Terhadap Pengenalan Aktifitas Manusia. 2020;19(2):302–11.
 14. Watratan AF, B AP, Moeis D, Informasi S, Makassar SP. JOURNAL OF APPLIED COMPUTER SCIENCE AND TECHNOLOGY (JACOST) Implementasi Algoritma Naive Bayes Untuk Memprediksi Tingkat Penyebaran Covid-19 Di Indonesia. 2020;1(1):7–14.
 15. Reviantika F, Informatika JT, Malang UM, Azhar Y, Informatika JT, Malang UM, et al. Analisis Klasifikasi SMS Spam Menggunakan Logistic Regression. 2021;04(03):155–60.
 16. Puspita R, Widodo A. Perbandingan Metode KNN , Decision Tree , dan Naïve Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS. 2021;5(4):646–54.
 17. Sari L, Romadloni A, Listyaningrum R. Penerapan Data Mining dalam Analisis Prediksi Kanker Paru Menggunakan Algoritma Random Forest. Infotekmesin. 2023;14(1):155–62.

LAMPIRAN

BIODATA KETUA PENELITI DAN ANGGOTA

1. Identitas Diri

- a. Nama Lengkap dan Gelar : Musriatun Napiah, M.Kom
- b. NIDN : 0318059601
- c. Jabatan Fungsional : Asisten Ahli
- d. Program Studi : Ilmu Komputer
- e. Perguruan Tinggi : Universitas Bina Sarana Informatika
- f. Jangka Waktu Penelitian : 8 bulan

2. Riwayat Pendidikan

	S-1	S-2
Nama Penguruan Tinggi	Universitas Nusa Mandiri	Universitas Nusa Mandiri
Tahun Masuk - Lulus	2015-2019	2020-2022

Semua data yang saya isikan dan tercantum dalam biodata ini adalah benar dan dapat dipertanggungjawabkan secara hukum. Apabila di kemudian hari ternyata dijumpai ketidaksesuaian dengan kenyataan, saya sanggup menerima sanksi.

Demikian biodata ini saya buat dengan sebenarnya untuk memenuhi salah satu persyaratan dalam pengajuan Penelitian Mandiri.

Jakarta, 10 Agustus 2024

Peneliti



Musriatun Napiah, M.Kom

NIP. 202210126

1. Identitas Diri

- a. Nama Lengkap dan Gelar : Sujiliani Heristian, M.Kom
- b. NIDN : 0330109501
- c. Jabatan Fungsional : Asisten Ahli
- d. Program Studi : Teknologi Komputer
- e. Perguruan Tinggi : Universitas Bina Sarana Informatika
- f. Jangka Waktu Penelitian : 8 bulan

2. Riwayat Pendidikan

	S-1	S-2
Nama Perguruan Tinggi	Universitas Nusa Mandiri	Universitas Nusa Mandiri
Tahun Lulus	2016-2017	2017-2019

Semua data yang saya isikan dan tercantum dalam biodata ini adalah benar dan dapat dipertanggungjawabkan secara hukum. Apabila di kemudian hari ternyata dijumpai ketidaksesuaian dengan kenyataan, saya sanggup menerima sanksi.

Demikian biodata ini saya buat dengan sebenarnya untuk memenuhi salah satu persyaratan dalam pengajuan Penelitian Mandiri.

Jakarta, 10 Agustus 2024

Pengusul

(Sujiliani Heristian, M.Kom)

NIP. 201911385

SURAT TUGAS
1642/B.01/LPPM-UBSI/V/2024

Tentang

PENELITIAN YANG DIPUBLIKASIKAN DALAM JURNAL ILMIAH
Periode Maret - Agustus 2024

Menulis Pada Jurnal Infortech Volume 6 No. 1 Juni 2024 (E-ISSN : 2715-8160)

Judul :

Perbandingan Algoritma Machine Learning pada Klasifikasi Penyakit Jantung

- Menimbang :
1. Bahwa perlu diadakan pelaksanaan Tridharma Perguruan Tinggi dalam bentuk Penelitian.
 2. untuk Keperluan pada butir 1 (satu) diatas, maka perlu dibentuk tugas yang berkaitan dengan penelitian yang dipublikasikan dalam Jurnal Ilmiah.

MEMUTUSKAN

- Pertama : Menugaskan kepada saudara
Musriatun Napiah M.Kom
Sebagai Penulis yang mempublikasikan Penelitiannya pada Jurnal Ilmiah.
- Kedua : Mempunyai tugas sbb:
Melaksanakan Tugas yang diberikan dengan penuh rasa tanggung jawab.
- Ketiga : Keputusan ini berlaku sejak tanggal ditetapkan, dengan ketentuan apabila dikemudian hari terdapat kekeliruan akan diubah dan diperbaiki sebagaimana mestinya

Jakarta, 2 Mei 2024

LPPM Universitas Bina Sarana Informatika

Ketua




Dr. Taufik Baidawi, M.kom

Tembusan

- Rektor Universitas Bina Sarana Informatika
- Arsip
- Ybs

SURAT TUGAS
1732/B.01/LPPM-UBSI/V/2024

Tentang

PENELITIAN YANG DIPUBLIKASIKAN DALAM JURNAL ILMIAH
Periode Maret - Agustus 2024

Menulis Pada Jurnal Infortech Volume 6 No. 1 Juni 2024 (E-ISSN : 2715-8160)

Judul :

Perbandingan Algoritma Machine Learning pada Klasifikasi Penyakit Jantung

- Menimbang :
1. Bahwa perlu diadakan pelaksanaan Tridharma Perguruan Tinggi dalam bentuk Penelitian.
 2. untuk Keperluan pada butir 1 (satu) diatas, maka perlu dibentuk tugas yang berkaitan dengan penelitian yang dipublikasikan dalam Jurnal Ilmiah.

MEMUTUSKAN

- Pertama : Menugaskan kepada saudara
Sujiliani Heristian M.Kom
Sebagai Penulis yang mempublikasikan Penelitiannya pada Jurnal Ilmiah.
- Kedua : Mempunyai tugas sbb:
Melaksanakan Tugas yang diberikan dengan penuh rasa tanggung jawab.
- Ketiga : Keputusan ini berlaku sejak tanggal ditetapkan, dengan ketentuan apabila dikemudian hari terdapat kekeliruan akan diubah dan diperbaiki sebagaimana mestinya

Jakarta, 2 Mei 2024

LPPM Universitas Bina Sarana Informatika

Ketua




Dr. Taufik Baidawi, M.kom

Tembusan

- Rektor Universitas Bina Sarana Informatika
- Arsip
- Ybs