

Implementasi Algoritma *Naïve Bayes* dan *Support Vector Machine (SVM)* Pada Klasifikasi Penyakit Kardiovaskular

Anita Desiani, Muhammad Akbar*, Irmeilyana, Ali Amran

Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Sriwijaya
Jl. Raya Palembang – Prabumulih Km.32 Indralaya Indah, Kec. Indralaya, Kabupaten Ogan Ilir Sumatera Selatan
E-mail: muhammadakbar1771@gmail.com

Naskah Masuk: 26 Mei 2022; Diterima: 26 Juli 2022; Terbit: 18 Agustus 2022

ABSTRAK

Abstrak - Penyakit kardiovaskular adalah penyakit yang diakibatkan penyempitan atau penyumbatan pembuluh darah di jantung penyakit ini disebabkan gangguan fungsi jantung dan pembuluh darah. Sistem kardiovaskular terdiri dari jantung dan pembuluh darahnya. Penelitian ini bertujuan melakukan klasifikasi penyakit kardiovaskular untuk memprediksi suatu pola. Pada penelitian ini akan menggunakan metode *support vector machine* dan *naïve bayes* dengan metode latihan *percentage split* dan *k-fold cross validation*. Hasil akurasi pengolahan menggunakan Algoritma *Naïve Bayes* adalah sebesar 70% untuk metode latihan *percentage split* dan 71% untuk metode latihan *k-fold cross validation*. Kemudian dengan menggunakan algoritma *support vector machine* didapat akurasi 61% untuk metode latihan *percentage split* dan 65% untuk metode latihan *k-fold validation*. Hasil tersebut menunjukkan bahwa algoritma *naïve bayes* dengan metode latihan *k-fold validation* cukup baik dalam melakukan klasifikasi penyakit kardiovaskular.

Kata kunci: Klasifikasi, *K-Nearest Neighbor*, *Naïve Bayes*, Kardiovaskular

ABSTRACT

Abstract – cardiovascular disease is a disease that is experienced or disease of the blood in the heart, this disease is caused by impaired function of the heart and blood vessels. The cardiovascular system consists of the heart and its blood vessels. This study aims to classify cardiovascular disease to predict a pattern. In this study, the support vector machine and naïve Bayes methods will be used with the split percentage training method and k-fold cross validation. The results of processing using the Naïve Bayes algorithm are 70% for the split percentage training method and 71% for the k-fold cross validation training method. Then by using the support vector machine algorithm, the accuracy is 61% for the split percentage training method and 65% for the k-fold training validation method. These results indicate that the Naïve Bayes algorithm with the k-fold validation training method is quite good in classifying cardiovascular diseases.

Keywords: Classification, *K-Nearest Neighbor*, *Naïve Bayes*, Cardiovascular

1. PENDAHULUAN

Penyakit kardiovaskular merupakan penyakit tidak menular yang dapat menyebabkan kematian nomor satu di dunia. Lebih dari tiga per empat atau sekitar 80 % dari kematian akibat penyakit kardiovaskular terjadi di negara berpenghasilan rendah sampai menengah termasuk Indonesia. Penyakit kardiovaskular sebagai penyakit dengan gangguan pada jantung dan pembuluh darah seperti penyakit jantung koroner, gagal jantung, hipertensi dan stroke Penyakit kardiovaskular juga dapat menyebabkan kecacatan dan penurunan kualitas hidup terkait kesehatan [1]. Menurut [2] World Health Organization (WHO) (2002), melaporkan *Noncommunicable Disease (NCDs)* atau penyakit non infeksi menyumbang 60 persen mortalitas dan 47 persen beban penyakit di dunia dan akan terus meningkat dengan prediksi pada tahun 2020, kematian akibat NCDs adalah 73 persen dan 60 persen beban penyakit di dunia. peningkatan angka kematian setiap tahunnya. Dengan meningkatnya korban setiap tahunnya, maka dibutuhkan sebuah sistem klasifikasi yang dapat memberikan informasi tentang penyakit serta dapat melakukan pengecekan klasifikasi secara dini tentang penyakit serangan jantung yang dialami oleh seseorang [3]. Seiring dengan perkembangan ilmu pengetahuan dan teknologi informasi, kehadiran cabang ilmu baru di bidang komputer data mining [4]. Klasifikasi data mining adalah penempatan objek-objek ke salah satu dari beberapa

kategori yang telah ditetapkan sebelumnya [5]. Dalam penelitian ini digunakan klasifikasi dengan dua algoritma untuk membandingkan algoritma mana yang terbaik. Algoritma yang digunakan yaitu *Naïve Bayes* dan *Support Vector Machine*. *Naïve Bayes* adalah salah satu algoritma pembelajaran induktif yang paling efektif dan efisien untuk machine learning dan data mining. Performa *naïve bayes* yang kompetitif dalam proses klasifikasi [6]. *SVM* merupakan suatu teknik untuk menemukan hyperplane yang bisa memisahkan dua set data dari dua kelas yang berbeda.

Algoritma *naïve bayes* sendiri memiliki kelebihan yaitu. mudah di implementasikan dan mampu mendapatkan hasil baik dalam banyak kasus. Namun kekurangan *Naïve Bayes* yaitu tidak berlaku jika probabilitas kondisionalnya adalah nol, apabila nol maka probabilitas prediksi akan bernilai nol juga mengasumsikan variabel bebas [8]. Kemudian untuk algoritma *Support Vector Machine* memiliki kelebihan diantaranya adalah dalam menentukan jarak menggunakan *Support Vector Machine* sehingga proses komputasi menjadi cepat [7]. Adapun kekurangan *Support Vector Machine* (*SVM*) adalah sulit dipakai problem berskala besar. Dalam hal ini dimaksudkan dengan jumlah sampel yang diolah. [9]

Pada beberapa penelitian sebelumnya pernah dilakukan prediksi yang dilakukan oleh [10] menggunakan algoritma *naïve bayes* yang melakukan prediksi penyakit diabetes melitus dengan akurasi 90%. Kemudian [11] dengan *naïve bayes* yang melakukan klasifikasi penyakit jantung mendapatkan hasil akurasi 82% dan [12] melakukan penelitian klasifikasi data penyakit ayam menggunakan algoritma *Support Vector Machine* dengan tingkat akurasi 84%. [13] melakukan prediksi kepuasan pengunjung taman tabebuya menggunakan *Support Vector Machine* mendapatkan hasil dengan tingkat akurasi 84%. Dari beberapa penelitian sebelumnya dapat disimpulkan bahwa kedua algoritma cukup banyak digunakan dalam penelitian, dimana algoritma *naïve bayes* dapat melakukan klasifikasi dengan baik dengan tingkat akurasi yang cukup besar yaitu 90%, sementara *Support Vector Machine* juga telah banyak digunakan dalam penelitian sebelumnya dan memiliki tingkat akurasi yang sedikit lebih kecil dari *naïve bayes* yaitu 84% hingga 86%. Berdasarkan pengetahuan yang telah dijelaskan sebelumnya, maka pada penelitian ini akan melakukan klasifikasi dengan menggunakan algoritma *naïve bayes* dan *Support Vector Machine*.

2. METODE PENELITIAN

2.1. Pengumpulan Data

Data ini diambil dari website kaggle <https://www.kaggle.com/datasets/sulianova/cardiovascular-disease-dataset>. Data yang digunakan ini terdiri dari 462 data pasien penyakit kardiovaskular berupa format csv. Data yang digunakan terdapat 10 atribut yang digunakan untuk melakukan klasifikasi penyakit kardiovaskular adalah tekanan darah sistolik, tembakau, kolesterol lipoprotein densitas rendah, adipositas, riwayat keluarga penyakit jantung, faktor dengan tingkat "Absen" dan "Ada", perilaku tipe-A, kegemukan, konsumsi alkohol saat ini, usia, respon, penyakit jantung coroner. Secara lengkap atribut yang digunakan dapat dilihat pada tabel 1 berikut.

Tabel 1. Atribut

Atribut	Keterangan	Nilai	Data Kosong
Sbp	Tekanan darah sistolik (Hg)	101-218	Null
Tobacco	Tembakau (mg)	0-31	Null
Ldl	Kadar kolesterol densitas rendah(mg)	0.98-15.33	Null
Adiposity	Indeks Adipositas (imt)	6.74-42.49	Null
Famhist	Riwayat penyakit jantung	0: Absen/Tidak ada 1: Ada	Null
Type-A	Indeks Personality (mmpi) Tes psikologi yang menilai kepribadian	13-78	Null
Obesity	Indeks Obesitas (imt)	14-47	Null
Alcohol	Kadar alcohol (ml)	0-147	Null
Age	Usia	15-64	Null
Chd	Cardio/penyakit jantung	0 : Tidak Aktif 1 : Aktif	Null

2.2 Persiapan Data

Dalam data tersebut dilakukan *pre-processing* yaitu atribut Famhist yaitu riwayat penyakit jantung dari keluarga yaitu 0 untuk tidak ada dan 1 untuk ada. Kemudian atribut Chd yaitu indeks atau respon dari penyakit yang disimbolkan 0 untuk tidak aktif dan 1 untuk aktif. Data tersebut dilakukan klasifikasi

dengan pemodelan sesuai algoritma *K-Nearest Neighbor* dan *Naïve Bayes Classifier* dalam memprediksi penyakit kardiovaskular. Pada tahap selanjutnya, perlu pembagian data menjadi dua yaitu dengan menggunakan *percentage split* dengan ukuran data training 80% dan data testing 20% untuk algoritma *Naïve Bayes* dan *Support Vector Machine*. Kemudian menggunakan *K-Fold Validation* dimana mengambil nilai K sebesar 10 ke dalam kedua algoritma.

2.2 Penerapan Algoritma *Naïve Bayes*

Naïve Bayes merupakan metode pengklasifikasian probabilistik sederhana. Metode ini akan menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Metode *naïve bayes* menganggap semua atribut pada setiap kategori tidak memiliki ketergantungan satu sama lain (independen) [14].

Persamaan dari *Naïve Bayes* dapat dilihat pada persamaan 1 berikut:

1. Menghitung jumlah kategori dari setiap variabel
2. Menghitung peluang pada setiap kategori
3. Menentukan frekuensi atau jumlah kemunculan pada setiap kategori
4. Menentukan kategori dengan nilai maksimal

Perhitungan *Naïve Bayes* dapat dilakukan dengan rumus (1):

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

dimana:

- X : Data dengan kelas yang belum diketahui
- H : Hipotesis data X merupakan suatu kelas spesifik
- P(H | X) : Probabilitas hipotesis H berdasarkan kondisi X
- P(H) : Probabilitas hipotesis H
- P(X | H) : Probabilitas X berdasarkan kondisi pada hipotesis H
- P(X) : Probabilitas hipotesis X

2.3 Penerapan Algoritma *Support Vector Machine*

Support Vector Machine dapat dengan dijelaskan bahwa algoritma yang digunakan sebagai upaya untuk menemukan hyperplane optimal yang bertindak sebagai pembatas untuk dua kelas di ruang input. Data yang tersedia dinotasikan sebagai $x_i \in R^d$ kemudian label masing-masing dinotasikan dengan $y_i \in \{-1, +1\}$ untuk $i = 1, 2, \dots, m$ dimana m merupakan jumlah banyaknya data. Diasumsikan bahwa terdapat data learning dengan data points x_i ($i=1,2,\dots,m$) memiliki dua kelas $y_i = \pm 1$ yaitu kelas positif (+1) dan kelas negatif (-1), yang didefinisikan pada persamaan (2) dibawah ini [15]:

$$\vec{w}_i \cdot \vec{x}_i + b = 0 \quad (2)$$

Rumus yang digunakan untuk menghitung hasil prediksi dengan *hyperplane* tunggal dan sesuai dengan b dan w yang akan diperoleh didefinisikan pada persamaan (3) dan (4) sebagai berikut:

$$f(\phi(x)) = \text{sign}(w \cdot \phi(x)) + b \quad (3)$$

$$f(\phi(x)) = \text{sign}(\sum_{i=1}^n \alpha_i y_i \phi(x_i)^T \cdot \phi(x) + b) \quad (4)$$

Keterangan: w = nilai weight
 x = nilai variabel input
 b = nilai bias

Selanjutnya, lebar margin y atau jarak dari x^+ (data yang terletak pada kelas $y = +1$) ke *hyperplane* atau jarak dari x^- (data yang terletak pada kelas $y = -1$) ke *hyperplane* dilakukan dengan memaksimalkan $\|w\|$ dengan didefinisikan pada persamaan (5) dan (6) sebagai berikut:

$$w = \sum_{i=1}^n \alpha_i y_i K(x, x_i) \quad (5)$$

$$b = -\frac{1}{2}(w \cdot x^+ + w \cdot x^-) \quad (6)$$

Dengan kendala $y_i(\langle w, x_i \rangle + b \geq 1), i = 1, 2, \dots, n$ merupakan *hyperplane* dengan *margin* maksimum. Persamaan yang digunakan untuk mencari nilai *alpha* didefinisikan pada persamaan (7) berikut:

$$\alpha_n = \frac{n}{\sum K(n*n)} \quad (7)$$

2.4 Evaluasi Hasil

Evaluasi dari hasil klasifikasi dilakukan dengan *confusion matrix*. Metode ini dipresentasikan dari hasil klasifikasi yang dapat dilihat pada Tabel 2.

Table 2. *Confusion Matrix*

Kelas	Nilai Aktual		
	Positif	Negatif	
Nilai Prediksi	Positif	True Positif	False Negatif
	Negatif	False Positif	True Negatif

Ketika hasil berada pada kolom *True Positif* maka hasil bernilai benar dan teridentifikasi sebagai *positif*. ketika hasil berada pada kolom *False Positif* maka hasil adalah salah dan teridentifikasi sebagai *positif*. Ketika hasil berada pada kolom *False Negatif* maka hasil adalah salah dan teridentifikasi sebagai negatif dan ketika hasil berada di kolom *True Negatif*, maka akan bernilai benar dan teridentifikasi *negative* [16]. Akurasi klasifikasi dicari dengan menggunakan persamaan 8.

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (8)$$

Precision merupakan tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh system yang dicari menggunakan persamaan 9.

$$Presisi = \frac{TP}{TP+FP} \quad (9)$$

Recall merupakan pengukuran untuk data dengan klasifikasi positif yang benar oleh system yang dicari menggunakan persamaan 10.

$$Recall = \frac{TP}{TP+FN} \quad (10)$$

3. HASIL DAN PEMBAHASAN

Dari data yang digunakan dalam penelitian ini didapat dari situs Kaggle yang berjudul Penyakit Kardiovaskular. Terdapat 463 data terkait dalam dataset ini. Dalam proses penelitian terdapat 12 atribut diantara tekanan darah sistolik, tembakau, kolesterol lipoprotein densitas rendah, adipositas, riwayat keluarga penyakit jantung, faktor dengan tingkat "Absen" dan "Ada", perilaku tipe-A(personality), kegemukan, konsumsi alkohol saat ini, usia, respon, penyakit jantung coroner. Indeks dari pasien terjangkit penyakit kardiovaskular ya atau tidak. Metode yang digunakan adalah *Percentage split* dan *K-Fold cross-validation*. K yang dipilih adalah 10 dengan dibagi menjadi kelompok latihan 80% dan kelompok test 20%.

3.1 Algoritma *Naïve Bayes*

Hasil klasifikasi algoritma *Naïve Bayes* dalam penyakit kardiovaskular menggunakan pembagian data latihan dengan *percentage split* dan *K-fold cross validation* dimana tingkat keberhasilan yang dicapai dengan penggunaan diukur dengan *Matriks confusion*. Pembagian data dengan *percentage split* dapat dilihat pada Tabel 2 dibawah ini.

Tabel 2. *Percentage Split*

Kelas		Label Sebenarnya	
		0	1
Label	0	228	74
Prediksi	1	59	101

Berdasarkan tabel 2 dapat dilihat bahwa sebanyak 228 kasus yang positif diprediksi positif, 74 kasus yang sebenarnya positif diprediksi negatif, 59 kasus yang negatif diprediksi positif, dan 101 kasus yang negatif diprediksi negatif. Selain itu, perhitungan menggunakan *percentage split* mendapatkan tingkat akurasi yang sebesar 70%. Kemudian dengan metode latihan *K-cross validation* dengan nilai sebesar 10. *Matriks confusion* yang dihasilkan dapat dilihat pada tabel 3.

Tabel 3. *K-fold cross validation*

Kelas		Label Sebenarnya	
		0	1
Label	0	47	14
Prediksi	1	14	18

Berdasarkan tabel 3 dapat dilihat bahwa sebanyak 47 kasus yang positif diprediksi positif, 14 kasus yang sebenarnya positif diprediksi negatif, 14 kasus yang negatif diprediksi positif, dan 18 kasus yang negatif diprediksi negatif. Selain itu, perhitungan menggunakan *K-fold cross validation* mendapatkan tingkat akurasi yang sebesar 71%.

3.2 Algoritma *Support Vector Machine*

Hasil klasifikasi algoritma *Support Vector Machine* dalam penyakit kardiovaskular juga menggunakan pembagian data latihan dengan *percentage split* dan *K-fold cross validation* dimana tingkat keberhasilan yang dicapai dengan penggunaan diukur dengan *Matriks confusion*. Pembagian data dengan metode *percentage split* dapat dilihat pada Tabel 4 dibawah ini.

Tabel 4. *Percentage Split*

Kelas		Label Sebenarnya	
		0	1
Label	0	57	0
Prediksi	1	36	0

Berdasarkan tabel 4 dapat dilihat bahwa sebanyak 57 kasus yang positif diprediksi positif, 0 kasus yang sebenarnya positif diprediksi negatif, 36 kasus yang negatif diprediksi positif, dan 0 kasus yang negatif diprediksi negatif. Selain itu, perhitungan menggunakan *K-cross validation* mendapatkan tingkat akurasi yang sebesar 61%. Kemudian dengan metode latihan *K-cross validation* dengan nilai sebesar 10. *Matriks confusion* yang dihasilkan dapat dilihat pada tabel 5.

Tabel 5 *K-fold cross validation*

Kelas		Label Sebenarnya	
		0	1
Label	0	302	0
Prediksi	1	160	0

Berdasarkan tabel 5 dapat dilihat bahwa sebanyak 307 kasus yang positif diprediksi positif, 0 kasus yang sebenarnya positif diprediksi negatif, 160 kasus yang negatif diprediksi positif, dan 0 kasus yang negatif diprediksi negatif. Selain itu, perhitungan menggunakan *K-cross validation* mendapatkan tingkat akurasi yang sebesar 65%.

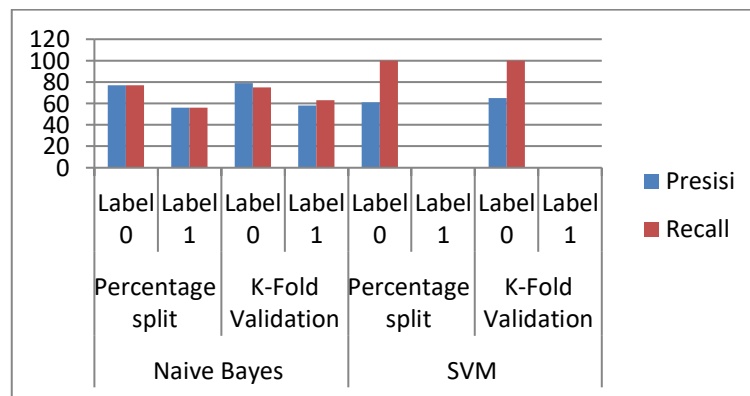
3.3 Perbandingan Kedua Algoritma

Hasil prediksi dari kedua algoritma *Naïve Bayes* dan *Support Vector Machine* dengan metode latihan *percentage split* dan *k-fold cross validation* menunjukkan bahwa algoritma SVM dengan kedua metode latihan tersebut cukup baik dalam memprediksi pengklasifikasian penyakit kardiovaskular. Perbandingan hasil prediksi dari kedua metode dapat dilihat pada Tabel 6 berikut.

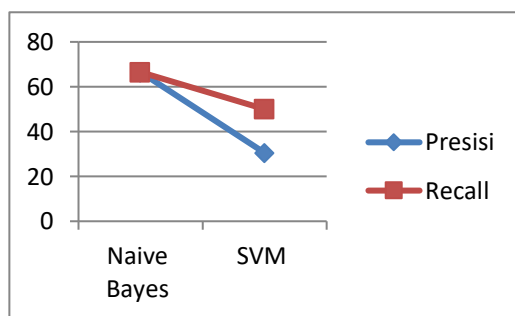
Tabel 6. Nilai Presisi, Recall, dan Akurasi *Naïve Bayes* dan *Support Vector Machine*

Algoritma	Pemodelan	Label	Presisi	Recall	Akurasi
<i>Naïve Bayes</i>	<i>Percentage split</i>	0	77%	77%	70%
		1	56%	56%	
	<i>K-Fold cross Validation</i>	0	79%	75%	71%
		1	58%	63%	
<i>SVM</i>	<i>Percentage split</i>	0	61%	100%	61%
		1	0%	0%	
	<i>K-Fold cross Validation</i>	0	65%	100%	65%
		1	0%	0%	
		1	0%	0%	

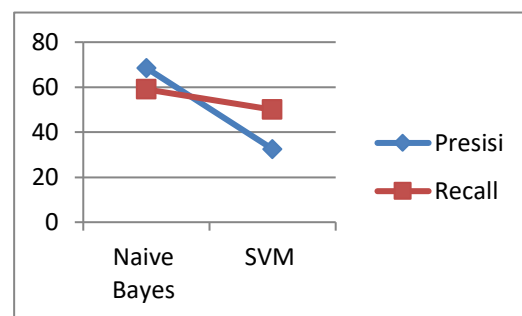
Berdasarkan Tabel 6 diketahui bahwa dari algoritma *Naïve Bayes* didapat dengan metode latih *K-fold cross validation* memiliki tingkat akurasi yang cukup baik, dimana memberikan nilai akurasi dan presisi yang cukup tinggi. Kemudian dari algoritma *SVM* mendapatkan tingkat akurasi yang cukup baik pada metode latih *K-fold cross validation* dimana memberikan nilai akurasi dan *recall* yang cukup tinggi meskipun tingkat akurasinya masih dibawah *Naïve Bayes*. Hasil akurasi, presisi, dan *recall* dari kedua metode latih tersebut dapat dilihat dalam Gambar 1.



Gambar 1. Nilai Presisi dan Recall dari Algoritma *Naïve Bayes* dan *K-Nearest Neighbor*



Gambar 2. Rata-Rata *Percentage split*



Gambar 3. Rata-rata *K-fold validation*

Pada Gambar 2 dan 3 terlihat bahwa nilai presisi dan *recall* dari rata-rata *Percentage split* dan *K-fold validation* yang digunakan tidak berbeda jauh meskipun memiliki sedikit perbandingan, dimana algoritma memiliki sedikit keunggulan dalam memprediksi klasifikasi penyakit kardiovaskular. Selain itu, perbandingan hasil pada penelitian ini terhadap penelitian sebelumnya mengenai klasifikasi kardiovaskular dapat di tabel 7 berikut.

Tabel 7. Perbandingan Algoritma Dalam Penyakit Kardiovaskular

Algoritma	Dataset	Hasil		
		Akurasi	Presisi	Recall
Decision Tree [17]	<i>Cardiovascular Disease Dataset</i>	0.84	0.79	0.82
KNN [17]	<i>Cardiovascular Disease Dataset</i>	0.83	0.1	0.29
Logistic Regression [17]	<i>Cardiovascular Disease Dataset</i>	0.68	0.69	0.68
ANN [18]	<i>Cardiovascular Disease Diagnosis Dataset</i>	0.73	0.71	0.71
Naïve Bayes [19]	Data Penyakit Jantung di Prabumulih	0.92	-	-
SVM [20]	Data Penyakit Jantung pasien di Cleveland	0.90	0.90	0.92

Berdasarkan tabel 7 dapat dilihat bahwa algoritma klasifikasi *Naïve Bayes* dan *SVM* memberikan hasil tingkat akurasi yang cukup tinggi dibandingkan dengan algoritma lain pada beberapa penelitian yang telah dilakukan. Hal ini menunjukkan bahwa algoritma *Naïve-Bayes* dan *SVM* merupakan algoritma klasifikasi yang cukup baik yang dapat digunakan untuk memprediksi klasifikasi penyakit kardiovaskular.

4. KESIMPULAN

Berdasarkan hasil yang diperoleh dalam penelitian ini, metode *Naïve-Bayes* ditemukan sangat baik pada klasifikasi penyakit kardiovaskular yang dilakukan dalam penelitian ini. kemudian dilihat presisi, *recall*, dan akurasi menggunakan metode *k-fold cross-validation* mendapatkan nilai yang lebih baik daripada menggunakan metode *percentage split*. Dengan demikian dapat disimpulkan algoritma *Naïve Bayes* cukup baik dalam prediksi klasifikasi penyakit kardiovaskular

REFERENSI

- [1] Yogyakarta, K. D. I., Farmako, J., Utara, S., Farmako, J., & Utara, S. (2020). *11096-31793-1-Sm. 13*(1), 1–12.
- [2] Rosjidi, C. H., & Isro'in, L. (2014). Perempuan lebih rentan terserang penyakit kardiovaskular. *Jurnal Florence, VII*(1), 1–10
- [3] Ari Bianto, M. (2019). Perancangan Sistem Klasifikasi Penyakit Jantung Menggunakan *Naïve Bayes* Designing a Heart Disease Classification System Using *Naïve Bayes*. *Citec Journal, 6*(1).
- [4] Dwi Septiani Program Studi Manajemen Informatika AMIK BSI Jakarta Jl Kramat Raya No, W., & Pusat, J. (2017). Dan *Naïve Bayes* Untuk Prediksi Penyakit Hepatitis. *Jurnal Pilar Nusa Mandiri, 13*(1), 76–84.
- [5] Susanto, H., & Sudiyatno, S. (2014). Data mining untuk memprediksi prestasi siswa berdasarkan sosial ekonomi, motivasi, kedisiplinan dan prestasi masa lalu. *Jurnal Pendidikan Vokasi, 4*(2), 222–231.
- [6] Syarli, & Muin, A. A. (2016). Metode *Naïve Bayes* Untuk Prediksi Kelulusan. *Jurnal Ilmiah Ilmu Komputer, 2*(1), 22–26.
- [7] Pushpita Anna Octaviani, Yuciana Wilandari, D. I. (2014). Penerapan Metode *SVM* Pada Data Akreditasi Sekolah Dasar Di Kabupaten Magelang. *Jurnal Gaussian, 3*(8), 811–820.
- [8] Suprianto, S. (2020). Implementasi Algoritma *Naive Bayes* Untuk Menentukan Lokasi Strategis Dalam Membuka Usaha Menengah Ke Bawah di Kota Medan (Studi Kasus: Disperindag Kota Medan). *Jurnal Sistem Komputer Dan Informatika (JSON), 1*(2), 125.
- [9] Hilmiyah, F. (2017). Prediksi Kinerja Mahasiswa Menggunakan *Support Vector Machine* untuk Pengelola Program Studi di Perguruan Tinggi (Studi Kasus: Program Studi Magister Statistika ITS). *Departemen Manajemen Teknologi Bidang Keahlian Manajemen Teknologi Informasi Fakultas Bisnis Dan Manajemen Teknologi Institut Teknologi Sepuluh Nopember Surabaya, 1–99*.
- [10] Ridwan, A. (2020). Penerapan Algoritma *Naïve Bayes* Untuk Klasifikasi Penyakit Diabetes Mellitus. *Jurnal SISKOM-KB (Sistem Komputer Dan Kecerdasan Buatan), 4*(1), 15–21..
- [11] Evanko, D. (2010). Optical imaging of the native brain. *Nature Methods, 7*(1), 34.
- [12] Nurcahya, E. D. (2017). Klasifikasi Penyakit Ayam Menggunakan Metode *Support Vector Machine*. *VOLT: Jurnal Ilmiah Pendidikan Teknik Elektro, 2*(1), 45.
- [13] Darmawan, A., Kustian, N., & Rahayu, W. (2018). Implementasi Data Mining Menggunakan Model *SVM* untuk Prediksi Kepuasan Pengunjung Taman Tabebuya. *STRING (Satuan Tulisan Riset Dan*

- Inovasi Teknologi*, 2(3), 299.
- [14] Devita, R. N., Herwanto, H. W., & Wibawa, A. P. (2018). Perbandingan Kinerja Metode Naive Bayes dan K-Nearest Neighbor untuk Klasifikasi Artikel Berbahasa Indonesia. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 5(4), 427.
- [15] Fiska, R. R. (2017). Penerapan Teknik Data Mining dengan Metode Support Vector Machine. *Sains Dan Teknologi Informasi (SATIN)*, 3(1).
- [16] Ningrum, H. C. S. (2018). Perbandingan Metode Support Vector Machine (SVM) Linear, Radial Basis Function (RBF), dan Polinomial Kernel dalam Klasifikasi Bidang Studi Lanjut Pilihan Alumni UII [skripsi]. *Tugas Akhir Statistika Universitas Islam Indonesia*, 1–90.
- [17] Minou, J., Mantas, J., Malamateniou, F., & Kaitelidou, D. (2020). Classification Techniques for Cardio-Vascular Diseases Using Supervised Machine Learning. *Medical Archives (Sarajevo, Bosnia and Herzegovina)*, 74(1), 39–41.
- [18] Ei, H., Cho, E., & Hlaing, N. Y. (2020.). *DATA MINING CLASSIFICATION TECHNIQUES FOR. 1*, 1–6.
- [19] Anggara, B., & D, R. M. N. H. (2021.). *PENERAPAN DATA MINING RUMAH SAKIT UMUM PRABUMULIH MENGGUNAKAN ALGORITMA NAIVE BAYES (STUDY KASUS: PENYAKIT JANTUNG)*. 209–220.
- [20] Permana, D. S., & Silvanie, A. (2021). Prediksi Penyakit Jantung Menggunakan Support Vector Machine Dan Python Pada Basis Data Pasien. *Jurnal Nasional Informatia*, 2(1), 29–34.