

Klasifikasi Penyakit Hati dengan Menggunakan Metode Naive Bayes

Wahyugi Fadri^{1✉}

¹Independent Researcher

wahyugifadri19@gmail.com

Abstract

Data mining is the process of finding relationships in data that are not known to the user and understandably presenting them so that these relationships can form the basis of decision-making. The purpose of this study was to design a data mining website using the Naive Bayes method to classify liver disease symptoms and to test a data mining website using the Naive Bayes method to assist in classifying liver disease. The data was processed by 50 medical records of liver disease in Lubuk Basung Hospital which consisted of symptoms and diseases. The method used is the Naive Bayes method which has a fairly high accuracy value. one method that can be used to classify data. Bayesian classification is a statistical classifier that can be used to predict the probability of belonging to a class. Then it was developed into a website-based form using the PHP Framework Codeigniter programming language and MySQL as the database. The system to be built is based on the Naive Bayes method which is one of the Data Mining methods. The test results from 50 medical record data of patients with liver disease, by applying the Naive Bayes method which is applied to data mining, is useful for comparing hospital data with system data. From 45 training data and 5 testing data, the accuracy rate is 60%.

Keywords: Data Mining, Classification, Symptom, Liver, Naive Bayes.

Abstrak

Data mining is the process of finding relationships in data that are not known to the user and presenting them in an understandable way so that these relationships can form the basis of decision making. The purpose of this study was to design a data mining website using the Naive Bayes method to classify liver disease symptoms and to test a data mining website using the Naive Bayes method to assist in classifying liver disease. The data was processed by 50 medical records of liver disease in Lubuk Basung Hospital which consisted of symptoms and diseases. The method used is the Naive Bayes method which has a fairly high accuracy value. one method that can be used to classify data. Bayesian classification is a statistical classifier that can be used to predict the probability of belonging to a class. Then it was developed into a website-based form using the PHP Framework Codeigniter programming language and MySQL as the database. The system to be built is based on the Naive Bayes method which is one of the Data Mining methods. The test results from 50 medical record data of patients with liver disease, with the application of the Naive Bayes technique in data mining, are useful for comparing hospital data with system data. From 45 training data and 5 testing data, the accuracy rate is 60%.

Kata kunci: Data Mining, Klasifikasi, Gejala, Hati, Naive Bayes.

Jidt is licensed under a Creative Commons 4.0 International License.



1. Pendahuluan

Knowledge Discovery in Database (KDD) adalah proses yang melibatkan pengumpulan dan penggunaan data historis untuk menemukan keteraturan, pola, atau hubungan dalam kumpulan data yang besar. [1]. Tahapan dalam menganalisis data mining dengan menggunakan pendekatan Knowledge Discovery in Database (KDD) Secara umum, ini mencakup beberapa langkah seperti data selection, preprocessing, transformation, data mining dan evaluation [2]. Data mining adalah proses menemukan hubungan dalam data yang tidak diketahui oleh pengguna dan menyajikannya dengan cara yang dapat dipahami sehingga hubungan tersebut dapat menjadi dasar pengambilan keputusan [3].

Hati adalah salah satu organ tambahan manusia, sehingga dapat melakukan pekerjaan yang sangat baik di dalam tubuh. Tapi ada stadium penyakit liver di

dalam tubuh manusia [4]. Penyakit hati dapat mempengaruhi fungsi hati, tetapi dapat mengidentifikasi gejala klinis dan fisik pasien [5]. Ada berbagai faktor yang menyebabkan penyakit liver ini terus meningkat. Diantaranya gaya hidup masyarakat yang rawan dan tidak terkontrol seperti : diabetes melitus, dislipidemia, hipertensi dan lain-lain [6].

Naive 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. Naive Bayes merupakan suatu kelas keputusan, dengan menggunakan perhitungan probabilitas matematika dengan syarat bahwa nilai keputusan adalah benar, berdasarkan informasi obyek [7]. Naive Bayes Classifier dapat dilatih secara efisien dalam pembelajaran dengan cara diawasi [8].

Klasifikasi penyakit Hepatocellular Carcinoma menggunakan metode Naive Bayes, berdasarkan tabel dari hasil pengujian terlihat bahwa nilai Akurasi sebesar 70,30%, Presisi sebesar 73,53% dan Recall sebesar 77,32% serta hasil performa yang diukur menggunakan Kurva ROC (Receiver Operating Characteristic) dengan nilai yang dihasilkan berupa nilai AUC sebesar 0.783 yang termasuk dalam kategori Fair Classification atau kategori Klasifikasi Cukup [9].

Nilai prediksi dihitung menggunakan classifier Naive Bayes, yang bertindak sebagai classifier dan memberikan output nilai probabilitas. Naive Bayes dapat dengan mudah mengklasifikasikan penyakit terlepas dari apakah pasien memiliki kemungkinan memiliki penyakit yang dipilih atau tidak. Selain itu, dapat membantu petugas kesehatan membuat keputusan dengan akurasi 80% berdasarkan diagnosis dokter. [10].

2. Metodologi Penelitian

2.1. Data Mining

Data mining memproses data dalam jumlah besar menjadi informasi atau pengetahuan, diperlukan suatu teknik/metode [11]. Data mining terkait dengan bidang ilmu lain, yaitu sistem database, data warehousing, statistik, machine learning, information retrieval, dan komputasi tingkat tinggi. Selain itu, data mining didukung oleh ilmu lain seperti neural network, pengenalan pola, spatial data analysis, image database, signal processing [12].

2.2. Klasifikasi

Klasifikasi adalah menemukan fungsi atau model sebagai konsep diferensial atau kelas data untuk memperkirakan kelas objek yang belum diketahui [13]. Salah satu Metode yang digunakan dalam melakukan teknik yang diterapkan dalam pengelompokan ini ialah teknik Naive Bayes, metode klasifikasi statistik untuk produk probabilitas sederhana untuk memprediksi probabilitas atau probabilitas dengan menambahkan kombinasi dan frekuensi nilai dalam suatu kumpulan data. [14]

2.3. Metode Naive Bayes

Metode Naive bayes adalah salah satu metode machine learning yang menggunakan klasifikasi menggunakan probabilitas dan statistik untuk memprediksi kemungkinan kejadian masa depan yang dilihat berdasarkan pengalaman dimasa sebelumnya dan metode naïve bayes dinilai baik dalam hal pengklasifikasian dibandingkan dengan metode klasifikasi lain yang dianggap baik dari segi akurasi dan efisiensi [15]. Aritmatika naive bayes dilakukan dengan menetapkan naive bayes classifier (nc) untuk setiap kelas. Asalkan jika dugaan sementara benar, maka sama dengan 1, dan jika dugaan sementara salah, maka 0 [16]. Secara umum Naive Bayes dinyatakan pada Persamaan (1)

$$D(JXF|CNV) = \frac{D(CNV|JXF) * D(JXF)}{D(CNV)} \quad (1)$$

Dimana nilai JXF adalah hipotesis khusus (benar atau salah), CNV merupakan bukti yang telah diamati, $D(JXF|CNV)$ adalah probabilitas hipotesis berdasarkan kondisi atau disebut juga probabilitas posterior. $D(CNV|JXF)$ merupakan probabilitas CNV berdasarkan hipotesis JXF, $D(JXF)$ merupakan probabilitas JXF disimpulkan sebelum bukti baru atau disebut juga probabilitas priot. $D(CNV)$ adalah probabilitas marginal CNV dalam semua hipotesis yang mungkin.

Berdasarkan Persamaan (1) maka dapat dihasilkan persamaan Naive Bayes, seperti Persamaan (2)

$$D(CNV) * [D(JXF_1 | CNV_1) * ..., D(GXF_1 | CNV_1)] \quad (2)$$

. Dimana $D(CNV)$ merupakan probabilitas marginal CNV dalam semua kemungkinan dugaan sementara dan $D(GXF|CNV)$ merupakan peluang yang disesuaikan dengan kondisi.

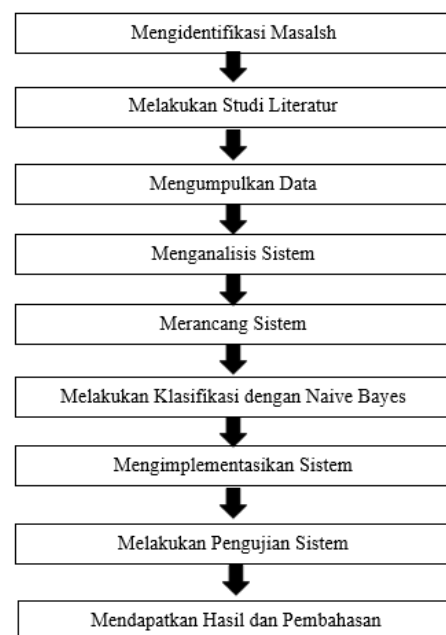
Selanjutnya adalah perhitungan nilai P setiap penyakit dengan rumus, terdapat pada Persamaan (3)

$$P(T) = \frac{h}{H} * 100\% \quad (3)$$

Dimana P merupakan probabilitas terjadinya peristiwa, T adalah peristiwa, h merupakan jumlah kejadian berhasil dan H merupakan jumlah semua kejadian.

2.4. Kerangka Penelitian

kegiatan penelitian yang dilakukan secara sistematis, berupa mencari, merumuskan maupun menganalisa sampai dapat menyusun sesuai dengan langkah-langkah yang telah ditentukan. Langkah-langkah tersebut dapat dilihat pada Gambar 1.



Gambar 1. Kerangka Penelitian

3. Hasil dan Pembahasan

3.1. Analisa Data

Data yang dianalisa bertujuan untuk mengklasifikasi penyakit hati pada pasien, agar hasil analisa dapat

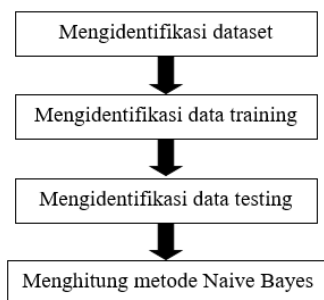
dijadikan acuan dalam pengambilan keputusan. Metode yang dipilih pada penelitian ini adalah metode Naïve Bayes. Data yang dianalisa berjumlah 50 data uji, namun yang disajikan sebanyak 18 data uji yang dapat dilihat pada Tabel 1.

Tabel 1. Tabel Data Rekam Medis Pasien

No	Tanggal Konsultasi	Nama	Umur (th)	L/P	Gejala	Penyakit
1.	27/01/2022	Pasien 1	40	L	Perut Kembung, Mual, muntah air, BAB bau kencur	Sirosis
2.	21/01/2022	Pasien 2	37	P	Sesak nafas, perut sakit, batuk berdahak, dada debar-debar, kedua kaki sakit.	Sirosis
3.	21/01/2022	Pasien 3	30	P	Sakit pinggang, perut keras, perut membesar lebih kurang 2 bulan, nafas sesak	Sirosis
4.	27/01/2022	Pasien 4	31	L	Muntah 1x, muntah darah, bab warna hitam mual	Sirosis
1.	21/01/2022	Pasien 5	33	P	Muntah darah, Muntah, BAB warna hitam, Mual	Sirosis
2.	15/01/2022	Pasien 6	39	L	Nyeri perut, Muntah, Mual	Sirosis
3.	14/01/2022	Pasien 7	30	P	Muntah darah, BAB warna hitam, Kurang nafsu makan	Sirosis
4.	07/01/2022	Pasien 8	36	P	Mual, Muntah, Perut membuncit	Sirosis
5.	07/01/2022	Pasien 9	30	L	Perut sakit, Nyeri dada, Perut pembesar	Sirosis
6.	06/01/2022	Pasien 10	40	P	Badan Lesu, Demam, Sakit Pinggang	Sirosis
7.	04/01/2022	Pasien 11	42	P	Lesu, Perut membesar, Kaki cedera, Kepala sakit	Sirosis
8.	04/01/2022	Pasien 12	43	P	Mual, Badan lelah, Nafsu makan kurang, Demam	Sirosis
9.	03/01/2022	Pasien 13	42	L	Pilek, Kedua Kaki sakit, BAB Warna Hitam, Diare	Sirosis
10.	09/03/2022	Pasien 14	41	P	Nafsu makan kurang	Sirosis
11.	31/01/2022	Pasien 15	39	L	Perut sakit, BAB berdarah, Kepala sakit	Abses Hati
12.	08/02/2022	Pasien 16	33	L	Demam, Perut nyeri, Mual, Tidak mau makan	Abses Hati
13.	25/02/2022	Pasien 17	43	L	Perut membesar, BAB Warna Gitam	Sirosis
14.	02/02/2022	Pasien 18	44	P	Demam, Perut kembung, Muntah air, Nyeri kepala BAB lendir	Sirosis
					Badan Lesu, Nafsu makan kurang, Kedua Kaki sakit, Perut sakit, Perut membesar	Sirosis

3.2. Analisa Sistem

Analisa sistem merupakan langkah-langkah perancangan sistem yang digambarkan pada bagan alir analisa menggunakan algoritma Naïve Bayes. Agar lebih jelas dapat dilihat pada Gambar 2.



Gambar 2. Flowchart Naive Bayes

3.3. Mengidentifikasi Dataset

Dataset pada penelitian memiliki jumlah instance sebanyak 50 data. Transformasi dataset diberikan nilai keterangan untuk kondisi Iya dnegan nilai 1 dan Tidak dengan Nilai 0.

Target diagnosa penyakit yang akan dicapai sebagai hasil akhir penilaian diklasifikasikan 3 target yaitu disajikan pada Tabel 2

Tabel 2. Tabel Klasifikasi Diagnosa Penyakit

No	Penyakit	Kode
1	Abses Hati	AH
2	Hepatitis	H
3	Sirosis	S

Data set pada penelitian ini memiliki jumlah instance sebanyak 50. Dataset memiliki atribut atau fitur yang mendeskripsikan setiap instance. Daftar dataset rekam medis pasien yang ditampilkan sebanyak 5 data dapat dilihat pada Tabel 3

Tabel 3. Tabel Dataset Rekam Medis Pasien

A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	A17	A18	Diagnosa
1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	Sirosis
0	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	Sirosis
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Sirosis
0	1	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	Sirosis
0	1	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	Sirosis

Data yang telah didapatkan pada Tabel 4 kemudian diolah pada data training. Total data yang dihasilkan berjumlah 45 data, dan untuk tabel data training hanya

ditampilkan 5 data. Hasil dari data testing dapat dilihat pada Tabel 4.

Tabel 4. Tabel Data Training

A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	A17	A18	Diagnosa
1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	Sirosis
0	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	Sirosis
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Sirosis
0	1	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	Sirosis
0	1	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	Sirosis

Data pada Tabel 4 kemudian diolah juga untuk data testing. Total data yang dihasilkan berjumlah 5 data. Tabel data testing dapat dilihat

Penyelesaian likehood untuk A5 (Diare) dapat dilihat pada Tabel 9.

Tabel 9. Tabel Hasil Perhitungan $D(x|c)$ A5

	AH	H	S	P(AH)	P(H)	P(S)
Iya	2	4	2	0,153846	0,363636	0,095238
Tidak	11	7	19	0,846154	0,636364	0,904762
Total	13	11	21			

3.4 Memperhitungkan Metode Naive Bayes

Probabilitas berdasarkan kondisi pada hipotesis likehood didapatkan menggunakan rumus $D(x|c)$. Perhitungan yang dilakukan sebanyak 18 atribut. Namun pada artikel ini, penulis hanya melakukan perhitungan pada 5 atribut.

a. A1 (BAB Kencur)

Penyelesaian likehood untuk A1 (BAB Kencur) dapat dilihat pada Tabel 5.

Tabel 5. Tabel Hasil Perhitungan $D(x|c)$ A1

	AH	H	S	P(AH)	P(H)	P(S)
Iya	1	0	2	0,076923	0	0,095238
Tidak	12	11	19	0,923077	1	0,904762
Total	13	11	21			

b. A2 (BAB Warna Hitam)

Penyelesaian likehood untuk A2 (BAB Warna Hitam) dapat dilihat pada Tabel 6.

Tabel 6. Tabel Hasil Perhitungan $D(x|c)$ A2

	AH	H	S	P(AH)	P(H)	P(S)
Iya	1	1	8	0,076923	0,090909	0,380952
Tidak	12	10	13	0,923077	0,909091	0,619048
Total	13	11	21			

c. A3 (Batuk Berdahak)

Penyelesaian likehood untuk A3 (Batuk Berdahak) dapat dilihat pada Tabel 7.

Tabel 7. Tabel Hasil Perhitungan $D(x|c)$ A3

	AH	H	S	P(AH)	P(H)	P(S)
Iya	0	0	3	0	0	0,142857
Tidak	13	11	18	1	1	0,857143
Total	13	11	21			

d. A4 (Dada Berdebar-debar)

Penyelesaian likehood untuk A4 (Dada Berdebar-debar) dapat dilihat pada Tabel 8.

Tabel 8. Tabel Hasil Perhitungan $D(x|c)$ A4

	AH	H	S	P(AH)	P(H)	P(S)
Iya	1	0	1	0,076923	0	0,047619
Tidak	12	11	20	0,923077	1	0,952381
Total	13	11	21			

e. A5 (Diare)

1.5 Menghitung Jumlah Class atau Label

Menghitung jumlah class atau label adalah tahapan pertama pada perhitungan Naive Bayes. Menghitung jumlah class dilakukan dengan cara menjumlahkan data sesuai dengan masing-masing class, lalu dibagi dengan keseluruhan jumlah data training. Hasil perhitungannya dapat dilihat pada Tabel 10.

Tabel 10. Tabel Menghitung Jumlah Class atau Label

Label	Jml Data	Jml Seluruh data	Hasil
Abses Hati	13	45	0,288889
Hepatitis	11	45	0,244444
Sirosis	21	45	0,466667

Hasil perhitungan nilai akhir dari setiap class Diagnosa Penyakit dapat dilihat pada Tabel 11.

Tabel 2. Tabel Rekapitulasi Hasil Perhitungan

No.	Pasien	Abses Hati	Hepatitis	Sirosis
1.	Pasien 46	0,012557	0,006566	0,001547
2.	Pasien 47	0,018602	0	0
3.	Pasien 48	0,008878	0,014071	0,001173
4.	Pasien 49	0,000273	0	0
5.	Pasien 50	0,007610	0,016885	0,000692

Berdasarkan perhitungan akhir dengan mengalikan nilai peluang dari kasus yang di angkat, kita melihat bahwa nilai $D(X|Diagnosa \text{ Penyakit}) = \text{"Abses Hati"}$ memperoleh nilai paling tinggi 0,012557 dibanding dengan class yang lain. Hasil perbandingan class dapat dilihat pada Tabel 12.

Tabel 12. Tabel Hasil Perbandingan Class

No.	Pasien	Diagnosa Dokter	Diagnosa Sistem
1.	Pasien 46	Abses Hati	Abses Hati
2.	Pasien 47	Abses Hati	Abses Hati
3.	Pasien 48	Abses Hati	Hepatitis
4.	Pasien 49	Abses Hati	Abses Hati
5.	Pasien 50	Abses Hati	Hepatitis

4. Kesimpulan

Berdasarkan hasil pengujian dari data 50 rekam medis pasien penyakit hati, mendapatkan dari metode naive bayes yang d aplikasikan pada data mining berguna untuk membandingkan antara data RSUD dengan data sistem. Dari 45 data training dan 5 data testing didapat ingkat akurasi sebesar 60 % menggambarkan bahwa sistem ini layak.

Daftar Rujukan

- [1] Panggabean, D. S. O., Buulolo, E., & Silalahi, N. (2020). Penerapan Data Mining Untuk Memprediksi Pemesanan Bibit Pohon Dengan Regresi Linear Berganda. *JURIKOM (Jurnal Riset Komputer)*, 7(1), 56-62. DOI: <http://dx.doi.org/10.30865/jurikom.v7i1.1947>
- [2] Takdirillah, R. (2020). Penerapan Data Mining Menggunakan Algoritma Apriori Terhadap Data Transaksi Penjualan Bisnis Ritel. *Edumatic: Jurnal Pendidikan Informatika*, 4(1), 37-46. DOI: <https://doi.org/10.29408/edumatic.v4i1.2081>
- [3] Khoirunnisa, N., Nugroho, I. M. R., & Muhyidin, Y. (2022). Penerapan data mining dengan metode naive bayes untuk memprediksi calon terjamin potensial. *Journal of Information System, Applied, Management, Accounting and Research*, 6(1), 149-161. DOI: <https://doi.org/10.52362/jisamar.v6i1.684>
- [4] Br. Saragih, N. (2022). Sistem Pakar Mendiagnosa Penyakit Gangguan Hati Pada Manusia Menggunakan Metode Naive Bayes Berbasis WEB. *Jurnal Ilmu Komputer Dan Sistem Informasi (JIKOMSI)*, 5(1), 11-19. DOI: <https://doi.org/10.55338/jikomsi.v5i1.202>
- [5] Imawati, I., Widiyanto, K., Aziz, F., Rifai, A., & Rahmawati, A. (2022). Implementasi artificial neural network dalam mendeteksi penyakit hati (liver). *Journal of Information System, Applied, Management, Accounting and Research*, 6(1), 193-198. DOI: <https://doi.org/10.52362/jisamar.v6i1.694>
- [6] Setiawan, S. I., & Kurniawan, J. (2021). Pilihan Tatalaksana Penyakit Perlemakan Hati Non-Alkohol (Non-Alcoholic Fatty Liver Disease/NAFLD). *Cermin Dunia Kedokteran*, 48(3), 173-175. DOI: <http://dx.doi.org/10.55175/cdk.v48i3.1336>
- [7] D Damuri, A., Riyanto, U., Rusdianto, H., & Aminudin, M. (2021). Implementasi Data Mining dengan Algoritma Naive Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako. *JURIKOM (Jurnal Riset Komputer)*, 8(6), 219-225. DOI: <http://dx.doi.org/10.30865/jurikom.v8i6.3655>
- [8] Silahudin, D., Henderi, & Holidin, A. (2020). Model expert system for diagnosis of COVID-19 using naive bayes classifier. *IOP Conference Series: Materials Science and Engineering*, 1007(1). DOI: <https://doi.org/10.1088/1757-899X/1007/1/012067>
- [9] Doni, B. T. R., Susanti, S., & Mubarak, A. (2021). Penerapan Data Mining Untuk Klasifikasi Penyakit Hepatocellular Carcinoma Menggunakan Algoritma Naive Bayes. *Jurnal Responsif: Riset Sains dan Informatika*, 3(1), 12-19. <https://doi.org/10.51977/jti.v3i1.403>
- [10] Zohra, F. T. (2020). Prediction of Different Diseases and Development of a Clinical Decision Support System using Naive Bayes Classifier. *International Journal for Research in Applied Science and Engineering Technology*, 8(5), 8-13. <http://doi.org/10.22214/ijraset.2020.5002>
- [11] Irmayani, W. (2021). VISUALISASI DATA PADA DATA MINING MENGGUNAKAN METODE KLASIFIKASI NAIVE BAYES. *Jurnal Khatulistiwa Informatika*, 9(1). DOI: <https://doi.org/10.31294/jki.v9i1.9593>
- [12] Indraputra, R. A., & Fitriana, R. (2020). K-Means Clustering Data COVID-19. *Jurnal Teknik Industri*, 10(3), 275-282. DOI: <https://doi.org/10.25105/jti.v10i3.8428>
- [13] Titimeidara, M. Y., & Hadikurniawati, W. (2021). Implementasi Metode Naive Bayes Classifier Untuk Klasifikasi Status Gizi Stunting Pada Balita. *JURNAL ILMIAH INFORMATIKA*, 9(01), 54-59. DOI: <https://doi.org/10.33884/jif.v9i01.3741>
- [14] Alfiah, N. (2021). Klasifikasi Penerima Bantuan Sosial Program Keluarga Harapan Menggunakan Metode Naive Bayes. *Respati*, 16(1), 32-40. DOI: <https://doi.org/10.35842/jtir.v16i1.386>
- [15] Dwiramadhan, F., Wahyuddin, M. I., & Hidayatullah, D. (2022). Sistem Pakar Diagnosa Penyakit Kulit Kucing Menggunakan Metode Naive Bayes Berbasis Web. *Jurnal JTIK (Jurnal Teknologi Informasi Dan Komunikasi)*, 6(3), 429-437. <https://doi:10.35870/jtik.v6i3.466>
- [16] Putra, D. W. T., Utami, A. O., Minarni, & Swara, G. Y. (2019). Accuracy Level of Diagnosis of ENT Diseases in Expert System. *Jurnal KomtekInfo*, 6(2), 127-134. <https://doi:10.35134/komtekinfo.v6i2.51>