

Comparison of the Results of the K-Nearest Neighbor (KNN) and Naïve Bayes Methods in the Classification of ARI Diseases (Case Study: RSUD Fauziah Bireuen)

Riska Yolanda Putri^{1*}, Zara Yunizar², Safwandi³

^{1,2,3} Universitas Malikussaleh, Indonesia

*Corresponding Author Email: riska.190170059@mhs.unimal.ac.id

ABSTRAK

Received: 29 December 2023
Revised: 31 December 2023
Accepted: 31 December 2023
Available online: 1 January 2024

Kata Kunci:

Klasifikasi, K-Nearest Neighbors (KNN), Naïve Bayes, Infeksi Saluran Pernapasan Akut (ISPA)

Infeksi Saluran Pernapasan Akut atau yang biasa disebut ISPA adalah salah satu penyakit yang disebabkan oleh bakteri atau virus. ISPA bisa menyerang semua umur terutama pada anak – anak. Penelitian ini bertujuan untuk melakukan perbandingan tingkat akurasi klasifikasi pada penyakit ISPA. Data yang digunakan yaitu data dari pasien yang terkena penyakit ISPA yang ada di RSUD Fauziah Bireuen. *K-Nearest Neighbors* dan juga *Naïve Bayes* dapat digunakan dalam pengklasifikasian penyakit ISPA. Pengukuran akurasi menggunakan *Confusion Matrix* pada metode *K-Nearest Neighbors* dengan pendekatan *Eulidean Distance* pada kasus klasifikasi penyakit ISPA didapatkan persentase *precision* sebesar 91%, *recall* 84% dan *accuracy* sebesar 88%. Sedangkan pada metode *Naïve Bayes* didapatkan persentase *precision* sebesar 95%, *recall* 78% dan *accuracy* sebesar 86%. Hasil perbandingan akurasi dari kedua metode tersebut menunjukkan bahwa metode *K-Nearest Neighbors* memiliki tingkat akurasi yang lebih tinggi dibandingkan dengan metode *Naïve Bayes*.

ABSTRACT

Keywords:

Classification, K-Nearest Neighbors (KNN), Naïve Bayes, Acute Respiratory Infection (ARI)

Acute Respiratory Infection or commonly called (ARI) is a disease caused by bacteria or viruses. (ARI) can attack all ages, especially children. This study aims to compare the accuracy of classification in (ARI) disease. The data used is data from patients affected by (ARI) disease at Fauziah Bireuen Hospital. K-Nearest Neighbors and Naïve Bayes can be used in the classification of (ARI) diseases. Measurement of accuracy using Confusion Matrix in the K-Nearest Neighbors method with the Eulidean Distance approach in the case of (ARI) disease classification obtained a percentage of precision of 91%, recall 84% and accuracy of 88%. While the Naïve Bayes method obtained a percentage of precision of 95%, recall 78% and accuracy of 86%. The results of the accuracy comparison of the two methods show that the K-Nearest Neighbors method has a higher accuracy rate than the Naïve Bayes method.

1. PENDAHULUAN

Penyakit Infeksi Saluran Pernapasan Akut (ISPA) merupakan penyakit saluran pernapasan yang paling banyak menular kepada orang yang memiliki sistem kekebalan tubuh rendah seperti orang lanjut usia dan anak - anak yang belum memiliki sistem kekebalan tubuh terbentuk sepenuhnya. Penyakit ini biasanya disebabkan oleh virus atau bakteri yang menyerang hidung atau trakea (saluran pernapasan) sehingga menyebabkan fungsi pernapasan menjadi terganggu.

Jenis penyakit ISPA dapat diklasifikasikan kedalam dua katagori yaitu ISPA ringan dan berat. Masing - masing ISPA tersebut memerlukan penanganan yang berbeda - beda. RSUD Fauziah Bireuen merupakan salah satu pusat pelayanan kesehatan masyarakat yang menjadi tempat pengobatan bagi masyarakat yang memiliki keluhan terhadap kesehatan terutama penyakit ISPA, RSUD Fauziah Bireuen harus dapat mengidentifikasi jenis penyakit ISPA yang tepat agar penanganan kepada penderita ISPA dapat diberikan secara tepat dan maksimal. RSUD Fauziah Bireuen membutuhkan informasi tentang jenis klasifikasi penyakit ISPA yang diderita

oleh masyarakat sehingga dapat membantu pihak rumah sakit dalam memberikan pengobatan yang tepat kepada penderita ISPA sesuai dengan jenis ISPA yang diderita, sehingga dibutuhkan sebuah sistem yang mampu memberikan informasi tersebut kepada masyarakat dan juga dapat membantu pihak RSUD Fauziah Bireuen untuk mendiagnosis penyakit ISPA.

Ada berbagai metode data mining yang dapat digunakan untuk mengklasifikasi penyakit ISPA di antaranya yaitu *K-Nearest Neighbor* (KNN) dan *Naive Bayes*. *K-Nearest Neighbors* (KNN) merupakan metode klasifikasi dataset berdasarkan data learning. *K-Nearest Neighbors* (KNN) adalah algoritma yang mengklasifikasikan data berdasarkan seberapa dekat (jarak) dengan data lain. Adapun metode *Naive Bayes* yaitu metode klasifikasi yang menggunakan teori *Bayes* Probabilitas untuk memprediksi kelas yang tidak diketahui. Pada *Naive Bayes* dataset di anggap paling independent.

Sebelumnya, Rahman et al. (2018) didalam jurnal telah melakukan penelitian yaitu melakukan komparasi antara algoritma *K-Nearest Neighbor* (KNN) dengan *Naive Bayes* untuk mengklasifikasi kualitas air bersih. Hasilnya *K-Nearest Neighbor* (KNN) memiliki nilai rata-rata akurasi lebih tinggi,

yaitu sebesar 82,42%, sedangkan *Naive Bayes* hanya mendapatkan nilai rata-rata akurasi sebesar 70,32%. Evaluasi tingkat keberhasilan akan didasarkan pada nilai akurasi, baik menggunakan algoritma *K-Nearest Neighbor* (KNN) maupun *Naive Bayes*, sehingga hasil yang didapat akan dibandingkan.

2. RESEARCH METHODS

2.1 Data Mining

Data mining merupakan proses pencarian atau pencarian secara otomatis demi memperoleh informasi penting dalam gudang data yang sangat besar. Data mining adalah serangkaian proses untuk menggali nilai tambahan berupa informasi yang selama ini tidak diketahui secara manual dari sebuah basis data dengan melakukan penggalian pola dari data dengan tujuan untuk memanipulasi data menjadi sebuah informasi yang lebih berharga yang diperoleh dengan cara mengekstraksi dan mengenali pola yang penting atau menarik dari data yang dapat dalam basis data.

2.2 Algoritma *K-Nearest Neighbor* (KNN)

K-Nearest Neighbor adalah metode pengambilan keputusan dengan menggunakan supervised learning dimana hasil dari data input baru diklasifikasikan berdasarkan data nilai. Algoritma *K-Nearest Neighbor* (KNN) merupakan metode untuk mengklasifikasikan objek berdasarkan data pembelajaran yang paling dekat dengan objek tersebut. KNN adalah algoritma pembelajaran terawasi dimana hasil dari *instance query* baru diklasifikasikan berdasarkan mayoritas kategori dalam algoritma KNN. Dimanakah kelas yang paling banyak muncul yang nantinya akan menjadi kelas hasil klasifikasi.

2.2 Algoritma *Naive Bayes*

Algoritma *Naive Bayes* adalah metode klasifikasi menggunakan metode probabilitas dan statistik yang diusulkan oleh seorang ilmuwan Inggris bernama Thomas Bayes. Algoritma *Naive Bayes* dapat memprediksi peluang masa depan berdasarkan pengalaman masa lalu sehingga dikenal dengan *Teorema Bayes*. *Naive Bayes* menggunakan cabang matematika yang dikenal dengan teori probabilitas untuk mencari peluang terbesar dari kemungkinan klasifikasi, dengan cara melihat frekuensi tiap klasifikasi pada data training. Klasifikasi *Naive Bayes* adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu *class*.

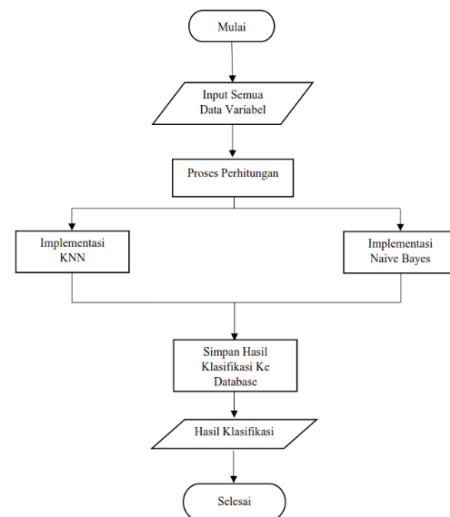
Naive Bayes adalah sebuah metode pengelompokan statistik yang bisa dipakai untuk memprediksi probabilitas anggota suatu *class*. *Naive Bayes* juga mempunyai akurasi dan kecepatan yang sangat kuat ketika diaplikasikan pada *database* dengan *big data*.

3. RESULT AND DISCUSSION

Perancangan sistem yakni rangkaian tindakan yang mendetail tentang cara sistem akan berfungsi. Tujuannya untuk menciptakan suatu perangkat lunak disesuaikan dengan kebutuhan dari pengguna. Pada proses merancang sistem, peneliti memakai berbagai jenis diagram UML guna mendapatkan gambaran menyeluruh tentang sistem yang

dikembangkan. Jenis-jenis diagram UML meliputi *use case diagram*, *sequence diagram*, dan *activity diagram*.

3.1 Skema Sistem



Gambar 1. Skema Sistem

Pada penelitian ini, penulis akan menguji algoritma *K-Nearest Neighbor* dan *Naive Bayes* untuk mengklasifikasikan penyakit ISPA pada RSUD Fauziah Bireuen. Algoritma ini akan mengklasifikasikan penyakit ISPA berdasarkan variable yang telah ditetapkan oleh admin.

3.2 Perhitungan Manual Algoritma KNN

Tabel 1. Tabel Kriteria

Kode Kriteria	Kriteria
X1	Pilek
X2	Demam
X3	Sesak
X4	Nyeri Dada
X5	Batuk
X6	Kondisi Tubuh
X7	Mual/Muntah

Selanjutnya menentukan nilai bobot dari setiap kriteria. Tabel 1 menunjukkan kriteria apa saja yang akan dibobotkan dan dapat dilihat pada tabel dibawah ini:

Tabel 2. Tabel Sub Kriteria

Atribut	Value	Bobot
Pilek	Tidak	1
	Ya	2
Demam	Demam	1
	Demam Tinggi	2
Sesak	Tidak	1
	Ya	2
Nyeri Dada	Tidak	1
	Ya	2
Batuk	Batuk	1
	Batuk Berat	2

Kondisi Tubuh	Mudah Lelah	1
	Lemas, Lesu	2
Mual/Muntah	Tidak	1
	Ya	2

Tabel 3. Tabel Normalisasi

Klasifikasi	Bobot
ISPA Ringan	1
ISPA Berat	2

Tujuan dari algoritma K-Nearest Neighbors adalah untuk mengklasifikasikan objek baru berdasarkan data latih yang ada. Dalam proses ini, perhitungan jarak antara dua titik, yaitu titik data latih dan titik data uji, dapat menggunakan rumus Euclidean Distance seperti yang dijelaskan berikut ini:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2 + (x_j - y_j)^2 + \dots} \quad (1)$$

Keterangan:

- x_i : Sampel Data
- y_i : Data *Testing*
- i : Variabel/Kriteria Data
- $d(x, y)$: Jarak
- n : Dimensi Data

Setelah mendapatkan hasil dari perhitungan algoritma k-nearest neighbor selanjutnya menghitung nilai akurasi dengan menggunakan Confusion Matrix yaitu mencari nilai precision, recall dan accuracy berdasarkan nilai TP, FP, FN dan TN yang sudah diperoleh sebelumnya dari penggunaan nilai $k = 3$. Berikut merupakan rumus Confusion Matrix:

1. Presisi (*Precision*)

Untuk mengukur tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem:

$$Precision = \frac{TP}{(TP+FP)} \quad (2)$$

2. Recall

Untuk mengukur tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi pada persamaan.

$$Recall = \frac{TP}{(TP+FN)} \quad (3)$$

3. Akurasi (*Accuracy*)

Untuk mengukur kinerja sebuah metode.

$$Accuracy = \frac{TP+TN}{(TP+TN+FP+FN)} \quad (4)$$

Tabel 4. *Predicted Class K = 3*

Class Prediksi	ISPA Ringan	ISPA Berat
ISPA Ringan	21	4
ISPA Berat	2	23

Tabel 5. Nilai *Precision*, *Recall* dan *Accuracy*

	<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>
K3	0,91	0,84	0,88
Persentase	91%	84%	88%

3.3 Perhitungan Manual Algoritma *Naïve Bayes*

Cara kerja dari proses perhitungan metode *Naïve Bayes* adalah menentukan data – data yang akan digunakan. Data yang dibutuhkan memiliki beberapa variabel penentu yang digunakan:

Tabel 6. Tabel Keterangan Data ISPA

ISPA Ringan	108
ISPA Berat	92

Langkah awal pada algoritma *naïve bayes* adalah dengan menghitung probabilitas setiap kriteria yang ada pada data:

Tabel 7. Tabel Keterangan Kriteria Data ISPA

Keterangan	ISPA Ringan	ISPA Berat
Pilek (Tidak)	34	55
Pilek (Ya)	74	37
Demam (Demam)	107	0
Demam (Demam Tinggi)	1	92
Sesak (Tidak)	107	0
Sesak (Ya)	1	92
Nyeri Dada (Tidak)	21	75
Nyeri Dada (Ya)	87	17
Batuk (Batuk)	107	13
Batuk (Batuk Berat)	1	79
Kondisi Tubuh (Mudah Lelah)	66	0
Kondisi Tubuh (Lemas, Lesu)	42	92
Mual/Muntah (Tidak)	107	34
Mual/Muntah (Ya)	1	68

Setelah menentukan kriteria, selanjutnya menghitung probabilitas dari atribut gejala pada ISPA tersebut. Bentuk umum dari teorema bayes seperti dibawah ini:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (5)$$

Keterangan:

- X : Data dengan *class* yang belum diketahui
- H : Hipotesis data X merupakan suatu *class* spesifik
- $P(H|X)$: Probabilitas hipotesis H berdasarkan kondisi X
- $P(H)$: Probabilitas hipotesis H (*prior probability*)
- $P(X|H)$: Probabilitas X berdasarkan kondisi pada hipotesis H
- $P(X)$: Probabilitas dari X

Setelah mendapatkan hasil dari perhitungan algoritma *naïve bayes* selanjutnya menghitung nilai akurasi dengan menggunakan Confusion Matrix yaitu mencari nilai precision, recall dan accuracy berdasarkan nilai TP, FP, FN dan TN yang sudah diperoleh sebelumnya dari hasil perhitungan probabilitas setiap data:

Tabel 8. *Predicted Class Naïve Bayes*

Class Prediksi	ISPA Ringan	ISPA Berat
ISPA Ringan	22	6
ISPA Berat	1	21

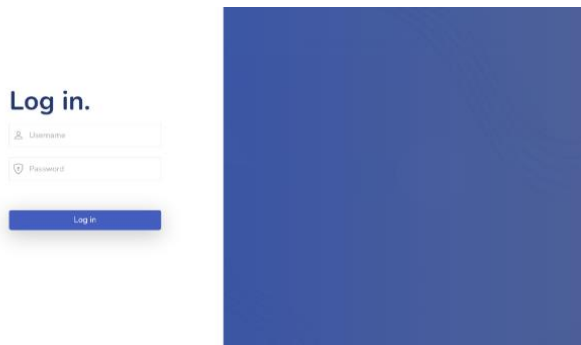
Tabel 9. Nilai *Precision*, *Recall* dan *Accuracy*

	<i>Precision</i>	<i>Recall</i>	<i>Accuracy</i>
	0,95	0,78	0,86
Persentase	95%	78%	86%

3.4 Implementasi Sistem

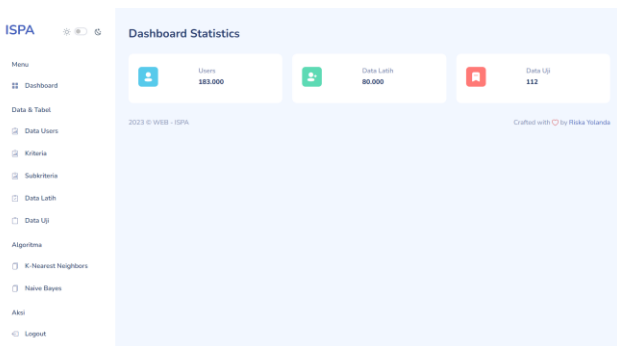
Implementasi adalah proses penerapan sistem dalam penyelesaian masalah yang ada. Implementasi bertujuan untuk mengetahui apakah sistem tersebut berjalan baik atau masih ada hal yang harus diperbaiki lagi:

1. Halaman Login



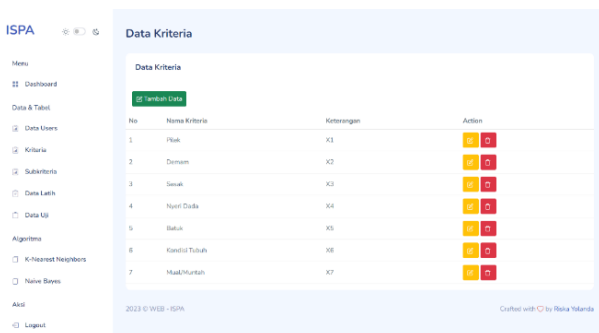
Gambar 2. Halaman Login

2. Halaman Menu Utama



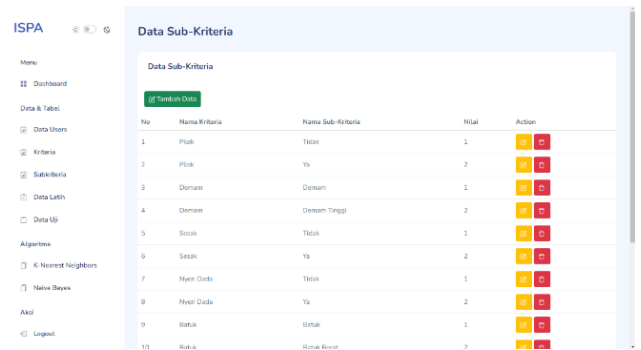
Gambar 3. Halaman Menu Utama

3. Tampilan Halaman Data Kriteria



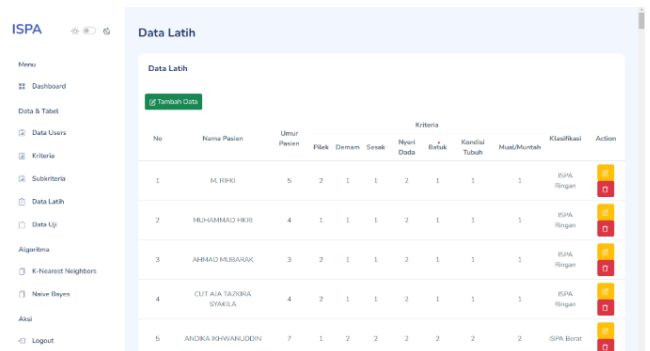
Gambar 4. Tampilan Halaman Data Kriteria

4. Tampilan Halaman Data SubKriteria



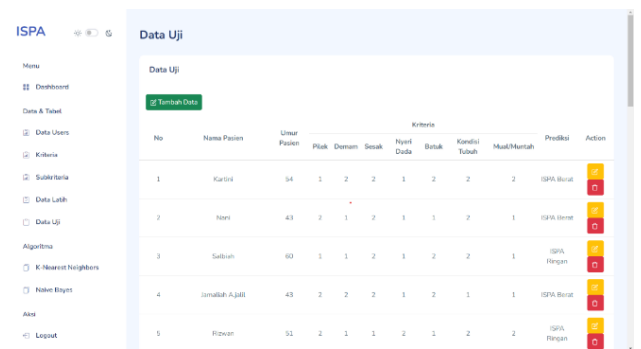
Gambar 5. Tampilan Halaman Data SubKriteria

5. Tampilan Halaman Data Latih



Gambar 6. Tampilan Halaman Data Latih

6. Tampilan Halaman Data Uji



Gambar 7. Tampilan Halaman Data Uji

7. Tampilan Halaman Hasil Klasifikasi Algoritma K-Nearest Neighbors



Gambar 8. Tampilan Halaman Hasil Klasifikasi Algoritma K-Nearest Neighbors

8. Tampilan Halaman Hasil Klasifikasi Algoritma *Naïve Bayes*

No	Nama Pasien	Kriteria							Prediksi	Prediksi Naive Bayes
		Pilek	Demam	Sesak	Nyeri Dada	Batuk	Kondisi Tubuh	Mual/Muntah		
1	Karlisi	1	2	2	1	2	2	2	ISPA Berat	ISPA Berat
2	Nani	2	1	2	1	1	2	1	ISPA Berat	ISPA Ringan
3	Sabbiah	1	2	2	1	2	2	1	ISPA Ringan	ISPA Berat
4	Jamalah Azali	2	2	2	1	2	1	1	ISPA Berat	ISPA Ringan
5	Rizwan	2	1	1	2	1	1	2	ISPA Ringan	ISPA Ringan
6	Nur Azqah Sabi	1	2	2	2	1	2	1	ISPA Berat	ISPA Berat
7	Muklis	1	2	1	2	1	1	2	ISPA Berat	ISPA Ringan
8	Anwarah	2	2	1	2	1	1	2	ISPA Ringan	ISPA Ringan
9	Andhyan	2	1	2	1	1	1	1	ISPA Ringan	ISPA Ringan
10	Irene Dadman	2	2	2	1	2	2	2	ISPA Berat	ISPA Berat

Gambar 9. Tampilan Halaman Hasil Klasifikasi Algoritma *Naïve Bayes*

4. CONCLUSION

Sistem klasifikasi penyakit ISPA ini dirancang menggunakan UML (Unified Modelling Language), dan dibangun dengan bahasa pemrograman PHP dan Database MYSQL.

Dalam penerapannya sistem klasifikasi penyakit ISPA ini dibuat dengan menggunakan metode K-Nearest Neighbors dan Naïve Bayes menggunakan 7 (tujuh) kriteria yaitu pilek, demam, sesak nyeri dada, batuk, kondisi, tubuh dan mual/muntah. Perbandingan nilai akurasi yang diperoleh menunjukkan bahwa algoritma K-Nearest Neighbor lebih baik dibandingkan dengan metode Naïve Bayes dalam mengklasifikasi penyakit ISPA.

REFERENCES

[1] Arifin, Z. (2019). Penerapan Metode Knn (K-Nearest Neighbor) Dalam Sistem Pendukung Keputusan Penerimaan Kip (Kartu Indonesia Pintar) Di Desa Pandean Berbasis Web Dan Mysql. *NJCA (Nusantara Journal of Computers and Its Applications)*, 4(1). <https://doi.org/10.36564/njca.v4i1.101>

[2] Bari, M., Sitorus, S. H., Ristian, U., Rekayasa, J., Komputer, S., Mipa, F., Tanjungpura, U., Prof, J., Hadari, H., & Pontianak, N. (2018). Penyebaran Wabah Penyakit ISPA (Studi Kasus: Wilayah Kota Pontianak).

In Jurnal Coding, Sistem Komputer Untan (Vol. 06, Issue 03).

[3] Darnila, E., Maryana, M., & Azmi, M. (2021). Aplikasi Klasifikasi Status Gizi Balita Menggunakan Metode Naive Bayes Berbasis Android. *METHOMIKA Jurnal Manajemen Informatika Dan Komputerisasi Akuntansi*, 5(2),135–141. <https://doi.org/10.46880/jmika.Vol5No2.pp135-141>

[4] Dinata, R.K., Adek, R.T., Hasdyna, N., Retno, S. (2023). K-nearest neighbor classifier optimization using purity. *AIP Conference Proceedings*. 2431(1).

[5] Hua, W., Xiaofeng, L., Zhenqiang, B., Jun, R., Ban, W., & Liming, L. (2020). Consideration on the strategies during epidemic stage changing from emergency response to continuous prevention and control. *Chinese Journal of Endemiology*, 41(2), 297–300. <https://doi.org/10.3760/cma.j.issn.0254-6450.2020.02.003>

[6] Hasdyna, N., Retno, S. (2022). Machine Learning Approach to Determine the Drug-Prone Areas in Lhokseumawe City, Indonesia. *International Journal of Multidisciplinary Research and Analysis*. 5(9): 2354-2464.

[7] Prayoga Permana, A., Ainiyah, K., & Fahmi Hayati Holle, K. (2021). Analisis Perbandingan Algoritma Decision Tree, kNN, dan Naive Bayes untuk Prediksi Kesuksesan Start-up. In *JISKa (Vol. 6, Issue 3)*. <https://www.kaggle.com/manishkc06/startup-success-prediction>.

[8] Retno, S., Dinata, R.K., Hasdyna, N. (2023). Evaluasi model data chatbot dalam natural language processing menggunakan k-nearest neighbor. *Jurnal CoSciTech (Computer Science and Information Technology)*. 4(1): 146-153.

[9] Susanti, E., Al-Kautsar Aidilof, H., & Priyanto, D. (2022). Comparison of Naive Bayes and Dempster Shafer Methods in Expert System for Early Diagnosis of COVID-19. *Teknik Informatika Dan Rekayasa Komputer*, 22(1), 217–230. <https://doi.org/10.30812/matrik.v22i1.22>

[10] Yanosma, D., Johar T, A., & Anggriani, K. (2016). Implementasi Metode K-Nearest Neighbor (KNN) Dan Simple Additive Weighting (SAW) Dalam Pengambilan Keputusan Seleksi Penerimaan Anggota Paskibraka (Studi Kasus: Dinas Pemuda Dan Olahraga Provinsi Bengkulu). In *Jurnal Rekursif (Vol. 4, Issue 2)*.