



INNOVATIVE: Journal Of Social Science Research

Volume 5 Nomor 1 Tahun 2024 Page 5311-5323

E-ISSN 2807-4238 and P-ISSN 2807-4246

Website: <https://j-innovative.org/index.php/Innovative>

Analisa Prediksi Penyakit Diabetes Menggunakan Metode Naive Bayes dan K-NN

Nugraha Rahmansyah^{1✉}, Shary Armonitha Lusia², Ilmawati³

Teknik informatika, Universitas Putra Indonesia YPTK

Jln. Raya Lubuk Begalung, Padang, Sumatera Barat

Email: nugraharahmansyah@upiyptk.ac.id^{1✉}

Abstrak

Diabetes mellitus adalah salah satu penyakit kronis yang semakin umum di dunia termasuk di Indonesia. Mendeteksi penyakit ini sejak dini sangatlah penting untuk mencegah munculnya komplikasi serius. Penelitian ini dilakukan untuk menganalisis prediksi penyakit diabetes menggunakan dua metode pembelajaran mesin, yaitu Naive Bayes dan K-Nearest Neighbor (KNN). Naive Bayes dikenal sebagai metode klasifikasi yang sederhana namun efektif berdasarkan probabilitas, sementara KNN merupakan metode klasifikasi berbasis instance yang menggunakan kedekatan data dengan data yang sudah ada. Dataset yang digunakan mencakup 8 atribut kesehatan, seperti usia, jenis kelamin, hipertensi, penyakit jantung, riwayat merokok, BMI, kadar HbA1c, dan kadar glukosa darah, dengan label target berupa status diabetes (1 untuk diabetes, 0 untuk non-diabetes). Tahapan tersebut mencakup pengolahan awal data, seperti mengisi nilai kosong, menormalisasi atribut numerik, dan mengkode atribut kategori. Naive Bayes memanfaatkan distribusi probabilitas, sedangkan KNN mengelompokkan data berdasarkan jarak ke tetangga terdekat. KNN lebih baik digunakan untuk dataset dengan pola yang rumit, sementara Naive Bayes lebih efisien dalam hal komputasi

Kata Kunci : *Prediksi, Diabetes, Naive Bayes, KNN.*

Abstract

Diabetes mellitus is a chronic disease that is becoming increasingly common worldwide, including in Indonesia. Early detection of this disease is crucial to prevent the onset of serious complications. This study aims to analyze diabetes prediction using two machine learning methods, namely Naïve Bayes and K-Nearest Neighbor (KNN). Naïve Bayes is known as a simple yet effective classification method based on probability, while KNN is an instance-based classification method that uses data proximity to existing data. The dataset used includes 8 health attributes such as age, gender, hypertension, heart disease, smoking history, BMI, HbA1c levels, and blood glucose levels, with a target label representing diabetes status (1 for diabetes, 0 for non-diabetes). The process includes data preprocessing steps, such as handling missing values, normalizing numerical attributes, and encoding categorical attributes. Naïve Bayes leverages probability distributions, while KNN groups data based on the distance to the nearest neighbors. KNN performs better for datasets with complex patterns, while Naïve Bayes is more computationally efficient

Key word : *Prediction, Diabetes, Naive Bayes, KNN.*

PENDAHULUAN

Diabetes melitus, atau yang sering disebut sebagai penyakit kencing manis oleh masyarakat umum, adalah penyakit kronis yang dapat bertahan seumur hidup. Diabetes terdiri dari dua tipe, yaitu diabetes melitus tipe 1 yang disebabkan oleh reaksi autoimun terhadap protein sel pulau pankreas, dan diabetes tipe 2 yang disebabkan oleh faktor genetik yang terkait dengan gangguan sekresi insulin, resistensi insulin, serta faktor lingkungan seperti obesitas, pola makan yang tidak sehat, kurangnya aktivitas fisik, stres, dan penuaan (Lestari et al., 2021). Ketidakstabilan kadar gula darah, kurang berolahraga, dan kebiasaan makan yang tidak terjaga semuanya merupakan penyebab diabetes, dan semua ini berdampak negatif pada kualitas hidup karena sulit untuk mempertahankan pola makan yang seimbang. Dari lima pilar pengobatan diabetes, dua di antaranya adalah aktivitas fisik dan pola makan. Cara untuk menurunkan kadar gula darah adalah dengan meningkatkan aktivitas fisik dan mengubah pola makan (Suryawan et al., 2023).

Hal ini senada dengan kesepakatan Perhimpunan Endokrinologi Indonesia (PERKENI, 2011), strategi utama dalam mengontrol diabetes mellitus mencakup aktivitas fisik, pengaturan nutrisi medis, penanganan obat, dan penyuluhan. Salah satu faktor penting dalam menentukan keberhasilan pengendalian penyakit diabetes melitus adalah tingkat kepatuhan pasien dalam menjalani pola makan atau diet sehari-hari (Petersmann et al., 2018). Secara prinsipnya penderita Diabetes melitus memerlukan pengaturan dalam pola

makan. Ini bisa dilakukan dengan memperhatikan total kalori dan jenis bahan yang diperlukan untuk nutrisi makanan dan pola makan teratur (Astutisari et al., 2022).

Metode Naïve Bayes dan K-NN merupakan beberapa metode machine learning dalam pengklasifikasian dan prediksi suatu kondisi. Naïve Bayes adalah suatu metode statistik untuk memperkirakan probabilitas dari sebuah kelas tertentu, menghitung kemungkinan dari sebuah hipotesis, menghitung probabilitas dari sebuah kelas dalam setiap kelompok atribut, serta menentukan kelas yang paling optimal (Syarli & Muin, 2020). Pendekatan Naive Bayes memiliki keunggulan karena menggunakan jumlah data training yang sedikit untuk memperoleh estimasi parameter yang dibutuhkan dalam proses klasifikasi. Itulah sebabnya, meskipun variabel independen yang dipertimbangkan, hanya variasi dari variabel kelas yang dibutuhkan untuk menentukan kategorisasi, bukan semua variabel kelas (Pebdika et al., 2023). Dalam statistik Naive Bayes, data dipasangkan dengan hasil yang mungkin dan parameter terhadap variabel acak. Teorema Bayes digunakan sebagai dasar dari metode Naïve Bayes (Murni Pratiwi et al., 2024).

Teorema Bayes memiliki bentuk umum sebagai berikut (Purwati & Dwi Januanti, 2021),

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

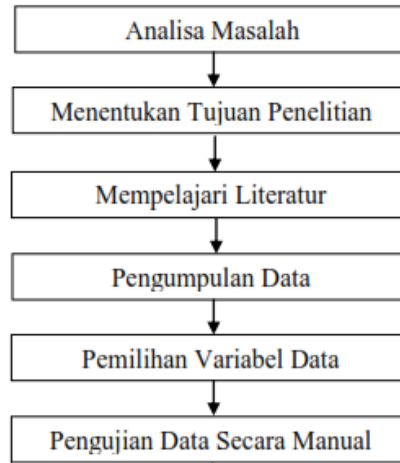
Dengan :

- X : Data dengan kelas yang belum diketahui
- H : Hipotesis data X merupakan suatu label kelas tertentu
- $P(H|X)$: Probabilistik hipotesis H berdasarkan kondisi X (posteriori probability)
- $P(H)$: Probabilistik hipotesis H (prior probability)
- $P(X|H)$: Probabilitas X berdasarkan kondisi pada hipotesis H
- $P(X)$: Probabilistik X

Algoritma KNN adalah sebuah teknik yang dipakai untuk mengklasifikasikan data berdasarkan jarak terdekat ke objek data (Cholil et al., 2021). Algoritma ini termasuk jenis algoritma *supervised learning* (Syihabuddin Azmil Umri, 2021). K-NN didasarkan pada konsep pembelajaran melalui analogi. Data belajar adalah data yang dijelaskan dengan atribut numerik n-dimensi. Setiap data pembelajaran mewakili sebuah titik, yang ditandai dengan c, dalam ruang ndimensi. Jika sebuah data query dengan label yang tidak diketahui diinputkan, maka KNearest Neighbor akan mencari k data pembelajaran yang memiliki jarak terdekat dengan data query dalam ruang n-dimensi (Ghani Muttaqin et al., 2020).

METODE PENELITIAN

Dalam menganalisa data, kita harus memikirkan tahapan atau langkah-langkah kerja. Dalam hal ini, penulis memaparkan beberapa tahap dalam menganalisa data yang dapat dilihat pada gambar 1 berikut ini :



Gambar 1. Metode Penelitian

Adapun data yang digunakan pada penelitian ini di ambil dari data set Kaggle sebanyak 200 data, dimana data set ini berisi *gender*, *age*, *hypertension*, *heart_disease*, *smoking_history*, *bmi*, *HbA1c_level*, *blood_glucose_level*, dan *diabetes*. Berikut data set yang digunakan yang dapat dilihat pada tabel dibawah ini :

Tabel 1. Data Set Diabetes (Data Training)

No	gender	age	hypertension	Heart disease	Smoking history	bmi	HbA1c level	Blood glucose level	diabetes
1	Female	48	0	0	Never	20.19	7	145	1
2	Female	80	0	0	Former	36.18	6.5	200	1
3	Male	28	0	0	Never	27.32	5.7	158	0
4	Female	42	0	0	Never	24.81	9	159	1
5	Male	76	1	1	Current	20.14	4.8	155	0
6	Female	20	0	0	Never	27.32	6.6	85	0
7	Female	44	0	0	Never	19.31	6.5	200	1
8	Female	79	0	0	No Info	23.86	5.7	85	0
9	Male	71	0	0	Never	27.09	8.2	200	1
10	Female	32	0	0	Never	27.32	5	100	0
11	Female	53	0	0	Never	27.32	6.1	85	0
12	Female	52	1	0	Never	50.3	6.6	155	1
13	Female	78	0	0	Former	36.05	5	130	0

14	Female	67	0	0	Never	25.69	5.8	200	0
15	Female	76	0	0	No Info	27.32	5	160	0
16	Male	80	0	1	Former	24.36	7.5	280	1
17	Male	15	0	0	Never	30.36	6.1	200	0
18	Female	42	0	0	Never	24.48	5.7	158	0
19	Female	42	0	0	No Info	27.32	5.7	80	0
20	Male	59	0	0	Current	29.2	8.2	220	1
...
180	Male	25	0	0	Never	27.32	6.1	140	0
181	Female	59	0	0	Never	25.8	5.7	90	0
182	Male	39	0	0	Never	27.32	5.7	159	0
183	Female	25	0	0	Never	26.02	4	160	0
184	Male	9	0	0	No Info	27.32	6.5	85	0
185	Female	27	0	0	No Info	27.32	6.1	145	0
186	Male	41	0	0	Former	23.56	6.5	140	0
187	Female	29	0	0	Never	24.65	4.5	145	0
188	Female	40	0	0	No Info	27.32	5.7	160	0
189	Female	21	0	0	Current	25.11	6.6	145	0
190	Female	77	0	0	Never	27.32	6.5	140	0
191	Male	27	0	0	No Info	27.32	6	90	0
192	Male	65	0	0	Not Current	30.37	3.5	145	0
193	Female	56	0	0	Never	21.54	6	85	0

19									
4	Female	26	0	0	Never	40.9	6.1	140	0
19									
5	Female	41	0	0	Never	26.7	6.2	130	0
19									
6	Female	45	0	0	Former	42.24	6.1	130	0
19									
7	Male	80	0	0	No Info	27.32	6	140	0
19									
8	Male	76	0	0	Never	23.93	5.8	130	0
19									
9	Female	3	0	0	No Info	15.48	5.7	130	0
20									
0	Male	37	0	0	Never	37.24	7	126	1

HASIL DAN PEMBAHASAN

A. Algoritma Naive Bayes

Langkah pertama adalah pemilihan data set (data training) yang melibatkan semua atribut dari set data asli yang dipilih untuk memperoleh atribut yang dibutuhkan untuk proses penambangan data berikutnya yang telah di tampilkan pada tabel 1. Langkah selanjutnya menghitung jumlah status diabetes dari total data yang ada.

- $H1:P(\text{Diabetes}="0")= 166/200 = 0.83$
- $H2:P(\text{Diabetes}="1")= 34/200 = 0.17$

Menghitung probabilitas dari setiap atribut data X, kemudian dibagi dengan banyaknya jumlah diabetes "0" dan "1" :

Tabel 2. Data X

No	gender	age	hypertension	Heart disease	Smoking history	BMI	HbA1c Level	Blood Glucose Level
1	Female	70	0	0	Former	26.53	6.2	280

- $P(\text{Gender} = \text{"Female"}) = 109/166 = 0.65662651$
- $P(\text{Gender} = \text{"Female"}) = 17/34 = 0.50000000$

- $P(\text{Age} = \text{"70"}) = 0/166 = 0.0$
- $P(\text{Age} = \text{"70"}) = 1/34 = 0.02941176$

- $P(\text{hypertension} = \text{"0"}) = 157/166 = 0.94578313$
- $P(\text{hypertension} = \text{"0"}) = 29/34 = 0.85294118$

- $P(\text{Heart disease} = \text{"0"}) = 160/166 = 0.96385542$
- $P(\text{Heart disease} = \text{"0"}) = 31/34 = 0.91176471$

- $P(\text{Smoking history} = \text{" Former"}) = 16/166 = 0.09638554$
- $P(\text{Smoking history} = \text{" Former"}) = 6/34 = 0.17647059$

- $P(\text{BMI} = \text{"26.53"}) = 0/166 = 0.0$
- $P(\text{BMI} = \text{"26.53"}) = 1/34 = 0.02941176$

- $P(\text{HbA1c Level} = \text{"6.2"}) = 12/166 = 0.07228916$
- $P(\text{HbA1c Level} = \text{"6.2"}) = 3/34 = 0.08823529$

- $P(\text{Blood Glucose Level} = \text{"280"}) = 0/166 = 0.0$
- $P(\text{Blood Glucose Level} = \text{"280"}) = 4/34 = 0.11764706$

Hitung total masing-masing nilai pada setiap atribut Data X :

- $P(X|\text{Diabetes} = \text{"0"})$
 $= 0.65662651 * 0.0 * 0.94578313 * 0.96385542 * 0.09638554 * 0.0 * 0.07228916 * 0.0$
 $= 0.0$
- $P(X|\text{Diabetes} = \text{"1"})$
 $= 0.50000000 * 0.02941176 * 0.85294118 * 0.91176471 * 0.17647059 * 0.02941176 * 0.08823529 * 0.11764706$

$$= 0.00000062$$

B. Algoritma K-Nearest Neighbor (KNN)

Langkah pertama yang dilakukan dalam menghitung menggunakan KNN adalah menentukan parameter K (jumlah tetangga yang paling dekat). Nilai K diperoleh dengan mencoba-coba. Disini penulis mencoba nilai K=1, K=3, K=5, dan K=7. Setelah nilai K diperoleh, langkah selanjutnya adalah mencari nilai kuadrat jarak euclid (*euclidean distance*) masing-masing objek terhadap data *training*. Sebelum menghitung nilai tersebut, data haruslah bersifat angka. Oleh karena itu data *gender* dan *smoking history* haruslah dirubah menjadi data angka. Data tersebut dapat dilihat pada tabel dibawah ini.

Tabel 3. Data *Gender*

No	Nilai Huruf	Nilai Angka
1	Male	0
2	Female	1

Tabel 4. Data *Smoking History*

No	Nilai Huruf	Nilai Angka
1	no info	0
2	never	1
3	former	2
4	current	4
5	not current	3

Langkah selanjutnya adalah menghitung jarak antara Data X dan semua data yang ada di data training. Contoh perhitungan manual Data X.

$$d_1 = \sqrt{\frac{(1 - 1)^2 + (0.599 - 0.874)^2 + (0.0 - 0.0)^2 + (0.0 - 0.0)^2 + (0.250 - 0.500)^2 + (0.125 - 0.253)^2 + (0.636 - 0.490)^2 + (0.295 - 0.909)^2}{8}} = 0.743$$

$$d_3 = \sqrt{\frac{(0 - 1)^2 + (0.349 - 0.874)^2 + (0.0 - 0.0)^2 + (0.0 - 0.0)^2 + (0.250 - 0.500)^2 + (0.269 - 0.253)^2 + (0.400 - 0.490)^2 + (0.354 - 0.909)^2}{8}} = 1.286$$

Untuk perhitungan keseluruhan semua data training dapat dilihat pada tabel dibawah ini :

Tabel 5. Data Nilai *Euclidean Distance* Data Testing

No	Gender	age	Hyper tension	Heart disease	Smoking history	bmi	HbA1c level	Blood glucose level	Euclidean Distance
1	1	0.59959960	0	0	0.25000000	0.12527783	0.63636364	0.29545455	0.74323241
2	1	1.00000000	0	0	0.50000000	0.44837341	0.54545455	0.54545455	0.43460738
3	0	0.34934935	0	0	0.25000000	0.26934734	0.40000000	0.35454545	1.28635804
4	1	0.52452452	0	0	0.25000000	0.21863003	1.00000000	0.35909091	0.86494323
5	0	0.94994995	1	1	1.00000000	0.12426753	0.23636364	0.34090909	1.91309473
6	1	0.24924925	0	0	0.25000000	0.26934734	0.56363636	0.02272727	1.11583690
7	1	0.54954955	0	0	0.25000000	0.10749646	0.54545455	0.54545455	0.56993553
8	1	0.98748749	0	0	0.00000000	0.19943423	0.40000000	0.02272727	1.02931880
9	0	0.88738739	0	0	0.25000000	0.26469994	0.85454545	0.54545455	1.15206224
10	1	0.39939940	0	0	0.25000000	0.26934734	0.27272727	0.09090909	1.00292399
11	1	0.66216216	0	0	0.25000000	0.26934734	0.47272727	0.02272727	0.94550123
12	1	0.64964965	1	0	0.25000000	0.73368357	0.56363636	0.34090909	1.29307131
13	1	0.97497497	0	0	0.50000000	0.44574662	0.27272727	0.22727273	0.74799903
14	1	0.83733734	0	0	0.25000000	0.23641140	0.41818182	0.54545455	0.44913006
15	1	0.94994995	0	0	0.00000000	0.26934734	0.27272727	0.36363636	0.77525160
16	0	1.00000000	0	1	0.50000000	0.20953728	0.72727273	0.90909091	1.43994675
17	0	0.18668669	0	0	0.25000000	0.33077389	0.47272727	0.54545455	1.29408426
18	1	0.52452452	0	0	0.25000000	0.21196201	0.40000000	0.35454545	0.70904606
19	1	0.52452452	0	0	0.00000000	0.26934734	0.40000000	0.00000000	1.09895902
20	0	0.73723724	0	0	1.00000000	0.30733482	0.85454545	0.63636364	1.21592201
...
180	0	0.31181181	0	0	0.25000000	0.26934734	0.47272727	0.27272727	1.33607039
181	1	0.73723724	0	0	0.25000000	0.23863407	0.40000000	0.04545455	0.91421766
182	0	0.48698699	0	0	0.25000000	0.26934734	0.40000000	0.35909091	1.23449435
183	1	0.31181181	0	0	0.25000000	0.24307941	0.09090909	0.36363636	0.91496823
184	0	0.11161161	0	0	0.00000000	0.26934734	0.54545455	0.02272727	1.61908657
185	1	0.33683684	0	0	0.00000000	0.26934734	0.47272727	0.29545455	0.95740269
186	0	0.51201201	0	0	0.50000000	0.19337240	0.54545455	0.27272727	1.24225794
187	1	0.36186186	0	0	0.25000000	0.21539705	0.18181818	0.29545455	0.89398668
188	1	0.49949950	0	0	0.00000000	0.26934734	0.40000000	0.36363636	0.83483328
189	1	0.26176176	0	0	1.00000000	0.22469186	0.56363636	0.29545455	1.00427576
190	1	0.96246246	0	0	0.25000000	0.26934734	0.54545455	0.27272727	0.69163595
191	0	0.33683684	0	0	0.00000000	0.26934734	0.45454545	0.04545455	1.51225984
192	0	0.81231231	0	0	0.75000000	0.33097595	0.00000000	0.29545455	1.29999070
193	1	0.69969970	0	0	0.25000000	0.15255607	0.45454545	0.02272727	0.94356533
194	1	0.32432432	0	0	0.25000000	0.54374621	0.47272727	0.27272727	0.92477299
195	1	0.51201201	0	0	0.25000000	0.25681956	0.49090909	0.22727273	0.81182344

196	1	0.56206206	0	0	0.50000000	0.57082239	0.47272727	0.22727273	0.81475473
197	0	1.00000000	0	0	0.00000000	0.26934734	0.45454545	0.27272727	1.29313267
198	0	0.94994995	0	0	0.25000000	0.20084866	0.41818182	0.22727273	1.24139501
199	1	0.03653654	0	0	0.00000000	0.03010709	0.40000000	0.22727273	1.21482693
200	0	0.46196196	0	0	0.25000000	0.46979188	0.63636364	0.20909091	1.33827734

Berdasarkan hasil perhitungan *Euclidean Distance* pada tabel diatas, maka langkah selanjutnya adalah mencari perbandingan berdasarkan parameter K yang telah ditentukan dimana jumlah K yang di tentukan adalah K=1, K=3, K=5, dan K=7 yang hasilnya dapat dilihat pada tabel dibawah ini :

Tabel 6. Perhitungan *Euclidean Distance* terhadap K

No	Euclidean Distance	K=1	K=3	K=5	K=7
1	0.743232408				
2	0.43460738			1	1
3	1.286358043				
14	0.449130065			0	0
15	0.775251601				
64	0.56834958				0
65	1.017462604				
83	0.522518655				0
84	1.22311812				
95	0.377549123		1	1	1
96	1.515503862				
126	0.289930755		1	1	1
127	1.602331933				
143	0.231446607	1	1	1	1

SIMPULAN

Berdasarkan hasil perhitungan yang telah dipaparkan pada bab sebelumnya, maka dapat disimpulkan bahwa prediksi penyakit Diabetes dapat dianalisa menggunakan algoritma *Naïve Bayes* dan *K-Nearest Neighbor* (KNN). Hal ini dapat kita lihat berdasarkan hasil probabilitas $P(X|Diabetes="1")$ Data X dengan nilai 0.00000062 pada algoritma *Naïve Bayes* dan perhitungan sejumlah parameter K pada algoritma KNN yang mayoritas menunjukkan hasil Diabetes = "1".

DAFTAR PUSTAKA

- Astutisari, I. D. A. E. C., AAA Yulianti Darmini, A. Y. D., & Ida Ayu Putri Wulandari, I. A. P. W. (2022). Hubungan Pola Makan Dan Aktivitas Fisik Dengan Kadar Gula Darah Pada Pasien Diabetes Melitus Tipe 2 Di Puskesmas Manggis I. *Jurnal Riset Kesehatan Nasional*, 6(2), 79–87. <https://doi.org/10.37294/jrkn.v6i2.350>
- Cholil, S. R., Handayani, T., Prathivi, R., & Ardianita, T. (2021). Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa. *IJCIT (Indonesian Journal on Computer and Information Technology)*, 6(2), 118–127. <https://doi.org/10.31294/ijcit.v6i2.10438>
- Ghani Muttaqin, A., Auliasari, K., & Santi Wahyuni, F. (2020). Penerapan Metode K-Nearest Neighbor Untuk Prediksi Penjualan Berbasis Web Pada Pt.Wika Industri Energy. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 4(2), 1–6. <https://doi.org/10.36040/jati.v4i2.2728>
- Lestari, Zulkarnain, Sijid, & Aisyah, S. (2021). Diabetes Melitus: Review Etiologi, Patofisiologi, Gejala, Penyebab, Cara Pemeriksaan, Cara Pengobatan dan Cara Pencegahan. *UIN Alauddin Makassar*, 1(2), 237–241. <http://journal.uin-alauddin.ac.id/index.php/psb>
- Murni Pratiwi, I., Fauzi, A., Arum Puspita Lestari, S., Cahyana, Y., Ilmu Komputer, F., & Buana Perjuangan Karawang, U. (2024). Penerapan Algoritma Naïve Bayes Untuk Prediksi Penerimaan Karyawan. *Jurnal TEKINKOM*, 7(1), 236–243. <https://doi.org/10.37600/tekinkom.v7i1.1282>
- Pebdika, A., Herdiana, R., & Solihudin, D. (2023). Klasifikasi Menggunakan Metode Naive Bayes Untuk Menentukan Calon Penerima Pip. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(1), 452–458. <https://doi.org/10.36040/jati.v7i1.6303>
- Petersmann, A., Nauck, M., Müller-Wieland, D., Kerner, W., Müller, U. A., Landgraf, R., Freckmann, G., & Heinemann, L. (2018). Definition, classification and diagnostics of diabetes mellitus. *Journal of Laboratory Medicine*, 42(3), 73–79. <https://doi.org/10.1515/labmed-2018-0016>
- Purwati, N., & Dwi Januanti, A. (2021). Aplikasi Data Mining Dengan Algoritma Naive Bayes Untuk Memprediksi Tingkat Kelulusan Mahasiswa. *Jurnal Pepadun*, 2(1), 123–137. <https://doi.org/10.23960/pepadun.v2i1.38>
- Suryawan, N. W., Bachrun, E., Prayitno, S., Bhakti, S., Mulia, H., Relationship, T., Parenting, B., & Behavior, S. (2023). *JPKM Jurnal Profesi Kesehatan Masyarakat*. 4(1), 1–7.
- Syarli, & Muin, A. A. (2020). Metode Naive Bayes Untuk Prediksi Kelulusan. *Jurnal Ilmiah Ilmu Komputer*, 2(1), 22–26. <https://media.neliti.com/media/publications/283828-metode->

naive-bayes-untuk-prediksi-kelulu-139fcfea.pdf

Syihabuddin Azmil Umri, S. (2021). Analisis Dan Komparasi Algoritma Klasifikasi Dalam Indeks Pencemaran Udara Di Dki Jakarta. *JIKO (Jurnal Informatika Dan Komputer)*, 4(2), 98–104. <https://doi.org/10.33387/jiko.v4i2.2871>