

Klasifikasi Penyakit Migrain dengan Metode *Naïve Bayes* pada *Dataset Kaggle*

Michael Yuichi ^{1*}, Yeremia Alfa Susetyo ²

^{1*,2} Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana, Kota Salatiga, Provinsi Jawa Tengah, Indonesia.

Emai: 672021024@student.uksw.edu ^{1*}, yeremia.alfa@uksw.edu ²

Histori Artikel:

Dikirim 23 Oktober 2024; Diterima dalam bentuk revisi 20 November 2024; Diterima 30 November 2024; Diterbitkan 10 Januari 2025. Semua hak dilindungi oleh Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) STMIK Indonesia Banda Aceh.

Abstrak

Migrain merupakan salah satu penyakit neurologis yang umum dijumpai di masyarakat dan memiliki dampak disabilitas yang signifikan. Berdasarkan Global Burden of Disease Study, migrain adalah salah satu gangguan neurologis paling umum di dunia, dengan beban disabilitas yang lebih besar dibandingkan gangguan neurologis lainnya. Dalam data science, klasifikasi data berperan penting untuk menentukan kategori atau kelas objek berdasarkan sejumlah kelas yang tersedia. Salah satu metode pengklasifikasian yang sering digunakan adalah Naïve Bayes, yang memanfaatkan probabilitas matematika dengan asumsi bahwa keputusan yang diambil akurat berdasarkan data yang diberikan. Penelitian ini mengembangkan model klasifikasi menggunakan algoritma Naïve Bayes untuk memprediksi dan mengklasifikasikan data pasien migrain. Model ini menghasilkan nilai klasifikasi dengan akurasi sebesar 88.51% saat dilakukan training dari 280 data train yang diberikan dan 89.02% saat dilakukan testing dari 120 data test yang diberikan. Hasil klasifikasi dapat mendukung dunia medis dalam mendiagnosa jenis migrain secara lebih akurat, mengoptimalkan pengobatan, menghemat biaya kesehatan, serta menjadi dasar bagi penelitian dan pengembangan lebih lanjut dalam bidang neurologi.

Kata Kunci: Migrain; Neurologis; Global Burden of Disease Study; Klasifikasi; Naïve Bayes; Prediksi.

Abstract

Migraine is one of the most common neurological diseases in society and has a significant disability impact. According to the Global Burden of Disease Study, migraine is one of the most common neurological disorders in the world, with a greater disability burden than other neurological disorders. In data science, data classification plays an important role in determining the category or class of an object based on a number of available classes. One frequently used classification method is Naïve Bayes, which utilizes mathematical probabilities with the assumption that the decision made is accurate based on the given data. This research develops a classification model using the Naïve Bayes algorithm to predict and classify migraine patient data. This model produces classification values with an accuracy of 88.51% when training from 280 train data given and 89.02% when testing from 120 test data given. The classification results can support the medical world in diagnosing migraine types more accurately, optimizing treatment, saving health costs, and becoming the basis for further research and development in the field of neurology.

Keyword: Migraine; Neurological; Global Burden of Disease Study; Classification; Naïve Bayes; Prediction.

1. Pendahuluan

Migrain merupakan salah satu gangguan kesehatan yang banyak dijumpai dan sering dialami oleh masyarakat. Menurut *Global Burden of Disease Study*, migrain termasuk dalam kategori gangguan neurologis yang paling umum di seluruh dunia dan memiliki dampak disabilitas yang lebih besar dibandingkan gangguan neurologis lainnya (WHO, 2019). Kejadian migrain cenderung lebih sering dialami oleh perempuan dibandingkan laki-laki, dengan perbandingan 3:1 (Lipton *et al.*, 2001). Meskipun migrain tidak mengancam jiwa, intensitas dan frekuensi serangan dapat menyebabkan disabilitas yang signifikan serta penurunan kualitas hidup, bahkan saat pasien tidak mengalami serangan aktif (A. K. Eigenbrodt *et al.*, 2021). Permasalahan utama dalam penelitian ini adalah klasifikasi tipe-tipe migrain berdasarkan gejala yang muncul, menggunakan metode *Naïve Bayes*. Metode ini dipilih karena memiliki kemampuan untuk membangun model secara cepat, serta efektif dalam memprediksi dan menganalisis data (Rish, 2001). Tanpa adanya klasifikasi yang tepat, diagnosis dan pengobatan migrain dapat menjadi tidak akurat atau bahkan berpotensi mengarah pada misdiagnosis. Hal ini dapat menyebabkan pemberian perawatan yang tidak sesuai, yang bukan hanya mengurangi kualitas hidup penderita, tetapi juga meningkatkan beban ekonomi akibat perawatan yang tidak tepat (Nasawida & Sari, 2022). Oleh karena itu, penting untuk menggunakan teknik pembelajaran terawasi (supervised learning) seperti *Naïve Bayes*, yang dapat memberikan kontribusi terhadap penelitian dengan mengonfigurasi penerapan model algoritma serta mekanisme validasi untuk menghasilkan tingkat akurasi, presisi, dan recall yang lebih tinggi (Kumar *et al.*, 2020).

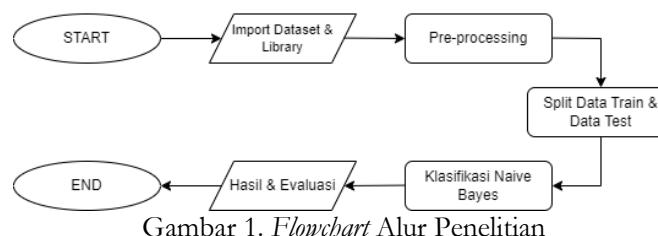
Metode yang akan digunakan untuk menganalisis migrain adalah klasifikasi, yang bertujuan untuk memasukkan data ke dalam kategori tertentu dari sejumlah kelas yang ada. Salah satu metode klasifikasi yang populer dalam *data science* adalah *Naïve Bayes*. *Naïve Bayes* menggunakan teori probabilitas untuk mengambil keputusan berdasarkan informasi yang tersedia, dengan asumsi bahwa semua fitur saling independen (McCallum & Nigam, 1998). Metode ini memanfaatkan pengalaman masa lalu untuk memprediksi kemungkinan kejadian di masa depan, yang dikenal sebagai *Teorema Bayes* (Bayes, 1763). Berdasarkan kemampuan tersebut, *Naïve Bayes* sangat cocok untuk memprediksi penyakit migrain yang memiliki beberapa kategori, seperti migrain tanpa aura, migrain dengan aura, atau kategori spesifik lainnya. Dengan klasifikasi yang tepat, diagnosis migrain menjadi lebih akurat, sehingga penanganan yang diberikan lebih tepat sasaran, mengurangi risiko misdiagnosis dan meningkatkan kualitas hidup pasien (K. Putri Paramita Abyuda & S. Nandar Kurniawan, 2021). Penelitian ini bertujuan untuk mengklasifikasikan migrain menggunakan metode *Naïve Bayes*, yang dapat meningkatkan akurasi diagnosis, optimasi pengobatan, penghematan biaya kesehatan, serta memberikan dukungan bagi penelitian dan pengembangan lebih lanjut dalam bidang ini. Dengan menggunakan model yang dikembangkan melalui algoritma *Naïve Bayes*, penelitian ini diharapkan dapat membantu dunia medis dalam mengidentifikasi kemungkinan jenis migrain berdasarkan gejala yang dimiliki pasien.

Metode *Naïve Bayes* dan *Naïve Bayes Updateable* dalam *statistical learning* telah terbukti memberikan hasil yang efisien dengan akurasi mencapai 93,50%. Dibandingkan dengan model pembelajaran fungsional, seperti Quadratic Discriminant Analysis (QDA), *Naïve Bayes* menunjukkan akurasi yang lebih tinggi sebesar 93% dan unggul dalam beberapa metrik seperti Positive Predictive Value (PPV) dan True Positive Rate (TPR), yang masing-masing mencapai nilai tertinggi 0,93. Untuk Area Under the ROC Curve (AUCROC), QDA dan Multi-Layer Perceptron mencatatkan nilai tertinggi sebesar 0,97. Dalam *statistical learning*, *Naïve Bayes* dan *Naïve Bayes Updateable* mencatatkan nilai tertinggi untuk Area Under the PRC Curve (AUCPRC) dan F1-Score, yakni 0,97 dan 0,93 (Kannan & Thangarasu, 2022). Berdasarkan temuan-temuan tersebut, metode *Naïve Bayes* terbukti efektif dalam klasifikasi data migrain, dengan akurasi dan metrik lainnya yang bersaing dengan algoritma lain. Oleh karena itu, metode ini dipilih untuk mengklasifikasikan jenis migrain dengan tujuan memperluas temuan terkait model klasifikasi yang dapat diandalkan berdasarkan dataset yang tersedia di Kaggle. Klasifikasi menggunakan *Naïve Bayes* juga menunjukkan keunggulan dalam hal akurasi jika dibandingkan dengan metode lain seperti Regresi Logistik Multinomial. Sebagai contoh, *Naïve Bayes* mencapai akurasi

88,73%, sedangkan Regresi Logistik Multinomial hanya memperoleh 80,44% (Narayan & Khatoon, 2023). Hal ini menunjukkan bahwa *Naïve Bayes* lebih unggul dalam menganalisis dan mendeteksi jenis migrain. Penggunaan teknik pembelajaran mesin lainnya juga menunjukkan peningkatan kinerja yang signifikan setelah penerapan *data augmentation*. Sebagai contoh, akurasi *Naïve Bayes* meningkat dari 67% menjadi 93%, sementara SVM meningkat dari 90% menjadi 95%, dan Adaboost dari 93% menjadi 94%. *Data augmentation* terbukti meningkatkan efisiensi algoritma *Naïve Bayes* dalam klasifikasi migrain (Khan *et al.*, 2024). Penerapan *data augmentation* pada *Naïve Bayes* menghasilkan peningkatan kinerja yang lebih signifikan dibandingkan dengan model lainnya, menjadikannya metode yang lebih tepat saat menangani data migrain dengan jumlah data yang lebih besar. Berdasarkan penelitian-penelitian tersebut, tiga teknik pembelajaran mesin yang digunakan untuk mengklasifikasikan penyakit migrain berdasarkan gejala yang ada menunjukkan bahwa penerapan klasifikasi menggunakan *Naïve Bayes* menghasilkan nilai kinerja yang meningkat secara signifikan saat diberikan *data augmentation* dibandingkan dengan fitur lainnya. Hasil penelitian ini mengonfirmasi bahwa model *Naïve Bayes* adalah algoritma yang paling tepat digunakan dalam mengklasifikasikan migrain dengan penerapan *data augmentation*.

2. Metode Penelitian

Untuk mengetahui hasil klasifikasi migrain menggunakan metode *Naïve Bayes*, maka langkah-langkah yang diperlukan dapat dilihat pada Gambar 1.



Berdasarkan alur dari metodologi diatas, tahap mempersiapkan dataset diambil dari situs <https://www.kaggle.com/datasets/weinoose/migraine-classification>. Di dalam dataset yang digunakan, terdapat 24 atribut dan 400 baris data, dan 7 jenis migrain, yaitu *Typical Aura with Migraine*, *Typical Aura without Migraine*, *Migraine without Aura*, *Basilar Type Aura*, *Sporadic Hemiplegic*, *Familial Hemiplegic Migraine*, dan *Other*. Keterangan atribut dapat dilihat pada Tabel 1.

Tabel 1. Keterangan Atribut pada Dataset

No	Atribut	Keterangan
1	<i>Age</i>	Usia penderita migrain
2	<i>Duration</i>	Migrain yang dialami berdasarkan durasi dibagi menjadi 3, jumlah nilai yang diberi adalah 1 hingga 3
3	<i>Frequency</i>	Migrain yang dialami berdasarkan frekuensi dibagi menjadi 8, jumlah nilai yang diberi adalah 1 hingga 8
4	<i>Location</i>	Migrain yang dialami berdasarkan lokasi dibagi menjadi 3, jumlah nilai yang diberi adalah 0 hingga 2
5	<i>Character</i>	Migrain yang dialami berdasarkan karakter dibagi menjadi 3, jumlah nilai yang diberi adalah 0 hingga 2
6	<i>Intensity</i>	Migrain yang dialami berdasarkan intensitas dibagi menjadi 4, jumlah nilai yang diberi adalah 0 hingga 3

7	<i>Nausea</i>	Migrain yang dialami berdasarkan <i>Nausea</i> atau gejala perasaan tidak nyaman pada bagian belakang tenggorokan atau lambung dibagi menjadi 2, jumlah nilai yang diberi adalah 0 hingga 1
8	<i>Vomit</i>	Migrain yang dialami berdasarkan <i>vomitus</i> atau muntah dibagi menjadi 2, jumlah nilai yang diberi adalah 0 hingga 1
9	<i>Phonophobia</i>	Migrain yang dialami berdasarkan <i>phonophobia</i> atau ketakutan terhadap suara dibagi menjadi 2, jumlah nilai yang diberi adalah 0 hingga 1
10	<i>Photophobia</i>	Migrain yang dialami berdasarkan <i>photophobia</i> atau mata terasa sakit dibagi menjadi 2, jumlah nilai yang diberi adalah 0 hingga 1
11	<i>Visual</i>	Migrain yang dialami berdasarkan <i>visual</i> atau kondisi penglihatan dibagi menjadi 5, jumlah nilai yang diberi adalah 0 hingga 4
12	<i>Sensory</i>	Migrain yang dialami berdasarkan <i>sensory</i> atau kondisi indra dibagi menjadi 3, jumlah nilai yang diberi adalah 0 hingga 2
13	<i>Dysphasia</i>	Migrain yang dialami berdasarkan <i>dysphasia</i> atau gangguan berbicara, menulis, dan berbahasa dibagi menjadi 2, jumlah nilai yang diberi adalah 0 hingga 1
14	<i>Dysarthria</i>	Migrain yang dialami berdasarkan <i>dysarthria</i> atau kekuatan otot untuk berbicara menjadi lemah dibagi menjadi 2, jumlah nilai yang diberi adalah 0 hingga 1
15	<i>Vertigo</i>	Migrain yang dialami berdasarkan <i>vertigo</i> atau pusing seperti berputar dibagi menjadi 2, jumlah nilai yang diberi adalah 0 hingga 1
16	<i>Tinnitus</i>	Migrain yang dialami berdasarkan <i>tinnitus</i> atau telinga berdengung dibagi menjadi 2, jumlah nilai yang diberi adalah 0 hingga 1
17	<i>Hypoacusis</i>	Migrain yang dialami berdasarkan <i>hypoacusis</i> atau kehilangan pendengaran total dibagi menjadi 2, jumlah nilai yang diberi adalah 0 hingga 1
18	<i>Diplopia</i>	Migrain yang dialami berdasarkan <i>diplopia</i> atau penglihatan yang melihat 1 objek terasa ganda dibagi menjadi 2, jumlah nilai yang diberi adalah 0 hingga 1
19	<i>Defect</i>	Migrain yang dialami berdasarkan <i>defect</i> atau cacat dibagi menjadi 2, jumlah nilai yang diberi adalah 0 hingga 1
20	<i>Ataxia</i>	Migrain yang dialami berdasarkan <i>ataxia</i> atau gangguan keseimbangan dibagi menjadi 2, jumlah nilai yang diberi adalah 0 hingga 1
21	<i>Conscience</i>	Migrain yang dialami berdasarkan <i>conscience</i> atau kemampuan membedakan tindakan yang benar atau salah dibagi menjadi 2, jumlah nilai yang diberi adalah 0 hingga 1
22	<i>Paresthesia</i>	Migrain yang dialami berdasarkan <i>paresthesia</i> atau kesemutan dibagi menjadi 2, jumlah nilai yang diberi adalah 0 hingga 1
23	DPF	Migrain yang dialami berdasarkan DPF atau <i>Decreased Palpebral Fissure</i> atau reduksi jarak antara pinggir mata atas dan bawah relatif terhadap kontralateral mata dibagi menjadi 2, jumlah nilai yang diberi adalah 0 hingga 1
24	Type	Tipe dari Migrain

Tabel 1 menunjukkan atribut-atribut yang terdapat pada dataset migrain yang akan digunakan dalam penelitian ini. Atribut-atribut tersebut dibagi menjadi dua jenis, yaitu atribut yang menggambarkan kondisi dan atribut yang menggambarkan kelas. Atribut yang menggambarkan kondisi dapat dilihat pada Tabel 1, nomor 1 hingga 23, sementara atribut kelas dapat ditemukan pada nomor 24 dalam tabel yang sama. Dataset migrain ini telah dikemas dan disimpan dalam format *Comma Separated Values* (CSV) untuk memudahkan pengolahan data lebih lanjut. Setelah dataset disiapkan, tahap selanjutnya adalah melakukan *pre-processing* untuk memastikan bahwa data yang digunakan bersih, terstruktur dengan baik, dan siap untuk dianalisis lebih lanjut dalam penelitian ini. Tahap *pre-processing* dilakukan untuk memastikan bahwa dataset migrain bebas dari nilai yang hilang (*missing values*) dan siap untuk dianalisis lebih lanjut. Dataset migrain yang terdiri dari 400 baris data kemudian akan dibagi menjadi dua bagian, yaitu data untuk *training* dan data untuk *testing*. Pembagian data ini dapat dilihat pada Tabel 2.

Tabel 2. Pembagian Data *Train* dan *Test*

Algoritma	Persentase Data <i>Train</i>	Jumlah Data <i>Train</i>	Persentase Data <i>Testing</i>	Jumlah Data <i>Test</i>
Naïve Bayes	70%	280	30%	120

Tabel 2 menunjukkan pembagian data antara *training* dan *testing* dengan perbandingan 70:30. Dalam hal ini, data *training* terdiri dari 70% atau sebanyak 280 data, sedangkan data *testing* mencakup 30% atau sebanyak 120 data. Dengan demikian, total keseluruhan data dalam dataset adalah 400, yang memastikan bahwa pembagian data tersebut proporsional. Setelah proses pembagian data (*splitting*) antara data *training* dan *testing* dilakukan, langkah berikutnya adalah penerapan metode *modeling*, yaitu penggunaan metode Naïve Bayes. Dalam penelitian ini, yang digunakan adalah Gaussian Naïve Bayes. Naïve Bayes diterapkan untuk memprediksi probabilitas berdasarkan penerapan *Teorema Bayes* (Bayes, 1763). Metode ini disebut Naïve Bayes karena mengasumsikan bahwa semua variabel input berkontribusi secara independen terhadap klasifikasi, yang dikenal sebagai *class conditional independence* (Rish, 2001). Persamaan perhitungan *Teorema Bayes* dapat dilihat pada Rumus (1):

$$P(y) = \frac{P(x|y)P(y)}{P(x)}$$

Dimana:

- y = data kelas yang belum diketahui
- x = hipotesis data y
- $P(x|y)$ = probabilitas hipotesis x berdasarkan kondisi y
- $P(x)$ = probabilitas hipotesis x
- $P(y|x)$ = probabilitas y berdasarkan kondisi hipotesis x
- $P(y)$ = probabilitas hipotesis y

Setelah proses *modeling* dengan metode *Naïve Bayes* selesai, tahap selanjutnya adalah evaluasi model menggunakan *confusion matrix*. Tujuan dari evaluasi ini adalah untuk mengukur kinerja dan akurasi model yang telah dibangun (Fawcett, 2006). Pengukuran kinerja model dengan menggunakan *confusion matrix* dapat dilihat pada Tabel 3.

Tabel 3. *Confusion Matrix*

	<i>Classified Positive</i>	<i>Classified Negative</i>
<i>Actual Positive</i>	TP	FN
<i>Actual Negative</i>	FP	TN

True Positive (TP) atau *True Negative* (TN) menunjukkan bahwa model klasifikasi berhasil melakukan prediksi dengan benar, sementara *False Positive* (FP) dan *False Negative* (FN) menunjukkan adanya kesalahan dalam klasifikasi yang dilakukan (Fawcett, 2006). Hasil dari *confusion matrix* ini dapat digunakan untuk menghitung beberapa metrik evaluasi model, seperti *accuracy*, *precision*, *recall*, dan *F1-score*. *Accuracy* dihitung sebagai perbandingan antara jumlah prediksi yang benar dengan keseluruhan data yang digunakan. *Precision* mengukur proporsi prediksi positif yang benar-benar positif di antara seluruh prediksi yang positif. Sementara itu, *recall* merupakan rasio prediksi positif yang benar dibandingkan dengan keseluruhan data yang sebenarnya positif. Terakhir, *F1-score* adalah rata-rata harmonik antara *precision* dan *recall*, yang digunakan untuk memberikan gambaran umum tentang kinerja model (Sasaki, 2007). Rumus untuk menghitung nilai *accuracy* dapat dilihat pada Rumus (2):

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Kedua, rumus nilai presisi atau *precision* dapat dilihat pada rumus:

$$\text{Precision} = \frac{TP}{TP + FP}$$

Ketiga, rumus nilai *recall* dapat dilihat pada rumus:

$$\text{Recall} = \frac{TP}{TP + FN}$$

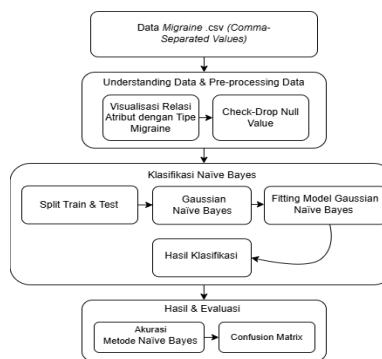
Keempat, rumus nilai *f1-score* dapat dilihat pada rumus:

$$F1 = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

3. Hasil dan Pembahasan

3.1 Hasil

Pada tahap pengklasifikasian, untuk mendapatkan hasil dari klasifikasi metode *Naïve Bayes* diperlukan tahap pengolahan pada dataset migrain yang dapat dilihat pada Gambar 2.



Gambar 2. Arsitektur Penelitian

Penelitian ini dimulai dengan membaca dataset penyakit migrain yang telah didapatkan melalui web resmi kaggle. Dataset tersebut diberi nama data.csv dan dapat dilihat seperti pada Tabel 4.

Table 4. Data Gejala Penyakit Migrain

Age	Duration	Frequency	...	Vertigo	Tinnitus	DPF	Type
0	30	1	5	...	0	0	Typical aura with migraine
1	50	3	5	...	1	0	Typical aura with migraine
2	53	2	1	...	0	0	Typical aura with migraine
3	45	3	5	...	1	0	Typical aura with migraine
4	53	1	1	...	0	0	Typical aura with migraine
5	49	1	1	...	0	0	Migraine without aura
6	27	1	5	...	1	1	Basilar-type aura
7	24	1	1	...	1	0	Typical aura with migraine
8	50	1	5	...	1	0	Typical aura with migraine
9	23	1	1	...	0	0	Typical aura with migraine

Data diatas akan dilakukan proses *cleaning* dengan melakukan *drop NaN* (*Not A Number*) atau null, dimana data yang bukan sebuah angka atau nilai yang hilang (*missing values*) akan dihapus atau dibersihkan barisnya sehingga data yang digunakan hanyalah data yang lengkap pada setiap atribut yang ada.

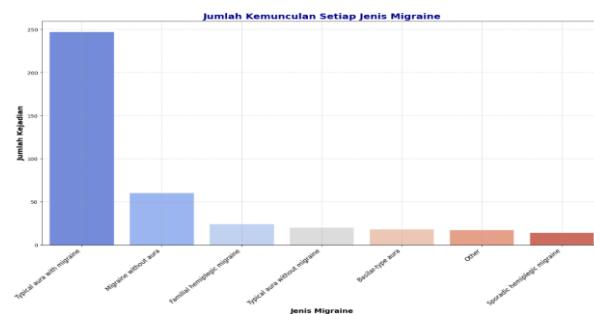
Kode Program 1 Preprocessing Data

```
1 df.isnull().sum().to_numpy()
2 df.dropna()
```

Pada kode program 1 diatas, baris pertama digunakan untuk mendekripsi dan menghitung jumlah nilai NaN atau null dalam *DataFrame*, dan kemudian mengubah hasilnya menjadi *array NumPy*. Pada baris kedua dilakukan proses drop pada setiap baris yang memiliki nilai NaN atau null, hasilnya tidak ada yang memiliki nilai NaN atau null. Data yang telah melalui proses *cleaning* kemudian akan divisualisasi untuk melihat banyaknya jenis migrain yang muncul di dalam dataset yang digunakan. Visualisasi menggunakan *library seaborn* pada aplikasi Google Colab.

Kode Program 2 Visualisasi

```
1 class_counts = df['Type'].value_counts()
2 plt.figure(figsize=(15, 10))
3 sns.barplot(x=class_counts.index, y=class_count.values,
4 palette='coolwarm')
5 plt.title('Jumlah Kemunculan Setiap Jenis Migraine', fontsize=18,
6 fontweight='bold', color='darkblue')
7 plt.xlabel('Jenis Migraine', fontsize=14, fontweight='bold',
8 color='black')
7 plt.ylabel('Jumlah Kejadian', fontsize=14, fontweight='bold',
9 color='black')
8 plt.xticks(rotation=45, ha='right', fontsize=12)
9 plt.tight_layout()
10 plt.show()
```



Gambar 3. Visualisasi Jumlah Data Setiap Jenis Migrain

Setelah itu, dilakukan proses *train* dan *test* untuk melatih model pada data *train* dan mengukur performa pada data *test*. Bisa dikatakan bahwa fungsi ini untuk menghitung seberapa baiknya hasil suatu prediksi dari model yang digunakan.

Kode Program 3 Training dan Testing

```
1 X = train_dataset.drop('Type', axis=1)
2 Y = train_dataset['Type']
3 X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.3,
random_state=0)
```

Pada baris pertama dan kedua dalam kode program, dataset dibagi menjadi fitur (*X*) dan target (*Y*), di mana *X* mencakup semua kolom kecuali kolom *Type*, sementara *Y* berisi kolom *Type* sebagai target. Pada baris ketiga, fungsi *train_test_split()* digunakan untuk membagi data menjadi data latih (*X_train*, *Y_train*) dan data uji (*X_test*, *Y_test*), dengan proporsi 70% data digunakan untuk pelatihan model dan 30% untuk pengujian model.

Kode Program 4 Hasil dan Evaluasi

```
1 gnb = GaussianNB()
2 gnb.fit(X_train_scaled, Y_train)
3 from sklearn.metrics import accuracy_score, classification_report,
confusion_matrix
4 import pandas as pd
5 def print_score(clf, X_train_scaled, Y_train, X_test_scaled, Y_test,
train=True):
6     if train:
7         pred = clf.predict(X_train_scaled)
8         clf_report = pd.DataFrame(classification_report(Y_train, pred,
output_dict=True))
```

Pada baris pertama dan kedua dalam kode program, algoritma Gaussian Naïve Bayes (*gnb*) diterapkan untuk klasifikasi dengan melatih model menggunakan data pelatihan (*X_train_scaled* dan *Y_train*). Fungsi *print_score()* yang didefinisikan dalam kode ini berfungsi untuk mengevaluasi kinerja model pada data *train* atau *test*, tergantung pada nilai parameter *train*. Jika *train=True*, fungsi ini akan menghasilkan prediksi pada data *train* dan menyajikan laporan klasifikasi dalam bentuk *DataFrame* menggunakan *classification_report* dari *sklearn*. Langkah ini bertujuan untuk memudahkan analisis hasil prediksi dan metrik evaluasi model. Berdasarkan hasil penelitian ini, metode Naïve Bayes menghasilkan akurasi sebesar 89,02% pada tahap pengujian (*testing*) untuk klasifikasi penyakit migrain, yang menunjukkan kinerja yang cukup baik. Penelitian sebelumnya yang berjudul “Analisis Perbandingan Kinerja Algoritma Multilayer Perceptron dan K-Nearest Neighbor pada Klasifikasi Tipe Migrain” menunjukkan bahwa *Multilayer Perceptron* (MLP) memiliki akurasi lebih tinggi sebesar 91%, sedangkan *K-Nearest Neighbor* (KNN) hanya mencapai akurasi 72%. Akurasi yang lebih tinggi pada MLP disebabkan oleh kemampuannya dalam menangani hubungan non-linear antar fitur dan secara iteratif mengoptimalkan bobot dalam jaringan saraf, yang meningkatkan ketepatan klasifikasi pada data yang lebih kompleks, seperti migrain. Sementara itu, metode Naïve Bayes bekerja dengan asumsi independensi antar fitur, yang kadang membatasi kinerjanya ketika terdapat korelasi antar gejala yang tidak sepenuhnya independen. Meskipun demikian, akurasi 89,02% yang diperoleh menunjukkan bahwa metode Naïve Bayes tetap memiliki performa yang kompetitif. Performa yang mendekati MLP ini dapat dijelaskan oleh kesederhanaan Naïve Bayes, yang menjadikannya kurang rentan terhadap *overfitting*, khususnya pada dataset dengan ukuran terbatas. Pada penerapan metode Naïve Bayes dalam kode program 4, hasil evaluasi untuk data *train* dan *test* menunjukkan akurasi yang cukup tinggi, masing-masing sebesar 88,51% dan 89,02%. Hasil ini dapat dilihat pada Gambar 4 dan Gambar 5.

```

Train Result:
=====
Accuracy Score: 88.51%

CLASSIFICATION REPORT:
          Basilar-type aura  Familial hemiplegic migraine \
precision      0.982456           0.770833
recall        0.767123           0.521127
f1-score      0.861538           0.621849
support       73.000000          71.000000

          Migraine without aura   Other  Sporadic hemiplegic migraine \
precision      0.926471     1.000000           0.882353
recall        1.000000     0.928571           0.869565
f1-score      0.961832     0.962963           0.875912
support       63.000000    70.000000          69.000000

          Typical aura with migraine  Typical aura without migraine \
precision      0.802752           1.0
recall        0.983146           1.0
f1-score      0.883838           1.0
support       178.000000          68.0

          accuracy  macro avg  weighted avg
precision  0.885135  0.909266  0.889507
recall    0.885135  0.867076  0.885135
f1-score   0.885135  0.881133  0.879743
support   0.885135  592.000000  592.000000

Confusion Matrix:
[[ 56 11  0  0  6  0  0]
 [ 0 37  0  0  0 34  0]
 [ 0  0 63  0  0  0  0]
 [ 0  0  5 65  0  0  0]
 [ 0  0  0  0 60  9  0]
 [ 1  0  0  0  2 175  0]
 [ 0  0  0  0  0  0 68]]

```

Gambar 4. Hasil Training

```

Test Result:
=====
Accuracy Score: 89.02%

CLASSIFICATION REPORT:
          Basilar-type aura  Familial hemiplegic migraine \
precision      1.000000           0.777778
recall        0.777778           0.482759
f1-score      0.875000           0.595745
support       27.000000          29.000000

          Migraine without aura   Other  Sporadic hemiplegic migraine \
precision      1.0           1.0           0.866667
recall        1.0           1.0           0.838710
f1-score      1.0           1.0           0.852459
support       37.0          30.0          31.000000

          Typical aura with migraine  Typical aura without migraine \
precision      0.770115           1.0
recall        0.971014           1.0
f1-score      0.858974           1.0
support       69.000000          32.0

          accuracy  macro avg  weighted avg
precision  0.890196  0.916366  0.896314
recall    0.890196  0.867180  0.890196
f1-score   0.890196  0.883168  0.884694
support   0.890196  255.000000  255.000000

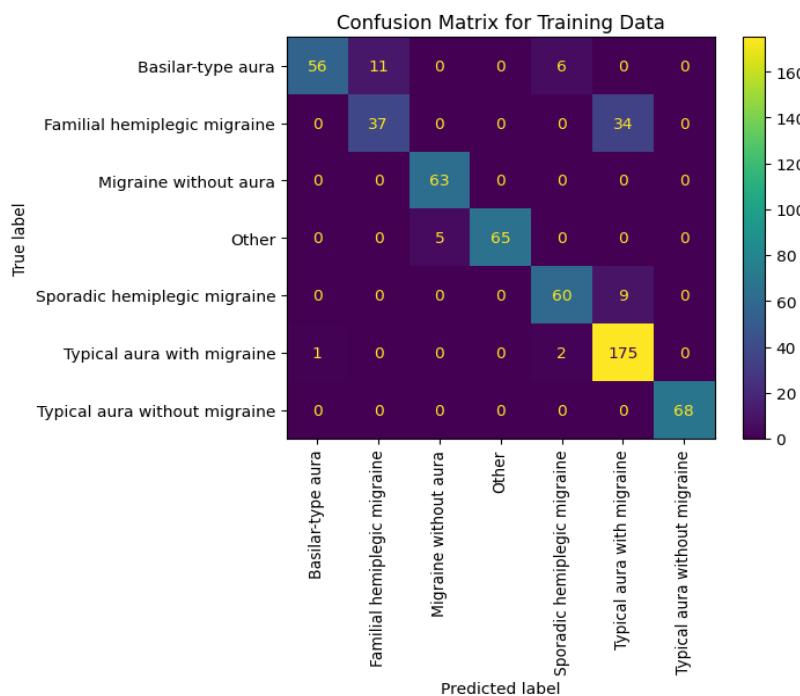
Confusion Matrix:
[[21  2  0  0  4  0  0]
 [ 0 14  0  0  0 15  0]
 [ 0  0 37  0  0  0  0]
 [ 0  0  0 30  0  0  0]
 [ 0  0  0  0 26  5  0]
 [ 0  2  0  0  0 67  0]
 [ 0  0  0  0  0  0 32]]

```

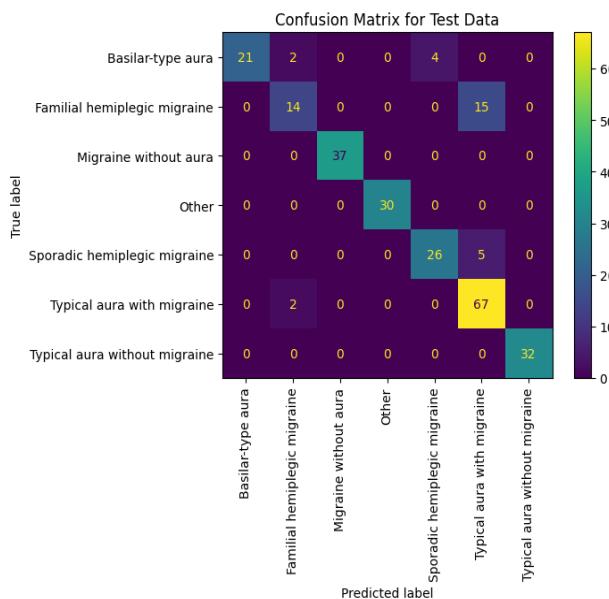
Gambar 5. Hasil Testing

Hasil pelatihan model klasifikasi menunjukkan skor akurasi sebesar 88,51%. Laporan klasifikasi mengidentifikasi berbagai tipe migrain, seperti *Typical Aura with Migraine*, *Typical Aura without Migraine*, *Migraine without Aura*, *Basilar Type Aura*, *Sporadic Hemiplegic*, *Familial Hemiplegic Migraine*, dan *Other*. Masing-masing tipe migrain memiliki metrik *precision*, *recall*, dan *F1-score* yang bervariasi. *Precision* tertinggi dicapai oleh kelas *Other* dengan nilai 1,0, sementara *recall* tertinggi juga tercatat pada kelas

Other serta beberapa kelas lainnya, yang masing-masing mencapai nilai 1,0. *F1-score* tertinggi juga dicapai oleh kelas *Other* dengan nilai 0,962963. Secara keseluruhan, rata-rata *precision*, *recall*, dan *F1-score* untuk seluruh kelas adalah sekitar 0,885135, 0,867076, dan 0,879743, yang menunjukkan bahwa model berhasil mengklasifikasikan tipe-tipe migrain dengan baik, meskipun terdapat variasi kinerja antar kelas. Hasil pengujian model klasifikasi menunjukkan skor akurasi sebesar 89,02%. Metrik *precision*, *recall*, dan *F1-score* juga bervariasi di antara kelas-kelas yang ada. *Precision* tertinggi dicapai oleh beberapa kelas, yaitu *Basilar Type Aura*, *Migraine without Aura*, dan *Other*, yang masing-masing memperoleh nilai 1,0. Begitu pula dengan *recall*, yang tertinggi pada kelas-kelas yang sama dengan nilai 1,0. *F1-score* tertinggi tercatat pada kelas *Other* dan *Migraine without Aura* dengan nilai 1,0. Rata-rata *precision*, *recall*, dan *F1-score* untuk seluruh kelas masing-masing adalah sekitar 0,890196, 0,867180, dan 0,884694. Hasil ini menunjukkan bahwa model memiliki performa yang baik dalam mengklasifikasikan tipe-tipe migrain, dengan sedikit variasi kinerja antar kelas. Penerapan metode Naïve Bayes dalam penelitian ini juga dapat dilihat pada hasil *confusion matrix*, yang digunakan untuk membandingkan hasil prediksi dengan kondisi data yang ada berdasarkan dataset yang digunakan. Hasil *confusion matrix* tersebut dapat dilihat pada Gambar 6 dan Gambar 7.

Gambar 6. *Confusion Matrix* dari Data Train

Kinerja pada tipe *Migraine without Aura* dan *Typical Aura without Migraine* menunjukkan hasil yang sangat baik, karena hampir seluruh prediksi benar, dengan nilai yang berada di luar diagonal *confusion matrix* sangat sedikit atau bahkan nol. Tipe *Other* juga menunjukkan kinerja yang baik, dengan mayoritas prediksi yang benar, meskipun terdapat beberapa kesalahan, yaitu lima prediksi yang salah mengklasifikasikan sebagai *Basilar-Type Aura*. Di sisi lain, tipe *Familial Hemiplegic Migraine* dan *Sporadic Hemiplegic Migraine* menunjukkan banyak prediksi yang salah, yang mengindikasikan bahwa model mengalami kesulitan dalam mengklasifikasikan kelas-kelas ini dengan benar. Terutama pada tipe *Familial Hemiplegic Migraine*, yang sering diprediksi sebagai *Typical Aura with Migraine*. Tipe *Basilar-Type Aura* dan *Typical Aura with Migraine* juga memiliki beberapa prediksi yang salah, meskipun mayoritas prediksi untuk kedua kelas ini tetap benar.



Gambar 7. Confusion Matrix dari Data Test

Tipe *Migraine without Aura* dan *Typical Aura without Migraine* menunjukkan kinerja baik karena prediksi sepenuhnya benar, tanpa ada nilai di luar diagonal. Sebaliknya, tipe *Familial Hemiplegic Migraine* dan *Sporadic Hemiplegic Migraine* memiliki banyak prediksi salah, menunjukkan model memiliki kesulitan dalam mengklasifikasikan kelas-kelas ini dengan benar. Tipe *Basilar-Type Aura* dan *Typical Aura with Migraine* juga memiliki beberapa prediksi salah, tetapi sebagian besar prediksinya benar.

3.2 Pembahasan

Pada tahap pengklasifikasian, penelitian ini menggunakan metode Naïve Bayes untuk mengklasifikasikan jenis migrain berdasarkan dataset yang diperoleh dari Kaggle. Dataset ini mencakup berbagai informasi terkait gejala migrain, seperti usia, durasi, frekuensi, serta gejala lainnya, termasuk vertigo dan tinnitus. Sebelum data digunakan untuk pelatihan model, dilakukan tahap *data cleaning* untuk menghapus nilai yang hilang atau tidak valid (*NaN*). Proses ini sangat penting untuk memastikan bahwa model dilatih menggunakan data yang bersih dan representatif, sesuai dengan praktik terbaik dalam pengolahan data (Abyuda & Kurniawan, 2021). Setelah data dibersihkan, dilakukan visualisasi distribusi jenis migrain untuk memahami komposisi dataset. Hasil visualisasi menunjukkan bahwa dataset mencakup berbagai jenis migrain dengan jumlah kemunculan yang bervariasi, yang memberikan gambaran tentang ketidakseimbangan kelas dalam data. Ketidakseimbangan kelas ini penting untuk diperhatikan, karena dapat mempengaruhi kinerja model klasifikasi, sebagaimana ditemukan dalam penelitian lain yang menggunakan Naïve Bayes untuk klasifikasi penyakit (Komputer, Buatan, & Ridwan, 2020). Pada tahap pelatihan dan pengujian model, dataset dibagi menjadi dua bagian: 70% untuk data pelatihan dan 30% untuk data pengujian. Pembagian ini memungkinkan penilaian performa model dengan membandingkan hasil prediksi terhadap data yang tidak digunakan saat pelatihan. Model Naïve Bayes yang diterapkan dalam penelitian ini berhasil mencapai akurasi sebesar 89,02% pada tahap pengujian, yang menunjukkan performa yang cukup baik. Perbandingan dengan algoritma lain, seperti Multilayer Perceptron (MLP), yang memiliki akurasi lebih tinggi (91%), mengindikasikan bahwa meskipun Naïve Bayes lebih sederhana, metode ini tetap memberikan hasil yang kompetitif (Rindri & Fitriyani, 2024). Meskipun Naïve Bayes mengasumsikan independensi antar fitur, yang mungkin membatasi kinerjanya, keunggulan kesederhanaan algoritma ini dapat mengurangi potensi *overfitting*, terutama pada dataset dengan ukuran terbatas. Dalam evaluasi kinerja, metrik seperti *precision*, *recall*, dan *F1-score* digunakan untuk mengukur kualitas klasifikasi model. Tipe migrain seperti *Migraine without Aura* dan *Typical Aura*

without Migraine menunjukkan hasil yang sangat baik, dengan *precision* dan *recall* yang hampir sempurna, yang menandakan kemampuan model dalam mengklasifikasikan kelas ini dengan akurat. Namun, beberapa jenis migrain yang lebih jarang, seperti *Familial Hemiplegic Migraine* dan *Sporadic Hemiplegic Migraine*, menunjukkan hasil yang lebih buruk, dengan banyak prediksi yang salah. Hal ini menunjukkan tantangan dalam mengklasifikasikan kelas yang lebih langka atau kompleks, yang juga diamati dalam penelitian serupa (Kannan & Thangarasu, 2022).

Visualisasi *confusion matrix* lebih lanjut mengungkapkan bahwa model mengalami kesulitan dalam mengklasifikasikan beberapa kelas migrain, terutama tipe-tipe migrain yang lebih jarang. Meskipun demikian, tipe migrain yang lebih umum, seperti *Migraine without Aura* dan *Typical Aura without Migraine*, berhasil diklasifikasikan dengan akurasi yang sangat tinggi. Hal ini mengindikasikan bahwa model Naïve Bayes efektif dalam mengklasifikasikan jenis migrain yang lebih sering muncul, namun perlu peningkatan untuk mengatasi ketidakseimbangan kelas dan meningkatkan klasifikasi untuk tipe migrain yang jarang terjadi (Narayan & Khatoon, 2023). Hasil penelitian ini menunjukkan bahwa Naïve Bayes merupakan algoritma yang efektif dan efisien untuk klasifikasi migrain, dengan akurasi yang kompetitif meskipun terdapat tantangan dalam klasifikasi jenis migrain yang lebih jarang. Penerapan teknik tambahan, seperti *data augmentation*, atau penggunaan model yang lebih kompleks, seperti MLP atau Support Vector Machine (SVM), dapat meningkatkan akurasi, terutama untuk kelas migrain yang kurang terwakili dalam dataset (Khan *et al.*, 2024).

4. Kesimpulan

Hasil penelitian klasifikasi penyakit migrain menggunakan metode Naïve Bayes pada dataset yang diperoleh dari Kaggle menunjukkan hasil yang cukup baik. Model Naïve Bayes menghasilkan akurasi sebesar 88,51% pada tahap pelatihan dengan 280 data *training*, dan 89,02% pada tahap pengujian dengan 120 data *testing*. Hingga saat ini, belum ditemukan penelitian sebelumnya yang menggunakan metode Naïve Bayes untuk klasifikasi penyakit migrain dengan hasil serupa atau lebih tinggi. Dengan tingkat akurasi yang diperoleh, metode Naïve Bayes diharapkan dapat membantu masyarakat dan dunia medis dalam mengidentifikasi jenis migrain berdasarkan gejala dengan akurasi yang lebih baik. Hasil pemodelan ini, yang dilakukan menggunakan Python, juga dapat dikembangkan lebih lanjut untuk mendukung peningkatan akurasi pada penelitian klasifikasi migrain di masa depan. Untuk pengembangan selanjutnya, disarankan untuk menguji metode Naïve Bayes pada dataset yang lebih besar dan lebih beragam guna memvalidasi konsistensi akurasi model dalam klasifikasi penyakit migrain. Selain itu, penelitian berikutnya dapat mengeksplorasi kombinasi Naïve Bayes dengan teknik *pre-processing* data yang lebih canggih, seperti seleksi fitur atau teknik *oversampling* pada kelas minoritas, untuk mengurangi kemungkinan bias dan meningkatkan sensitivitas terhadap variasi gejala. Pendekatan lain yang bisa dipertimbangkan adalah perbandingan performa Naïve Bayes dengan algoritma lain, seperti *Random Forest* atau *Support Vector Machine* (SVM), serta menguji model pada data *real-time* yang diperoleh dari aplikasi klinis. Pengembangan-pengembangan ini diharapkan dapat meningkatkan akurasi dan generalisasi model, sehingga hasil klasifikasi migrain dapat diterapkan lebih efektif dalam medis.

5. Daftar Pustaka

- Abyuda, K. P. P., & Kurniawan, S. N. (2021). Complicated migraine. *Journal of Pain, Headache and Vertigo*, 2(2), 28-33.
- Bianto, M. A., Kusrini, K., & Sudarmawan, S. (2020). Perancangan Sistem Klasifikasi Penyakit Jantung Menggunakan Naïve Bayes. *Creative Information Technology Journal*, 6(1), 75-83. DOI: <https://doi.org/10.24076/citec.2019v6i1.231>.

Eigenbrodt, A. K., Ashina, H., Khan, S., Diener, H. C., Mitsikostas, D. D., Sinclair, A. J., ... & Ashina, M. (2021). Diagnosis and management of migraine in ten steps. *Nature Reviews Neurology*, 17(8), 501-514. DOI: <https://doi.org/10.1038/s41582-021-00509-5>.

Fadri, W. (2023). Klasifikasi Penyakit Hati dengan Menggunakan Metode Naive Bayes. *Jurnal Informasi Dan Teknologi*, 32-37. DOI: <https://doi.org/10.37034/jidt.v5i1.230>.

Iskandar, J. W., & Nataliani, Y. (2021). Perbandingan Naïve Bayes, SVM, dan k-NN untuk Analisis Sentimen Gadget Berbasis Aspek. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 5(6), 1120-1126. DOI: <https://doi.org/10.29207/resti.v5i6.3588>.

Kannan, K. N., & Thangarasu, G. KNOWLEDGE DATA ANALYSIS ON MIGRAINE HEADACHES BY USING OPTIMAL CLASSIFIERS.

Khan, L., Shahreen, M., Qazi, A., Jamil Ahmed Shah, S., Hussain, S., & Chang, H. T. (2024). Migraine headache (MH) classification using machine learning methods with data augmentation. *Scientific Reports*, 14(1), 5180. DOI: <https://doi.org/10.1038/s41598-024-55874-0>.

Lasulika, M. E. (2019). *Komparasi Naïve Bayes, Support Vector Machine dan K-Nearest Neighbor untuk mengetahui akurasi tertinggi pada prediksi kelancaran pembayaran TV kabel*. ILKOM Jurnal Ilmiah, 11(1), 11–16. DOI: <https://doi.org/10.33096/ilkom.v11i1.408.11-16>.

Marutho, D. (2019). Perbandingan Metode Naive Bayes, KNN, Decision Tree Pada Laporan Water Level Jakarta. *Jurnal Ilmiah Infokam*, 15(2). DOI: <https://doi.org/10.53845/infokam.v15i2.175>.

Narayan, R., & Khatoon, U. M. (2023). *Migraine detection and analysis using fuzzy intelligent system*. DOI: <https://doi.org/10.21203/rs.3.rs-2705833/v1>.

Noviana, R., & Rasal, I. (2023). Penerapan Algoritma Naive Bayes Dan Svm Untuk Analisis Sentimen Boy Band Bts Pada Media Sosial Twitter. *Jurnal Teknik dan Science*, 2(2), 51-60.

Paramitha, N. Y., Nuryaman, A., Faisol, A., Setiawan, E., & Nurvazly, D. E. (2023). Klasifikasi Penyakit Stroke Menggunakan Metode Naïve Bayes. *Jurnal Siger Matematika*, 11-16.

Ridwan, A. (2020). Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus. *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, 4(1), 15-21. DOI: <https://doi.org/10.47970/siskom-kb.v4i1.169>.

Rindri, Y. A., & Fitriyani, A. (n.d.). *Analisis perbandingan kinerja algoritma Multilayer Perceptron dan K-Nearest Neighbor pada klasifikasi tipe migrain*. *Jurnal Teknologi dan Informasi*. DOI: <https://doi.org/10.34010/jati.v13i1>.

Yunita, R., & Kamayani, M. (2023). *Perbandingan algoritma SVM dan Naïve Bayes pada analisis sentimen penghapusan kewajiban skripsi*. *Indonesian Journal of Computer Science*, 12(5). DOI: <https://doi.org/10.33022/ijcs.v12i5.3415>.