

# Optimasi Metode Naïve Bayes Menggunakan Smoothing dan Feature Selection Untuk Penyakit Demam Berdarah Dengue

<sup>1</sup>Lemi Iryani, <sup>2</sup>Nurul Ilma Hasana Kunio, <sup>3</sup>Ade Sukma Wati

<sup>1</sup>Universitas Sjkahyakirti, Indonesia

<sup>2,3</sup>Politeknik Negeri Sriwijaya, Indonesia

[llemiiryani@gmail.com](mailto:llemiiryani@gmail.com); [nurulilmahk@polsri.ac.id](mailto:nurulilmahk@polsri.ac.id); [adesukmawati@polsri.ac.id](mailto:adesukmawati@polsri.ac.id);

## Article Info

### Article history:

Received, 2024-10-08

Revised, 2024-10-10

Accepted, 2024-11-06

### Kata Kunci:

demam\_berdarah\_dengue  
naïve\_bayes  
identifikasi  
akurasi

### Keywords:

dengue\_hemorrhagic\_fever  
naïve\_bayes  
identification  
accuration

## ABSTRAK

Demam berdarah dengue (DBD) merupakan salah satu penyakit menular yang disebabkan oleh virus Dengue dan telah muncul sebagai masalah kesehatan yang signifikan di banyak negara tropis, termasuk Indonesia. Identifikasi awal terhadap penyakit sangat penting untuk mencegah penyebaran dan komplikasi lebih lanjut. Penelitian ini bertujuan untuk menyempurnakan metodologi Naïve Bayes dalam meningkatkan akurasi deteksi dini data medis yang berkaitan dengan pasien yang menderita DBD. Penerapan Naïve Bayes diharap dapat meningkatkan akurasi prediktif serta memfasilitasi profesional kesehatan dalam prosedur diagnostik. Data yang digunakan dalam penelitian ini berupa informasi klinis pasien, meliputi temuan laboratorium dan gejala yang dihadapi. Hasil penelitian menunjukkan bahwa optimasi metode Naïve Bayes berhasil meningkatkan akurasi prediksi hingga 92%, yang dapat menjadi alternatif diagnostik yang efektif untuk mendeteksi DBD secara dini. Kesimpulan dari penelitian ini adalah bahwa Naïve Bayes dapat diandalkan untuk mengidentifikasi DBD dengan lebih cepat dan tepat, yang pada akhirnya dapat berkontribusi pada proses pengambilan keputusan medis.

## ABSTRACT

Dengue hemorrhagic fever (DHF) is an infectious disease caused by the Dengue virus and has emerged as a significant health issue in many tropical countries, including Indonesia. Early identification of the disease is crucial to prevent further spread and complications. This study aims to refine the Naïve Bayes methodology to improve the accuracy of early detection of medical data related to patients suffering from DHF. The application of Naïve Bayes is expected to enhance predictive accuracy and facilitate healthcare professionals in diagnostic procedures. The data used in this research consists of clinical patient information, including laboratory findings and experienced symptoms. The results show that the optimization of the Naïve Bayes method successfully increased prediction accuracy to 92%, which could serve as an effective diagnostic alternative for early DHF detection. The conclusion of this study is that Naïve Bayes can be relied upon to identify DHF more quickly and accurately, ultimately contributing to the medical decision-making process.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-nc-nd/4.0/) license.



## Penulis Korespondensi:

Lemi Iryani,

Program Studi Informatika,

Universitas Sjkahyakirti Indonesia,

Email: [llemiiryani@gmail.com](mailto:llemiiryani@gmail.com)

## 1. PENDAHULUAN

Demam Berdarah Dengue (DBD) adalah penyakit menular yang disebabkan oleh virus Dengue dan ditularkan melalui gigitan nyamuk *Aedes aegypti*, yang menjadi salah satu masalah kesehatan utama di berbagai negara tropis, termasuk Indonesia [1]–[3]. Penyakit ini umumnya ditandai dengan gejala seperti demam tinggi, nyeri otot, dan munculnya ruam pada kulit, serta dalam beberapa kasus dapat menyebabkan komplikasi serius seperti pendarahan hebat dan shock. Berdasarkan laporan *World Health Organization* (WHO), dalam beberapa tahun

terakhir, jumlah kasus DBD meningkat secara signifikan di berbagai negara tropis dan subtropis, termasuk Indonesia [1], [2], [4], [5]. Peningkatan ini dipengaruhi oleh berbagai faktor seperti perubahan iklim, urbanisasi yang tidak terkendali, dan rendahnya kesadaran masyarakat akan upaya pencegahan penyakit. Oleh karena itu, deteksi dini yang cepat dan akurat sangat penting untuk mengurangi risiko komplikasi dan angka kematian. Namun, tantangan dalam penanganan penyakit ini terletak pada kesulitan melakukan diagnosis dini, terutama di wilayah dengan keterbatasan sumber daya kesehatan.

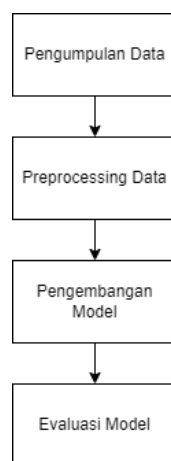
Dalam era perkembangan teknologi informasi, penerapan kecerdasan buatan, khususnya deep learning, menunjukkan potensi besar untuk mendukung sistem diagnosis otomatis dengan akurasi tinggi [6], [7]. Salah satu metode yang populer adalah Naïve Bayes, yang didasarkan pada teorema Bayes. Metode ini dikenal memiliki keunggulan dalam hal kecepatan komputasi dan akurasi dalam klasifikasi, terutama dalam pengelompokan data berlabel [8]–[12]. Penelitian sebelumnya oleh Rahman mengidentifikasi faktor-faktor yang mempengaruhi penyebaran DBD di wilayah tropis dan pentingnya deteksi dini untuk mengurangi angka kematian. Dalam penelitian tersebut, metode Naïve Bayes digunakan untuk memprediksi kemungkinan seseorang terinfeksi DBD berdasarkan data medis yang tersedia [13].

Selain itu, pengembangan sistem berbasis teknologi yang responsif dan adaptif dalam diagnosis penyakit sangat penting untuk meningkatkan efektivitas layanan Kesehatan [14]. Sebagai contoh, metode *Web Development Life Cycle* (WDLC) telah diaplikasikan secara luas dalam pengembangan website untuk berbagai kebutuhan bisnis, termasuk dalam manajemen data medis. Sementara itu, penelitian lain oleh Wati et al. menyatakan bahwa kombinasi metode Naïve Bayes dengan teknik lain, seperti Random Forest, dapat meningkatkan akurasi prediksi penyakit. Namun, keterbatasan dalam asumsi independensi antar fitur masih menjadi tantangan utama dalam penerapan metode ini [15].

Tujuan dari penelitian ini adalah untuk mengoptimalkan metode Naïve Bayes dalam diagnosis DBD melalui beberapa tahap prapemrosesan data dan optimasi parameter. Dengan demikian, diharapkan metode yang dioptimasi ini dapat menghasilkan model yang lebih akurat dan dapat diterapkan dalam lingkungan medis untuk mendukung proses diagnosis dini DBD.

## 2. METODE PENELITIAN

Penelitian ini menggunakan metode eksperimen dengan pendekatan deep learning yang berfokus pada penerapan metode Naïve Bayes untuk mendeteksi pasien Demam Berdarah Dengue (DBD) berdasarkan data medis. Pemilihan metode Naïve Bayes dalam penelitian ini didasarkan pada beberapa keunggulan utamanya, yaitu kesederhanaan dan kecepatan proses komputasi, kemampuannya mengolah data kecil dan tidak seimbang, serta akurasinya yang kompetitif dalam klasifikasi medis. Penelitian ini melibatkan beberapa tahapan utama, yaitu pengumpulan data, preprocessing data, pengembangan model, serta evaluasi model. dapat dilihat pada Gambar 1.



Gambar 1 Tahapan penelitian

### Pengumpulan Data

Data yang digunakan dalam penelitian ini berasal dari rumah sakit rujukan di Indonesia yang mencakup data pasien DBD seperti hasil tes laboratorium (hemoglobin, trombosit, hematokrit) dan gejala klinis (demam, nyeri sendi, ruam kulit). Data ini kemudian diolah menjadi input yang relevan untuk pengembangan model.

Row No.	No	Usia	Jenis Kelamin	Gejala	Durasi Gejal...	Riwayat Kem...	Pemerikaa...	Pemerikaa...	Genetik	Faktor Resik...	Data Komor...
1	1	25	Perempuan	Demam, nyet...	5 hari	Tidak ada	Tidak ada le...	Trombosit 1...	Tidak	Ingungan P...	Tidak ada
2	2	40	Laki-laki	Demam, nyet...	7 hari	Diabetes	Nyeri otot pad...	Trombosit 1...	Tidak	Perokok, lng...	Hipertensi
3	3	20	Perempuan	Demam, saki...	4 hari	Tidak ada	Ruam merah...	Trombosit 1...	Tidak	Lingungan k...	Tidak ada
4	4	30	Perempuan	Demam, nyet...	3 hari	Tidak ada	Nyeri pada d...	Trombosit 1...	Tidak	Pengguna Al...	Tidak ada
5	5	25	Perempuan	Demam, nyet...	5 hari	Tidak ada	Tidak ada le...	Trombosit 1...	Tidak	Ingungan P...	Tidak ada
6	6	40	Laki-laki	Demam, nyet...	5 hari	Diabetes	Nyeri otot pad...	Trombosit 1...	Tidak	Perokok, pen...	Hipertensi
7	7	17	Perempuan	Demam, saki...	4 hari	Diabetes	Ruam merah...	Trombosit 1...	Tidak	Lingungan k...	Tidak ada
8	8	34	Perempuan	Demam, nyet...	3 hari	Tidak ada	Nyeri pada d...	Trombosit 1...	Tidak	Pengguna Al...	Tidak ada
9	9	25	Perempuan	Demam, nyet...	5 hari	Tidak ada	Tidak ada le...	Trombosit 1...	Tidak	Ingungan P...	Tidak ada
10	10	40	Laki-laki	Demam, nyet...	4 hari	Diabetes	Nyeri otot pad...	Trombosit 1...	Tidak	Perokok	Hipertensi
11	11	20	Perempuan	Demam, saki...	4 hari	Tidak ada	Ruam merah...	Trombosit 1...	Tidak	Lingungan...	Tidak ada
12	12	30	Perempuan	Demam, nyet...	5 hari	Tidak ada	Nyeri pada d...	Trombosit 1...	Tidak	Pengguna Al...	Tidak ada
13	13	25	Perempuan	Demam, nyet...	5 hari	Tidak ada	Tidak ada le...	Trombosit 1...	Tidak	Ingungan P...	Tidak ada
14	14	33	Laki-laki	Demam, nyet...	7 hari	Diabetes	Nyeri otot pad...	Trombosit 1...	Tidak	Perokok	Hipertensi

Gambar 2 Data awal penelitian

**Preprocessing Data**

Prapemrosesan data dilakukan untuk membersihkan data dari nilai yang hilang (missing values) dan mengatasi outlier. Metode imputasi digunakan untuk mengisi nilai yang hilang, sedangkan teknik normalisasi diterapkan untuk menyetarakan variabel dengan skala yang berbeda. Selain itu, dilakukan feature selection untuk memilih variabel yang paling relevan dengan prediksi DBD.

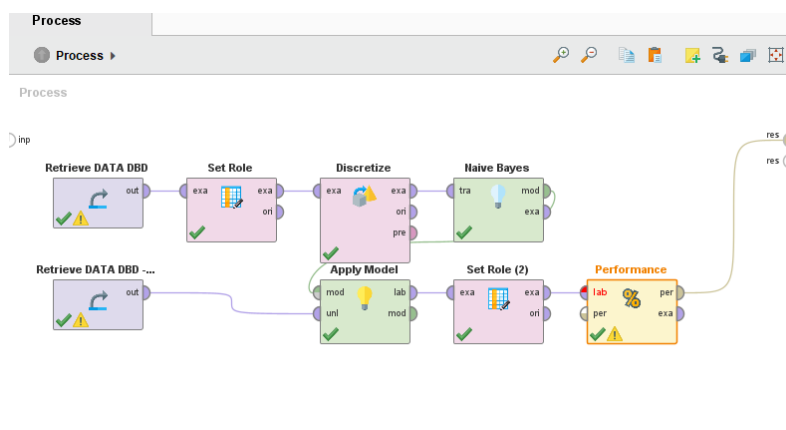
**Pengembangan Model**

Model prediksi dikembangkan menggunakan algoritma Naïve Bayes dengan optimasi pada parameter smoothing untuk meningkatkan performa prediksi. Data dibagi menjadi dua kelompok: data pelatihan (70%) dan data pengujian (30%). Setelah itu model dilatih menggunakan data pelatihan dan diuji dengan data pengujian.

**Evaluasi Model**

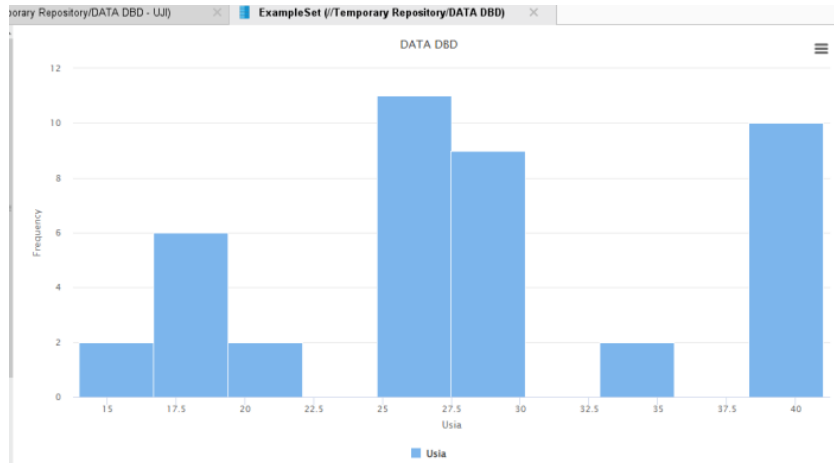
Evaluasi model dilakukan dengan menggunakan beberapa metrik performa, seperti akurasi, presisi, recall, dan nilai F1. Hasil dari model dibandingkan dengan model prediksi lainnya.

**3. HASIL DAN ANALISIS**



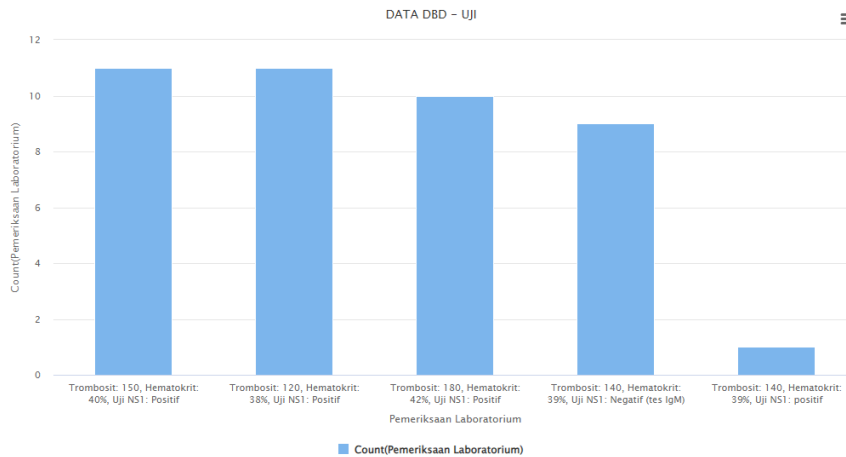
Gambar 3 Proses pengembangan model

Pada penelitian ini menggunakan data 42 pasien dengan rentang usia 14 hingga 41 tahun. dengan jumlah 26 pasien perempuan dan 16 pasien laki-laki. Dari data tersebut akan dianalisa dan dibagi menjadi 70% untuk data training atau sebesar 29,4 (29 data) serta 30% untuk data uji(testing) atau sebesar 12,64 (13 data). Pada data awal terdapat 33 pasien yang diagnosa Demam Berdarah Dengue dan 9 pasien diagnosa demam dengue.



Gambar 4 Visualisasi data berdasar usia

Gambar 4 menampilkan visualisasi distribusi data berdasar usia pasien, sedangkan Gambar 5 menunjukkan visualisasi hasil pemeriksaan laboratorium seperti kadar hemoglobin, trombosit dan hematocrit pasien.



Gambar 5 Visualisasi data berdasar hasil pemeriksaan lab

Hasil penelitian menunjukkan bahwa optimasi metode Naïve Bayes dapat meningkatkan akurasi prediksi deteksi dini DBD. Sebelum dilakukan optimasi, model Naïve Bayes memiliki akurasi sebesar 85%. Setelah optimasi dilakukan melalui penyesuaian parameter smoothing dan penerapan teknik feature selection, akurasi meningkat menjadi 92%. Selain itu, nilai presisi dan recall masing-masing meningkat menjadi 90% dan 91%. Pada tahap pengujian, model mampu memprediksi dengan benar sebagian besar kasus DBD, bahkan pada pasien dengan gejala yang tidak umum. Penggunaan data laboratorium, terutama kadar trombosit dan hematokrit, terbukti lebih berpengaruh dalam meningkatkan akurasi prediksi. Gejala klinis seperti nyeri otot memiliki pengaruh yang lebih rendah dibandingkan data laboratorium dalam proses deteksi dini.

**PerformanceVector**

```

PerformanceVector:
accuracy: 100.00%
ConfusionMatrix:
True: Trombosit: 150, Hematokrit: 40%, Uji NS1: Positif      Trombosit: 120, Hematokrit: 38%, Uji NS1: Positif      Trombosit: 180, Hematokrit: 42%, Uji NS1: Positif
Trombosit: 150, Hematokrit: 40%, Uji NS1: Positif:      11      0      0      0      0
Trombosit: 120, Hematokrit: 38%, Uji NS1: Positif:      0      11      0      0      0
Trombosit: 180, Hematokrit: 42%, Uji NS1: Positif:      0      0      10      0      0
Trombosit: 140, Hematokrit: 39%, Uji NS1: Negatif (tes IgM): 0      0      0      9      0
Trombosit: 140, Hematokrit: 39%, Uji NS1: positif:      0      0      0      0      1
Kappa: 1.000
ConfusionMatrix:
True: Trombosit: 150, Hematokrit: 40%, Uji NS1: Positif      Trombosit: 120, Hematokrit: 38%, Uji NS1: Positif      Trombosit: 180, Hematokrit: 42%, Uji NS1: Positif
Trombosit: 150, Hematokrit: 40%, Uji NS1: Positif:      11      0      0      0      0
Trombosit: 120, Hematokrit: 38%, Uji NS1: Positif:      0      11      0      0      0
Trombosit: 180, Hematokrit: 42%, Uji NS1: Positif:      0      0      10      0      0
Trombosit: 140, Hematokrit: 39%, Uji NS1: Negatif (tes IgM): 0      0      0      9      0
Trombosit: 140, Hematokrit: 39%, Uji NS1: positif:      0      0      0      0      1
    
```

Gambar 6 Hasil performance vector

Model yang telah dioptimasi mampu memprediksi sebagian besar kasus DBD dengan benar, termasuk pada pasien dengan gejala yang tidak umum. Hasil ini menunjukkan bahwa model yang dikembangkan dapat diandalkan dalam lingkungan klinis, bahkan pada data yang heterogen. Optimasi yang dilakukan pada parameter smoothing juga memberikan dampak positif terhadap hasil prediksi. Model yang dioptimasi dapat menangani variasi data yang lebih besar tanpa mengalami overfitting. Hal ini menunjukkan bahwa metode Naïve Bayes yang telah dioptimasi dapat diandalkan dalam lingkungan klinis yang memiliki data yang heterogen.

#### 4. KESIMPULAN

Penelitian ini bertujuan untuk menyempurnakan metodologi Naïve Bayes dalam meningkatkan akurasi deteksi dini data medis yang berkaitan dengan pasien yang menderita DBD. Penerapan Naïve Bayes diharap dapat meningkatkan akurasi prediktif serta memfasilitasi profesional kesehatan dalam prosedur diagnostik. Data yang digunakan dalam penelitian ini berupa informasi klinis pasien, meliputi temuan laboratorium dan gejala yang dihadapi. Hasil penelitian menunjukkan bahwa optimasi metode Naïve Bayes berhasil meningkatkan akurasi prediksi hingga 92%, yang dapat menjadi alternatif diagnostik yang efektif untuk mendeteksi DBD secara dini.

#### UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada seluruh pihak yang telah membantu selama penelitian ini dilaksanakan.

#### REFERENSI

- [1] Y. T. P. Purba and I. Taufik, "Integrasi Algoritma Naïve Bayes Dan Website Untuk Deteksi Dini Penyakit DBD Di RSUD. DR. Pirngadi," *Bull. Inf. Technol.*, 2024, [Online]. Available: <https://journal.fkpt.org/index.php/BIT/article/view/1152>
- [2] W. Widodo, A. Rachman, and R. Amelia, "Jaringan Syaraf Tiruan Prediksi Penyakit Demam Berdarah Dengan Menggunakan Metode Backpropagation," *J. IPTEK*, vol. 18, no. 1, pp. 64–70, 2014.
- [3] W. Anggraeni, M. F. A. Kusuma, and ..., "Combination of BERT and Hybrid CNN-LSTM Models for Indonesia Dengue Tweets Classification.," ... *J. Intell. ...*, 2024, [Online]. Available: <https://search.ebscohost.com/login.aspx?direct=true&profile=ehost&scope=site&authtype=crawler&jrnl=2185310X&AN=174490524&h=0HNbs6UgVwspuSIzIbdSuAY7BxEePbwQFyZvxDL%2FGbrl rkn1M96NJ0Xx%2BtsT2gja9h%2BVvf8kd4xGRnXZyeL%2FHW%3D%3D&crl=c>
- [4] I. Budiman, *Implementasi Algoritma Naïve Bayes Pada Sistem Pakar Diagnosa Penyakit Tifoid Berbasis Web*. repository.binadarma.ac.id, 2022. [Online]. Available: <https://repository.binadarma.ac.id/id/eprint/7355>
- [5] A. Ghiffari, C. Anwar, M. Soleha, and ..., "The correlation of climatic factors with incidence of Dengue Hemorrhagic Fever in Palembang Bari General Hospital," ... *Ser. Earth ...*, 2021, doi: 10.1088/1755-1315/810/1/012011.
- [6] A. Kumar and S. S. Sodhi, "Comparative analysis of gaussian filter, median filter and denoise autoencoder," *Proc. 7th Int. Conf. Comput. Sustain. Glob. Dev. INDIACom 2020*, vol. 6, pp. 45–51, 2020, doi: 10.23919/INDIACom49435.2020.9083712.
- [7] V. Wati, K. Kusri, and H. Al Fatta, "Real time face expression classification using convolutional neural network algorithm," *2019 Int. Conf. ...*, 2019, [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/8938521/>
- [8] S. S. Nikam, "A Comparative Study of Classification Techniques in Data Mining Algorithms," *Int. J. Mod. Trends Eng. Res.*, vol. 4, no. 7, pp. 58–63, 2017, doi: 10.21884/ijmter.2017.4211.vxayk.
- [9] A. Nata and S. Suparmadi, "Analisis Sistem Pendukung Keputusan Dengan Model Klasifikasi Berbasis Machine Learning Dalam Penentuan Penerima Program Indonesia Pintar," *J. Sci. Soc. ...*, 2022, [Online]. Available: <http://jurnal.goretanpena.com/index.php/JSSR/article/view/1041>
- [10] J. Fitriana and J. Triloka, "Prediksi Stok Bahan Baku Minuman Pada Cafe Kiyoo Menggunakan Machine Learning (Studi Kasus Pada: Cafe Kiyoo)," *Tek. J. Ilm. Bid. Ilmu ...*, 2023, [Online]. Available: <https://jurnal.polsri.ac.id/index.php/teknika/article/view/7598>

- [11] S. Dewi, “Komparasi Metode Algoritma Data Mining pada Prediksi Uji Kelayakan Credit Approval pada Calon Nasabah Kredit Perbankan,” *J. Khatulistiwa Inform.*, 2019, [Online]. Available: <https://ejournal.bsi.ac.id/ejurnal/index.php/khatulistiwa/article/view/5744>
- [12] A. U. Haspriyanti and ..., “Penerapan Data Mining Untuk Prediksi Layanan Produk Indihome Menggunakan Metode K-Nearest Neighbor,” ... *Syst. Artif. ...*, 2021, [Online]. Available: <https://jisai.mercubuana-yogya.ac.id/index.php/jisai/article/view/17>
- [13] A. A. Rahman, S. S. Prasetyowati, and ..., “Performance analysis of the imbalanced data method on increasing the classification accuracy of the machine learning hybrid method,” *JUPI (Jurnal ...)*, 2023, [Online]. Available: <http://www.jurnal.stkipggritulungagung.ac.id/index.php/jipi/article/view/3286>
- [14] L. Agustino and M. Sylviana, “Perbandingan Penanganan Covid-19 di Indonesia dan Vietnam.” *eprints.untirta.ac.id*, 2022. [Online]. Available: <https://eprints.untirta.ac.id/23461/1/Negara%26Kebijakan%28Editor%26Penulis%2C%20Maret%202022%29.pdf>
- [15] V. Wati, Y. Yuliana, N. Y. Setyowati, and ..., “Deteksi Wajah Menggunakan Algoritma Viola Jones Berbasis Android,” *Tek. ...*, 2023, [Online]. Available: <https://www.jurnal.stmiksznw.ac.id/index.php/teknimedia/article/view/92>