



DETEKSI DENGUE SHOCK SYNDROME PADA PASIEN DEMAM BERDARAH MENGUNAKAN ALGORITMA RANDOM FOREST

Lisa N. Nisa¹, Lailil Muflikhah², Novanto Yudistira³

^{1,2,3}Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya

¹lisaliza0104@gmail.com

Abstrak

Demam Berdarah Dengue (DBD) merupakan penyakit yang ditularkan oleh nyamuk yang paling sering ditemui di dunia. DBD disebarkan oleh virus nyamuk *Aedes aegypti* dan *Aedes Albopictus*. DBD memiliki gejala klinis ringan hingga berat seperti Dengue Shock Syndrom, Dengue Shock Syndrom atau DSS adalah syok/ renjatan disertai kegagalan sirkulasi. Penanganan DSS apabila tidak ditangani lebih awal dapat membahayakan nyawa pasien oleh karena penting untuk membuat deteksi Dengue Shock Syndrom pada pasien DBD. Penelitian pada dengue shock syndrom pada pasien demam berdarah menggunakan algoritma random forest karena kemampuannya mengurangi jumlah variabel yang diperlukan untuk mendeteksi, meningkatkan efisiensi. Data yang akan digunakan dalam penelitian ini berasal dari data pasien penderita demam dengue Rumah Sakit Umum Daerah dr. Saiful Anwar tahun 2014 sampai 2022 yang berjumlah 501 data. Tetapi terdapat permasalahan pada data yang dipakai yaitu dataset imbalanced. Karena data yang digunakan merupakan data imbalanced maka penulis menggunakan random oversampling, smote. Setelah di oversampling data menjadi 401 data syock dan 401 data tidak syock. Kemudian penulis memilih feature yang relevan untuk membangun model random forest. Feature selection yang digunakan adalah Principal component analysis (PCA) dan information gain. Pengujian pada penelitian ini mencari hyperparameter, metode oversampling dan ekstraksi fitur yang optimal. Hasil penelitian menunjukkan algoritma menggunakan PCA dan random oversampling memiliki hasil terbaik. Akurasi hasil pengujian tertinggi pada pada 0,911 dengan metode oversampling dan PCA.

Kata kunci: Demam Berdarah Dengue, Dengue Shock Syndrom, Principal Component Analysis, SMOTE, Oversampling, Information gain, Hyperparameter

Abstract

Dengue Hemorrhagic Fever (DHF) is the most common mosquito-borne disease in the world. Dengue fever viruses are spread by the *Aedes aegypti* and *Aedes Albopictus* mosquitoes. DHF has mild to severe clinical symptoms such as Dengue Shock Syndrome, Dengue Shock Syndrome, or DSS, which is a shock

Article History

Received: September 2024

Reviewed: September 2024

Published: September 2024

Plagiarism Checker No 234

Prefix DOI : Prefix DOI :

10.8734/Kohehi.v1i2.365

Copyright : Author

Publish by : Kohehi



This work is licensed
under a [Creative
Commons Attribution-
NonCommercial 4.0
International License](https://creativecommons.org/licenses/by-nc/4.0/)



accompanied by circulatory failure. If DSS is not treated early, it can endanger the patient's life because it is important to make predictions about Dengue Shock Syndrome in DHF patients. Research on dengue shock syndrome in dengue fever patients uses the random forest algorithm because of its ability to reduce the number of variables needed for prediction, increasing efficiency. The data used in this research comes from patients suffering from dengue fever at the Regional General Hospital, dr. Saiful Anwar from 2014 to 2022, totaling 501 data. However, there is a problem with the data used, namely the imbalanced dataset. Because the data is imbalanced, the author uses random oversampling and SMOTE. After oversampling, the data became 401 shock data and 401 non-shock data. The author selects relevant features to build a random forest model. The feature selection used is Principal component analysis (PCA) and information gain. The tests in this research look for hyperparameters, oversampling methods, and optimal feature extraction. The research results show that the algorithm using PCA and random oversampling has the best results. The highest accuracy of test results is 0.911 with the oversampling and PCA methods.

Keywords: *Dengue Hemorrhagic Fever, Dengue Shock Syndrome, Principal Component Analysis, SMOTE, Oversampling, Information gain, Hyperparameter*

1. PENDAHULUAN

Demam Berdarah Dengue (DBD) merupakan penyakit yang ditularkan oleh nyamuk yang paling sering ditemui di dunia, WHO mencatat bahwa 2,5 miliar orang tinggal di tempat virus ini bersirkulasi. Penelitian memaparkan infeksi Virus Dengue atau DENV 70% berada di Asia (Prasetya, et al., 2017).

Virus Dengue dibawa oleh nyamuk *Aedes aegypti* dan *Aedes albopictus*. Kedua nyamuk tersebar di seluruh Indonesia dengan ketinggian maksimum 1000 meter di atas permukaan air laut. Wilayah Indonesia memiliki risiko terjangkit Demam Berdarah Dengue. Penyakit Demam Berdarah Dengue memiliki risiko tinggi penularan saat musim penghujan (Sukohar, 2014).

Jumlah penderita DBD mengalami kenaikan dari tahun ke tahun, WHO menyebutkan penderita DBD meningkat 8 kali dalam dua dekade terakhir. Pada tahun 2000 jumlah penderita sebanyak 505.430 dan pada tahun 2010 menjadi 2,4 juta jiwa (Podung, et al., 2021). WHO memaparkan bahwa setiap tahunnya 50 juta orang terjangkit DBD 2, 5% diantaranya meninggal dunia. Setiap tahun, Asia menempati urutan pertama dalam jumlah orang yang terinfeksi DBD. Sedangkan pada tahun 1968 hingga 2009, Indonesia tercatat sebagai negara yang memiliki kasus DBD tertinggi di Asia Tenggara (Khaidir, et al., 2022)

Data Kementerian Kesehatan (Kemenkes) menyebutkan 103.781 kasus DBD di pada tahun 2020, 53,11% kasus DBD terjadi pada laki-laki sedangkan 46,89% terjadi pada perempuan. Badan Pusat Statistik memaparkan angka penderita DBD di Indonesia mencapai angka 40 orang setiap 100.000 penduduk (Dihni, 2022). Pada 2022, Kemenkes mencatat bahwa kasus demam berdarah mencapai 131.265 kasus, dimana sekitar 40% adalah anak-anak berusia 0 hingga 14 tahun. Sedangkan jumlah kematian mencapai 1.135 kasus, dengan 73% terjadi pada anak berusia 0-14 tahun. (Indonesia, 2023)



DBD memiliki gejala klinis ringan hingga berat seperti Dengue Shock Syndrom, Dengue Shock Syndrom atau DSS adalah syok/ renjatan disertai kegagalan sirkulasi. 30% sampai 50% penderita DBD mengalami DSS dan berakhir dengan kematian apabila tidak ditangani dengan dini (Raveendran & Budiarta, 2016).

Kewaspadaan terhadap DSS harus diperhatikan secara serius karena jumlah kematian pasien DBD dengan DSS sepuluh kali lebih banyak daripada pasien tanpa DSS (Pujiarti, 2016). Penelitian yang dilakukan oleh Tantracheewathorn dkk mengemukakan bahwa terdapat beberapa faktor resiko terjadinya DSS antara lain pasien dapat mengalami infeksi dengue sekunder, hemokonsentrasi dan pendarahan. Dalam penelitian lain juga disebutkan jumlah trombosit berpengaruh dalam terjadinya DSS (Salsabila, et al., 2017).

Penanganan DSS apabila tidak ditangani lebih awal dapat membahayakan nyawa pasien oleh karena penting untuk membuat prediksi Dengue Shock Syndrom pada pasien DBD.

2. DASAR TEORI

2.1 Demam Berdarah Dengue

Demam Berdarah Dengue (DBD) merupakan penyakit yang disebabkan oleh virus dengue yang termasuk famili *Flaviridae* dan genus *Flavivirusdan*. DBD disebarkan oleh virus nyamuk *Aedes aegypti* dan *Aedes Albopictus*. Masa inkubasi virus dengue berkisar 3- 14 sampai gejala muncul. Gejala awal DBD berupa demam, pendarahan diatesis yaitu uji tourniquet positif, trombositopenia atau jumlah keping darah (trombosit rendah) dan kebocoran plasma (Candra, 2010).

2.2 Dengue Shock Syndrome

Dengue Shock Syndrome (DSS) adalah syok pada penderita demam Berdarah Dengue yang terjadi di fase ketiga dan keempat. Terdapat 4 variabel yang mempengaruhi DSS yaitu umur lebih dari 15 tahun, infeksi ulangan, hepatomegaly dan trombositopenia kurang dari 50.000/ μ L (Prasetya, et al., 2017). DSS memiliki gejala klinis seperti demam, suhu tubuh ≥ 37 , lemas, gusi berdarah dan kesadaran menurun (Raveendran & Budiarta, 2016).

2.3 Random Forest

Random Forest adalah metode machine learning yang digunakan untuk menyelesaikan permasalahan, Random tree merupakan metode Ensemble learning yang menggabungkan beberapa *Decision Tree*. (Cutler, et al., 2011). Random forest adalah algoritma supervisi yang digunakan untuk klasifikasi (classification) dan regresi (regression). Dalam random forest decision tree dilatih pada subset yang berbeda dari data latih dan subset acak dari fitur-fitur pada setiap tree. Data latih dalam *Random Forest* dipilih secara acak. Klasifikasi pada Random Forest dilakukan dengan pemilihan data latih secara acak, kemudian membangun Decision Tree dan melakukan proses pelatihan dan training selanjutnya dilakukan pengujian menggunakan data yang tidak digunakan pada proses trainin. Terakhir menentukan kelas data uji menggunakan voting.

2.4 Synthetic Minority Over-sampling Technique (SMOTE)

Smote adalah salah satu metode untuk mengatasi data imbalanced. Metode SMOTE merupakan metode yang mengembangkan oversampling, yaitu dengan mengambil sampel baru dari kelas minoritas untuk membuat proporsi data menjadi lebih seimbang. SMOTE mengambil contoh pada kelas minoritas dan kemudian mencari k-nearest neighbor pada setiap instance, lalu menghasilkan instance sintetik sehingga dapat menghindari masalah overfitting. (Sutoyo & Fadlurrahman, 2020)



2.5 K-Fold Cross Validation

K-Fold Cross Validation merupakan salah satu metode untuk pengujian metode klasifikasi. K-Fold Cross Validation membagi dataset menjadi k subset berukuran sama. model dievaluasi pada subset ke-k pada set pengujian. Performa model diukur berdasarkan metrik tertentu (misalnya akurasi, presisi, recall, f1-score). Performa model dievaluasi untuk setiap iterasi, dan hasilnya dirata-ratakan untuk mendapatkan perkiraan performa model secara keseluruhan.

2.6 Principal component analysis (PCA)

PCA atau Principal component analysis adalah suatu metode statistic yang berguna untuk mengekstraksi fitur tabel pada dataset menjadi fitur-fitur yang penting atau disebut juga principal components. Principal component memiliki kombinasi yang linear dari variabel asal yang dapat menjelaskan secara maksimal variansi seluruh variabel. (Greenacre, et al., 2022)

2.7 Information Gain

Information gain merupakan salah satu metode yang digunakan untuk seleksi fitur, information gain dapat menunjukkan fitur yang paling relevan terhadap kelas target. Information gain dapat mengurutkan fitur yang paling relevan. (Nur, et al., 2022) Information gain dihitung dengan entropy parent dikurangi dengan entropy child.

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} \text{Entropy}(S_i)$$

Keterangan:

S = himpunan kasus

A = atribut

n = jumlah partisi atribut A

|S_i| = jumlah kasus pada partisi ke - i

|S| = jumlah kasus dalam S

$$\text{Entropy}(S) = -\sum_{j=1}^n p_i \cdot \log_2 p_i$$

Keterangan:

S = himpunan kasus

n = jumlah partisi

p_i = proporsi S_i terhadap S

3. METODOLOGI PENELITIAN

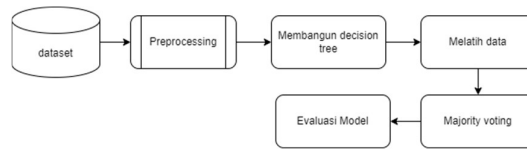
3.1 Tipe Penelitian

Tipe penelitian yang digunakan bersifat nonimplementatif dengan pendekatan analitis. Penelitian nonimplementatif menekankan pada variabel yang digunakan pada sebuah fenomena atau keadaan tertentu untuk menghasilkan tinjauan ilmiah. Penelitian nonimplementatif bisa dilakukan berdasarkan tinjauan dari ilmu yang sudah ada. Pendekatan analitis merupakan penelitian yang menjelaskan hubungan antara objek penelitian dengan fenomena yang sedang diteliti. Sehingga hasil analisa tersebut dapat menjawab pertanyaan pada rumusan masalah.



3.2 Strategi Penelitian

Penelitian ini digolongkan sebagai penelitian eksperimen. Tujuan dari penelitian eksperimen ini yaitu pengaruh penggunaan algoritme random forest pada prediksi Dengue shock syndrome pada pasien demam berdarah dengue.



Gambar 1. Contoh penggunaan gambar

Pada preprocessing dilakukan imputasi data kosong menggunakan modus dan knni. Pengisian menggunakan modus dilakukan pada data kategorikal dan pengisian dengan knni pada data numerik. Selanjutnya dilakukan oversampling data menggunakan SMOTE. Membangun decision tree dilakukan dengan mengurutkan information gain dan diambil fitur yang paling relevan sebagai root node. Iterasi selanjutnya hingga mencapai leaf node. Data uji dimasukkan dalam tree dan dicari kelas terbanyak.

3.3 Data Penelitian

Data yang akan digunakan dalam penelitian ini berasal dari data pasien penderita demam dengue Rumah Sakit Umum Daerah dr. Saiful Anwar tahun 2014 sampai 2022 yang berjumlah 501 data. Variabel yang digunakan dalam penelitian ini memiliki 40 fitur, yang meliputi gejala, kondisi fisik dan hasil test laboratorium pasien.

3.4 Teknik Pengujian dan Analisis Data

Tahap dalam pengujian dilakukan untuk mengetahui performa dari metode Random Forest dalam melakukan prediksi Dengue Shock Syndrome pada pasien Demam Berdarah Dengue. Pengujian meliputi analisis pengaruh jumlah decision tree yang digunakan terhadap akurasi. Pengujian dilakukan menggunakan metode K-fold cross-validation.

4. HASIL DAN PEMBAHASAN

Hyperparameter yang diuji adalah minimal sample split, banyak pohon dan maksimal kedalaman. sebanyak 55 kali dengan kombinasi minimal sample split, kedalaman maksimal dan jumlah pohon keputusan yang berbeda beda. Rata-rata hasil pengujian terdapat pada Tabel 1.

Tabel 1. Hasil pengujian jumlah pohon

Jumlah pohon	Rata-rata Accuracy
50	0,881412
100	0,883585
150	0,882401
200	0,885501
250	0,889239

Hasil pengujian rata rata akurasi tertinggi terdapat pada tree 250 yaitu 0,89 precision 0,92 recall 0,84 dan f1-score 0,88. Hyperparameter jumlah pohon Keputusan Jumlah pohon keputusan yang lebih banyak tidak selalu menghasilkan performa yang lebih baik tetapi menambah waktu komputasi.



Tabel 2. Hasil pengujian minimal sample split

Minimal sample split	Rata-rata Accuracy
3	0,884874
5	0,875826
7	0,882088

Hasil pengujian pada minimal sample split memiliki akurasi dari 0,864281 sampai 0,896996. Minimal sample dengan jumlah 3 memiliki rata-rata akurasi yang lebih baik daripada jumlah lain. Semakin kecil minimal sample split artinya node akan lebih sering terbagi hingga jumlah sample dalam node sama dengan minimal sample split mengakibatkan pohon keputusan akan semakin rumit dan meningkatkan akurasi, kemudian rata-rata akurasi hasil pengujian maksimal kedalaman terdapat pada Tabel 3.

Tabel 3. Hasil pengujian maksimal kedalaman

Maksimal Kedalaman	Rata-rata Accuracy
5	0,885484
7	0,896749
9	0,904537

Hasil pengujian pada maksimal kedalaman memiliki akurasi dari 0,872045 hingga 0,914244. Pada hasil pengujian maksimal kedalaman terbanyak memiliki akurasi lebih baik. Maksimal kedalaman yang memiliki nilai akurasi terbaik adalah 9. Hyperparameter maksimal kedalaman dengan jumlah yang lebih besar memiliki akurasi yang lebih besar. Semakin banyak jumlah maksimal kedalaman menandakan bahwa pohon keputusan semakin dalam menyebabkan masing-masing pohon keputusan menjadi lebih detail, tetapi memiliki waktu komputasi yang lebih lama.

Pengujian pada extraction fitur menggunakan metode metode PCA dan information gain. Extraction fitur pada jumlah fitur 20,25,30,35 dan 40. Rata-rata akurasi hasil pengujian metode ekstraksi fitur terdapat pada Tabel 4.

Tabel 4. Hasil pengujian metode PCA dan informasi gain

Metode	Jumlah fitur	Rata-rata Accuracy
PCA	20	0,911342
	25	0,903287
	30	0,899816
	35	0,907972
	40	0,904537
	20	0,899547
Information Gain	25	0,905453
	30	0,897357
	35	0,911080
	40	0,909186



Pada model random forest ekstraksi fitur menggunakan PCA memiliki performa yang lebih tinggi dibandingkan information gain, information gain memeringkat tingkat relevansi terhadap kelas. Information gain digunakan untuk seberapa baik dalam memisahkan kelas atau target dan seberapa informatif fitur pada kelas. Setelah memeringkat tingkat relevansi kelas information gain dapat mempersingkat waktu komputasi lebih cepat.

Sedangkan PCA tidak terkait dengan kelas tetapi pada data itu sendiri, PCA digunakan untuk mengurangi dimensi data tanpa kehilangan informasi penting. PCA dapat menghilangkan noise dalam dataset. PCA memilih fitur-fitur yang relevan dan menghilangkan fitur-fitur yang tidak relevan. Fitur yang relevan bisa meningkatkannya keakuratan pohon keputusan (Nasution, et al., 2018). Dalam algoritma random forest menggunakan information gain untuk mengurutkan fitur yang paling relevan, sehingga saat menggunakan PCA, dataset diekstraksi fitur dan kemudian dilakukan pemeringkatan relevansi fitur terhadap kelas.

Pengujian pada oversamplinRata-rata akurasi hasil pengujian metode untuk menangani data imbalance terdapat pada Tabel 5.

Tabel 5. Hasil pengujian metode oversample

Oversample	Rata-rata Accuracy
Oversampling	0,911342
SMOTE	0,850228
Tanpa oversampling	0,808

Oversampling memiliki peforma yang lebih baik daripada SMOTE karena oversampling cocok pada dataset yang sederhana seperti memiliki kelas tidak seimbang, hubungan antara fitur dapat dijelaskan dapat pola yang sederhana dan tidak memiliki variasi yang kompleks. Pada dataset yang digunakan dalam penelitian, metode oversampling yang tidak kompleks memungkinkan model random forest mempelajari pola lebih baik daripada Teknik oversampling yang kompleks. Pengambilan sampel yang berlebihan dan pemilihan fitur meningkatkan kinerja klasifikasi algoritma hutan acak pada data yang tidak seimbang. Oversampling dapat meningkatkan performa algoritma pada dataset imbalance dibandingkan dataset tanpa oversampling (Chamidah, et al., 2020).

5. KESIMPULAN DAN SARAN

Dalam penelitian ini, model random forest telah berhasil dalam memprediksi dengue syock syndrome pada pasien demam berdarah. Melalui serangkaian penelitian dan analisis yang telah dilakukan maka dapat ditarik kesimpulan yaitu.

Hyperparameter memiliki sedikit pengaruh terhadap performa model random forest dalam memeprediksi dengue syock syndrome. Perbedaan nilai masing masing hyperparameter dapat mempengaruhi evaluasi model random forest.Extraction future memiliki pengaruh terhadap performa random forest, information gain memiliki performa lebih baik daripada PCA pada setiap jumlah fitur Data imbalance yang telah dilakukan oversampling dapat meningkatkan performa evaluasi random forest dibandingkan data imbalance tanpa oversampling. Metode random oversampling memiliki hasil lebih baik daripada metode SMOTE pada dataset dengue shock syndrome.



6. DAFTAR PUSTAKA

- Alam, M. S. & Vuong, S. T., 2013. Random Forest Classification for Detecting Android Malware. *International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing*, pp. 663-669.
- Candra, A., 2010. Demam Berdarah Dengue: Epidemiologi, Patogenesis, dan. *Aspirator*, 2(2), pp. 110-119.
- Cutler, R., Stevens, J. R. & D, A. C., 2011. Random Forests. *Ensemble Machine Learning: Methods and Applications*, 5(1), pp. 157-176.
- Dihni, V. A., 2022. *Ini Provinsi dengan Angka Kesakitan DBD Tertinggi Nasional*. [Online] Available at: <https://databoks.katadata.co.id/datapublish/2022/03/02/ini-provinsi-dengan-angka-kesakitan-dbd-tertinggi-nasional> [Accessed 4 Maret 2023].
- Ferdiansyah, U., Lubis, A. I. & Erwan, K., 2022. Komparasi Metode K-Nearest Neighbor dan Random Forest Dalam Prediksi Akurasi Klasifikasi Pengobatan Penyakit Kutil. *Jurnal Media Informatika Budidarma*, 6(1), pp. 208-214.
- Greenacre, M. et al., 2022. Principal component analysis. *Nature Reviews Methods Primers*, 2(100).
- Indonesia, K. K. R., 2023. *sehatnegeriku*. [Online] Available at: <https://sehatnegeriku.kemkes.go.id/baca/umum/20230205/3642353/atasi-dengue-kemenkes-kembangkan-dua-teknologi-ini/#:~:text=Berdasarkan%20data%20Kementerian%20Kesehatan%2C%20di,anak%20usi a%200%2D14%20tahun>. [Accessed 26 Maret 2023].
- Khaidir, Zara, N. & Ikhsan, R., 2022. Gambaran Penyakit Demam Berdarah Dengue di Poliklinik Umum Puskesmas Muara Batu Aceh Utara. *Jurnal Kedokteran dan Kesehatan Malikussaleh*, 1(1), pp. 44-50.
- Linawati, S., Nurdiani, S., Handayani, K. & L., 2020. Prediksi Prestasi Akademik Mahasiswa Menggunakan Algoritma Random Forest dan C4.5. *Jurnal Khatulistiwa Informatika*, 8(1), pp. 47-52.
- Nasution, M. Z. F., Sitompul, O. S. & Ramli, M., 2018. PCA based feature reduction to improve the accuracy of decision tree c4.5 classification. *Journal of Physics: Conference Series*, Volume 978.
- Nguyen, C., Wang, Y. & Nguyen, H. N., 2013. Random forest classifier combined with feature selection for breast cancer diagnosis and prognostic. *Biomedical Science and Engineering*, Volume 6, pp. 551-560.
- Nur, F., Ahsan, M. & Harianto, W., 2022. KOMPARASI TINGKAT AKURASI INFORMATION GAIN DAN GAIN RATIO PADA METODE K-NEAREST NEIGHBOR. *Jurnal Mahasiswa Teknik Informatika*, 6(1), pp. 386 - 391.
- Podung, G. C. D., Tatura, S. N. N. & Mantik, M. F. J., 2021. Faktor Risiko Terjadinya Sindroma Syok Dengue pada Demam Berdarah. *Biomedik*, 13(2), pp. 161-166.
- Prasetya, D. I. et al., 2017. Faktor Karakteristik Klinis Host dan Sosiodemografik yang Berpengaruh Terhadap Kejadian Dengue Shock Syndrome. *Jurnal Epidemiologi Kesehatan Komunitas*, 2(2), pp. 99-108.
- Pujiarti, R., 2016. Faktor Faktor yang berhubungan dengan Kejadian Dengue Shock Syndrome pada Anak di Rumah Sakit Umum di Daerah Tugurejo Kota Semarang. *Jurusan Ilmu Kesehatan Masyarakat*.
- Raveendran, S. & Budiarta, I. G., 2016. Dengue Syok Sindrom.



- Salsabila, O., Shodikin, M. A. & Rachmawati, D. A., 2017. Analisis Faktor Risiko Terjadinya Sindrom Syok Dengue Pada Anakdi RSD dr. Soebandi Kabupaten Jember. *Journal of Agromedicine and Medical Sciences*, 3(1), pp. 56-61.
- Sandag, G. A., 2020. Prediksi Rating Aplikasi App Store Menggunakan Algoritma Random Forest. *Cogito Smart Journal*, 6(2), pp. 167-178.
- Speiser, J. L., Miller, M. E., Tooze, J. & Ip, E., 2019. A comparison of random forest variable selection methods for classification prediction modeling. *Expert Systems with Applications*, Volume 134, pp. 93-101.
- Sukohar, 2014. Demam Berdarah Dengue (DBD). *Medula*, 2(2), pp. 1-15.
- Sutoyo, E. & Fadlurrahman, M. A., 2020. Penerapan SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Television Advertisement Performance Rating Menggunakan Artificial Neural Network. *Jurnal Edukasi dan Penelitian Informatika*, 6(3), pp. 379-385.
- Xin, Y. & Ren, X., 2022. Predicting depression among rural and urban disabled elderly in China using a random forest classifier. *BMC Psychiatry*, pp. 1-11.