

KLASIFIKASI MAKANAN BERDASARKAN NILAI GIZI MENGGUNAKAN ALGORITMA RANDOM FOREST DAN TEKNIK SMOTE

Elsi Titasari Br Bangun^{1*)}, Bayu Anugerah Putra²⁾, Aryanto³⁾

^{1,2,3}Fakultas Ilmu Komputer, Universitas Muhammadiyah Riau
email: elsititasari@umri.ac.id^{*}, bayuanugerahputra@umri.ac.id², aryanto@umri.ac.id³
**Corresponding Author*

Abstract

Classifying food based on nutritional content is essential for developing personalized dietary recommendation systems and promoting healthier eating habits. This study aims to construct a food classification model using the Random Forest algorithm combined with the Synthetic Minority Oversampling Technique (SMOTE) to address class imbalance in the dataset. The dataset includes various nutritional attributes such as calories, protein, fat, carbohydrates, fiber, sugar, sodium, and cholesterol, along with additional information such as food category and mealtime. After preprocessing, the data were split into training and testing sets, with SMOTE applied to the training data to improve class representation. The model was trained using Random Forest and evaluated using accuracy, precision, recall, and F1-score. The results show that the model achieved an accuracy of 83.35% and an average F1-score above 0.80, with the best performance observed in majority classes. The confusion matrix analysis indicates that most predictions were accurate, although misclassifications occurred among classes with overlapping nutritional values. Protein, calories, and carbohydrates were identified as the most influential features in the classification process. These results show that combining Random Forest and SMOTE works well for creating food classification systems using nutritional data and could be useful in apps for diet recommendations and managing nutrition.

Keywords: *food classification, nutritional value, Random Forest, SMOTE, machine learning, recommendation system*

Abstrak

Klasifikasi makanan berdasarkan kandungan gizi merupakan langkah penting dalam pengembangan sistem rekomendasi nutrisi dan pengaturan pola makan yang personal. Penelitian ini bertujuan untuk membangun model klasifikasi makanan berdasarkan atribut nilai gizi menggunakan algoritma *Random Forest* yang dikombinasikan dengan teknik *Synthetic Minority Oversampling Technique* (SMOTE) guna mengatasi ketidakseimbangan kelas dalam data. Dataset yang digunakan terdiri dari berbagai informasi nutrisi seperti kalori, protein, lemak, karbohidrat, serat, gula, natrium, dan kolesterol, serta informasi tambahan seperti kategori makanan dan waktu konsumsi. Setelah melalui tahap pra-pemrosesan, data dibagi menjadi data latih dan uji, dan dilakukan *oversampling* pada data latih menggunakan SMOTE. Model dilatih menggunakan *Random Forest* dan dievaluasi berdasarkan akurasi, *precision*, *recall*, dan *f1-score*. Hasil evaluasi menunjukkan bahwa model mampu mencapai akurasi sebesar 83,35% dan *f1-score* rata-rata di atas 0,80, dengan performa terbaik pada kelas mayoritas. Analisis *confusion matrix* menunjukkan bahwa sebagian besar prediksi tepat sasaran, meskipun terdapat kesalahan klasifikasi pada kelas yang memiliki kemiripan kandungan gizi. Fitur protein, kalori, dan karbohidrat teridentifikasi sebagai atribut paling berpengaruh dalam proses klasifikasi. Temuan ini menunjukkan bahwa kombinasi algoritma *Random Forest* dan teknik SMOTE

efektif dalam membangun sistem klasifikasi makanan berbasis gizi, serta memiliki potensi diterapkan dalam sistem rekomendasi diet dan manajemen gizi digital.

Kata Kunci: klasifikasi makanan, nilai gizi, *Random Forest*, SMOTE, *machine learning*, sistem rekomendasi

PENDAHULUAN

Dalam *machine learning*, klasifikasi adalah teknik yang digunakan untuk mengelompokkan objek ke dalam kelas tertentu berdasarkan atribut atau karakteristik yang dimilikinya. Penerapan teknik ini dalam ilmu gizi dan teknologi informasi memungkinkan proses pengelompokan makanan berdasarkan nilai gizi seperti kandungan kalori, lemak, protein, maupun karbohidrat (Jamnekar et al. 2022). Dengan pendekatan ini, identifikasi makanan sesuai dengan kebutuhan nutrisi seperti makanan yang kaya protein, rendah karbohidrat, atau bernutrisi seimbang menjadi lebih praktis (Gilal et al., 2024). Penerapan sistem klasifikasi ini sangat penting dalam mendukung pengaturan pola makan yang lebih sehat dan dapat disesuaikan untuk individu sehat maupun mereka yang memiliki kondisi kesehatan tertentu.

Penerapan algoritma klasifikasi dalam pengolahan data nutrisi sangat penting karena realitas konsumsi pangan masyarakat Indonesia yang masih jauh dari ideal. Berdasarkan Survei Kesehatan Indonesia (SKI) 2023 yang mengintegrasikan Riskesdas dan Survei Status Gizi Balita Indonesia sebanyak 96,7% masyarakat Indonesia dikategorikan kurang mengonsumsi buah dan sayur sesuai rekomendasi gizi seimbang. Pada Riskesdas 2018, hanya 4,5% penduduk yang memenuhi panduan gizi seimbang yaitu lima porsi buah dan sayur per hari. Lebih lanjut, data Riskesdas 2018 mengungkapkan bahwa 95,5% penduduk usia ≥ 5 tahun mengonsumsi buah dan sayur di bawah anjuran WHO, dan untuk kelompok remaja 15–19 tahun angkanya mencapai 96,4%. Kondisi ini menunjukkan gap besar antara kebutuhan dan praktik konsumsi sehari-hari. Karena rendahnya literasi gizi serta minimnya data nutrisi yang tersedia secara personal, algoritma klasifikasi seperti *K-Nearest Neighbor*, *Decision Tree*, *AdaBoost*, dan *Random Forest* diperlukan untuk membantu mengelompokkan makanan berdasarkan kebutuhan individual (aktivitas harian, kondisi medis, dan preferensi), sehingga rekomendasi nutrisi menjadi lebih akurat dan

tepat sasaran (Orue-saiz et al. 2021), (Ajami & Teimourpour, 2023).

Algoritma *Random Forest* (RF) merupakan salah satu metode klasifikasi yang banyak digunakan dalam pengolahan data gizi. Pemilihan algoritma *Random Forest* dalam penelitian ini didasarkan pada jumlah data yang digunakan, yaitu sebanyak 10.000 entri dengan 14 fitur nutrisi yang memiliki karakteristik multivariabel dan heterogen. Skala data tersebut tergolong besar dalam konteks pengolahan data gizi, sehingga dibutuhkan algoritma yang mampu menangani kompleksitas dan volume data secara efisien. Keunggulan *Random Forest* terletak pada kemampuannya menangani data berskala besar, toleransi terhadap overfitting, serta kemampuannya untuk mengevaluasi pentingnya fitur dalam prediksi (Fatmawati et al. 2024). Beberapa penelitian telah membuktikan efektivitas algoritma ini, terutama ketika dikombinasikan dengan teknik penyeimbangan data seperti SMOTE (*Synthetic Minority Oversampling Technique*) untuk mengatasi ketidakseimbangan antar kelas.

Penelitian oleh (Miranda et al., 2024) mengembangkan model prediksi status gizi pada anak balita dengan memanfaatkan kombinasi algoritma *Random Forest* dan teknik *Synthetic Minority Oversampling Technique* (SMOTE) berdasarkan data rekam medis elektronik. Model yang dihasilkan menunjukkan kinerja yang sangat baik, ditunjukkan melalui nilai *f1-score* sebesar 84,7%, akurasi mencapai 88,9%, serta *area under the curve* (AUC) sebesar 93%, yang mencerminkan tingkat ketepatan klasifikasi yang tinggi dalam konteks data yang tidak seimbang.

Penelitian lain dilakukan oleh (Putri et al. 2021) mengenai klasifikasi status vaksinasi Hepatitis-B menunjukkan bahwa penerapan algoritma *Random Forest* yang dikombinasikan dengan teknik SMOTE secara signifikan meningkatkan akurasi model. Temuan serupa juga dilaporkan oleh (Hairani et al. 2023), yang membandingkan efektivitas SMOTE dan SMOTE-Tomek Link dalam klasifikasi data diagnosis diabetes. Hasil penelitian tersebut

menunjukkan bahwa integrasi SMOTE-Tomek Link dengan *Random Forest* menghasilkan performa klasifikasi yang lebih baik dibandingkan dengan penggunaan SMOTE standar.

Penelitian yang dilakukan oleh (Prasetya & Abdurakhman, 2023) dalam konteks klasifikasi kanker serviks menunjukkan bahwa penerapan algoritma *Random Forest* yang dikombinasikan dengan teknik SMOTE mampu mencapai tingkat akurasi yang sangat tinggi, yakni sebesar 96,28%. Temuan ini menegaskan efektivitas pendekatan tersebut dalam mengatasi permasalahan ketidakseimbangan kelas pada data medis.

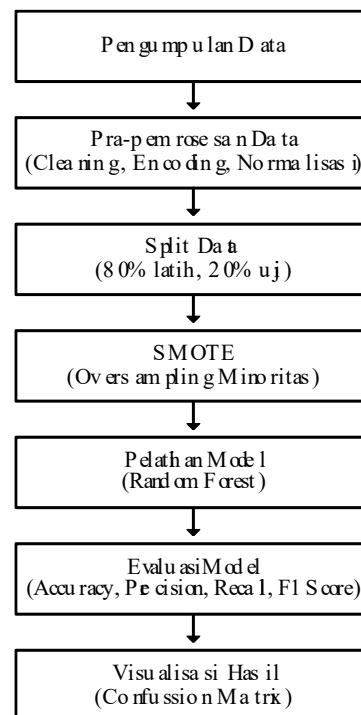
Lebih lanjut, kombinasi *Random Forest* dan SMOTE juga telah terbukti memberikan hasil yang positif di berbagai domain lainnya, seperti klasifikasi status gizi balita (Fatmawati et al., 2024), deteksi gangguan panik (Nurmalasari et al., 2024), klasifikasi kanker paru-paru (Latief et al., 2024), dan klasifikasi kanker payudara (Herawati et al. 2024). Dalam seluruh studi tersebut, integrasi SMOTE secara konsisten menunjukkan kontribusi signifikan dalam meningkatkan kinerja model klasifikasi, khususnya dalam memperbaiki representasi kelas minoritas yang sering terabaikan dalam distribusi data yang tidak seimbang.

Kombinasi algoritma *Random Forest* dan teknik SMOTE tidak hanya menunjukkan relevansi dari segi teknis, tetapi juga menawarkan manfaat praktis dalam pengembangan sistem informasi gizi. Kemampuan metode ini dalam mengakomodasi karakteristik data nutrisi yang tidak seimbang menjadikannya solusi yang unggul untuk merancang sistem klasifikasi makanan. Sistem tersebut dapat digunakan secara efektif untuk mendukung rekomendasi diet personal, manajemen kondisi medis berbasis nutrisi, serta penyusunan pola makan yang lebih presisi dan adaptif terhadap kebutuhan individu (Rachmatullah, 2022).

METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen komputasional berbasis *machine learning*. Penelitian ini dirancang untuk mengembangkan model klasifikasi makanan berdasarkan atribut nilai gizi, seperti kalori, protein, karbohidrat,

lemak, serat, gula, natrium, dan kolesterol. Data yang digunakan diperoleh dari kumpulan entri konsumsi makanan harian pengguna yang memuat informasi nutrisi, kategori makanan, waktu makan, serta volume asupan air. Proses pengolahan data dilakukan menggunakan lingkungan kerja *Google Colab* dan bahasa pemrograman *Python*, dengan bantuan pustaka seperti *pandas*, *numpy*, *scikit-learn*, *matplotlib*, *seaborn*, dan *imblearn*. Proses pengolahan data dilakukan secara bertahap mulai dari pengumpulan hingga visualisasi hasil, sebagaimana digambarkan dalam diagram alir pada tahapan penelitian dibawah ini.



Gambar 1. Alur Penelitian

Data dalam penelitian ini diperoleh dari platform *Kaggle*, yang menyediakan dataset terkait nutrisi makanan harian. Data awal masih dalam bentuk mentah sehingga perlu melalui tahap pra-pemrosesan, yang mencakup pembersihan data dari duplikasi dan nilai kosong, proses *encoding* untuk mengubah data kategorikal menjadi numerik, serta normalisasi agar skala fitur menjadi seimbang. Setelah data diproses, dilakukan *split data* menjadi data latih (untuk melatih model) sebesar 80% dan data uji (untuk menguji kinerja model) sebesar 20%. Untuk mengatasi ketidakseimbangan distribusi kelas, digunakan teknik *Synthetic Minority Oversampling Technique* (SMOTE), yang menghasilkan sampel sintesis baru pada kelas minoritas dengan menginterpolasi tetangga

terdekat dalam ruang fitur. Proses ini diterapkan hanya pada data latih untuk mencegah kebocoran data (*data leakage*).

Model klasifikasi dilatih menggunakan algoritma *Random Forest*. Secara matematis, prediksi dalam algoritma *Random Forest* dilakukan dengan menggabungkan hasil dari sejumlah pohon keputusan. Untuk kasus klasifikasi, prediksi akhir (\hat{y}) ditentukan berdasarkan mayoritas suara dari seluruh pohon yang dibentuk dalam ensemble, yang dirumuskan sebagai:

$$\hat{y} = \text{mode} \{h_1(x), h_2(x), \dots, h_T(x)\} \quad (1)$$

Sedangkan untuk regresi, prediksi dilakukan dengan menghitung rata-rata nilai prediksi dari semua pohon, yang diformulasikan sebagai:

$$\hat{y} = \frac{1}{T} \sum_{t=1}^T h_t(x) \quad (2)$$

Di mana $h_t(x)$ adalah prediksi dari pohon ke - t , dan T adalah jumlah total pohon dalam hutan acak. Beberapa pengembangan algoritma lebih lanjut juga mengintegrasikan bobot probabilistik terhadap masing-masing pohon berdasarkan performa atau akurasi individualnya (Zhang & Wang, 2021), (Yu et al. 2020), (Salman et al. 2024).

Implementasi dilakukan melalui pustaka *sklearn.ensemble*, dengan parameter awal $n_estimators = 100$ dan $random_state = 42$. Model dilatih menggunakan data latih yang telah diseimbangkan hasil dari SMOTE. Evaluasi model dilakukan menggunakan metrik akurasi, *precision*, *recall*, *f1-score*, dan visualisasi hasil dengan *confusion matrix*, untuk menilai kinerja klasifikasi secara menyeluruh.

HASIL DAN PEMBAHASAN

1. Pengumpulan dan Karakteristik Data

Dataset yang digunakan diperoleh dari platform Kaggle, terdiri dari 10.000 entri dengan 14 fitur nutrisi harian makanan. Data tersebut mencakup informasi seperti jumlah kalori, protein, lemak, karbohidrat, dan label kelas makanan. Dataset masih dalam bentuk mentah sehingga memerlukan tahap pembersihan dan transformasi data sebelum dianalisis lebih lanjut.

2. Pra-Pemrosesan Data

Tahap pra-pemrosesan dilakukan untuk meningkatkan kualitas data dan memastikan data siap digunakan dalam pelatihan model. Pada tahap ini, data dibersihkan dari duplikasi dan nilai kosong, kemudian dilakukan *encoding* terhadap fitur kategorikal seperti *Meal_Type* dan *Category* menggunakan teknik *one-hot encoding*. Fitur numerik dinormalisasi menggunakan metode standardisasi.

3. Split Data

Setelah pra-pemrosesan, data dibagi menjadi data latih dan data uji dengan rasio 80:20 menggunakan fungsi *train_test_split*, disertai parameter *stratify* untuk menjaga proporsi kelas. Distribusi kelas pada data latih menunjukkan ketidakseimbangan, di mana beberapa kelas jauh lebih dominan daripada yang lain. Untuk mengatasi hal ini, digunakan teknik *Synthetic Minority Oversampling Technique* (SMOTE). Hasil dari SMOTE menunjukkan bahwa jumlah sampel dari setiap kelas menjadi seimbang, yang membantu model agar tidak bias terhadap kelas mayoritas.

4. Model Random Forest

Model klasifikasi dibangun menggunakan algoritma *Random Forest*. Model ini dilatih menggunakan data latih yang telah diseimbangkan oleh SMOTE. Algoritma *Random Forest* dan teknik penyeimbangan data SMOTE menunjukkan kinerja yang cukup baik.

5. Evaluasi Kinerja Model

Evaluasi dilakukan terhadap data uji untuk menilai performa model klasifikasi. Hasil pengujian menunjukkan bahwa model mencapai tingkat akurasi sebesar 83,35%, yang berarti 83 dari setiap 100 prediksi model sesuai dengan label yang sebenarnya. Akurasi ini mengindikasikan bahwa model mampu belajar dengan efektif dari data latih yang telah diseimbangkan menggunakan SMOTE.

Selain akurasi, dilakukan juga evaluasi terhadap metrik lain seperti *precision*, *recall*, dan *f1-score*. Tabel 1 berikut menunjukkan ringkasan metrik evaluasi berdasarkan *classification report*:

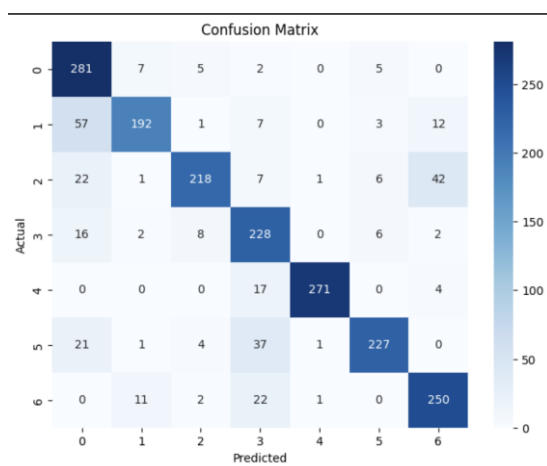
Tabel 1. Hasil Evaluasi Performa Model

Kelas	Precision	Recall	F1-Score	Support
0	0.71	0.94	0.81	300
1	0.90	0.71	0.79	272
2	0.92	0.73	0.81	297
3	0.71	0.87	0.78	2.62
4	0.99	0.93	0.96	292
5	0.92	0.78	0.84	291
6	0.81	0.87	0.84	286

Model memiliki performa terbaik pada kelas 4 dan kelas 2 dengan nilai f1-score tinggi, sedangkan kelas 1 menunjukkan nilai recall yang lebih rendah. Hal ini menunjukkan bahwa kelas 1 masih cukup sering salah diklasifikasikan sebagai kelas lain.

6. Visualisasi Confusion Matrix

Untuk memahami lebih jauh pola kesalahan klasifikasi, digunakan confusion matrix yang divisualisasikan dalam Gambar 2 di bawah ini.



Gambar 2. Confusion Matrix Hasil Klasifikasi

Gambar 2 menunjukkan confusion matrix dari hasil prediksi model klasifikasi *Random Forest* terhadap tujuh kelas makanan (kelas 0 hingga kelas 6) berdasarkan kandungan nutrisinya, seperti kalori, protein, lemak, dan karbohidrat. Sumbu vertikal menunjukkan label aktual (*Actual*), sementara sumbu horizontal menunjukkan label hasil prediksi model (*Predicted*). Angka-angka pada diagonal utama menunjukkan jumlah klasifikasi yang benar untuk masing-masing kelas, sedangkan angka-angka di luar diagonal mencerminkan kesalahan klasifikasi. Misalnya, untuk kelas 0, dari 295 sampel aktual, sebanyak 281 diklasifikasikan dengan benar, sementara sisanya diprediksi

salah ke kelas lain, terutama ke kelas 1 (7 sampel) dan kelas 2 (5 sampel). Kelas 1 memiliki 192 prediksi benar dari total 265 sampel, namun mengalami kekeliruan cukup besar ke kelas 0 sebanyak 57 sampel. Pada kelas 2, terdapat 218 prediksi benar dari 289 sampel, dengan kesalahan paling signifikan ke kelas 6 sebanyak 42 sampel. Kelas 3 memiliki performa yang sangat baik, dengan 228 dari 238 sampel diklasifikasikan secara akurat. Demikian pula kelas 4, yang menunjukkan akurasi tinggi dengan 271 dari 288 sampel diprediksi dengan tepat, meskipun terdapat 17 sampel yang salah diklasifikasikan ke kelas 3. Kelas 5 menunjukkan hasil cukup baik dengan 227 prediksi benar dari 286 sampel, dan kesalahan terbesar ke kelas 3 (37 sampel). Terakhir, kelas 6 memiliki 250 prediksi benar dari 281 sampel, namun mengalami kesalahan klasifikasi ke kelas 2 sebanyak 22 sampel. Secara keseluruhan, model menunjukkan performa klasifikasi yang baik, tercermin dari dominasi angka besar pada diagonal utama, namun masih terdapat ruang untuk perbaikan, terutama pada kelas-kelas yang menunjukkan pola kesalahan silang seperti antara kelas 0 dan 1, serta kelas 2 dan 6, yang kemungkinan disebabkan oleh kemiripan profil nutrisi antar kategori.

KESIMPULAN

Penelitian ini menunjukkan bahwa kombinasi algoritma *Random Forest* dan metode SMOTE mampu menghasilkan model klasifikasi makanan yang efektif berdasarkan kandungan nilai gizi. Model yang dikembangkan berhasil mencapai tingkat akurasi sebesar 83,35%, dengan performa klasifikasi yang cukup seimbang di berbagai kelas. Teknik SMOTE berhasil meningkatkan representasi pada kelas minoritas, sehingga menghasilkan perbaikan pada metrik evaluasi seperti *f1-score* dan *recall* dengan rata-rata di atas 0,80. Meskipun beberapa kelas masih mengalami kesalahan klasifikasi akibat kemiripan nilai fitur antar kategori, secara umum model dapat mengidentifikasi pola klasifikasi makanan secara andal. Analisis terhadap fitur juga memperlihatkan bahwa atribut seperti protein, kalori, dan karbohidrat merupakan indikator paling berpengaruh dalam penentuan kelas. Dengan demikian, model ini memiliki potensi untuk diterapkan dalam sistem rekomendasi makanan atau alat bantu personalisasi pola makan berbasis data.

DAFTAR PUSTAKA

- Ajami, A., & Teimourpour, B. (2023). A Food Recommender System in Academic Environments Based on Machine Learning Models. *Iranian Journal of Public Health*.
- Fatmawati, M., Herlambang, Bambang Agus, & Nada, N. Q. (2024). Random Forest Algorithm for Toddler Nutritional Status Classification Website. *Journal of Applied Informatics and Computing (JAIC)*, 8(2), 428–433. Retrieved from <http://jurnal.polibatam.ac.id/index.php/JAIC>
- Gilal, N. U., Al-Thelaya, K., Al-Saeed, J. K., Abdallah, M., Schneider, J., She, J., Awan, J. H., et al. (2024). Evaluating machine learning technologies for food computing from a data set perspective. *Multimedia Tools and Applications*, 83(11), 32041–32068. Springer.
- Hairani, H., Anggrawan, A., & Priyanto, D. (2023). Improvement Performance of the Random Forest Method on Unbalanced Diabetes Data Classification Using Smote-Tomek Link. *INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION*, 7, 258–264. Retrieved from www.joiv.org/index.php/joiv
- Herawati, B. C., Hairani, H., & Guterres, J. X. (2024). SMOTE Variants and Random Forest Method: A Comprehensive Approach to Breast Cancer Classification. *International Journal of Engineering Continuity*, 3(1), 12–23. Sultan Publisher.
- Jamnekar, R., Keole, R. R., Mohod, S. W., Mahore, T. R., & Pande, S. (2022). A Survey on Food Classification Using Data Mining Techniques. *A Survey on Food Classification Using Data Mining Techniques*. Proceedings of the International Conference on Innovative Computing & Communication (ICICC). Retrieved from <https://ssrn.com/abstract=4143847>
- Latief, M. A., Nabila, L. R., Miftakhurrahman, W., Ma'rufatullah, S., & Tantyoko, H. (2024). Handling Imbalance Data using Hybrid Sampling SMOTE-ENN in Lung Cancer Classification. *International Journal of Engineering and Computer Science Applications (JECSA)*, 3(1), 11–18. Universitas Bumigora.
- Miranda, E., Aryuni, M., Zakiyyah, A. Y., Kurniawati, Y. E., Sano, A. V. D., & Kumbangsila, M. (2024). An early prediction model for toddler nutrition based on machine learning from imbalanced data. *Procedia Computer Science* (Vol. 245, pp. 263–271). Elsevier B.V.
- Nurmalasari, D., Yuliantoro, H. R., Hidayatul Qudsi, D., Informasi, T., Caltex Riau, P., & Perpajakan, A. (2024). Improving Panic Disorder Classification Using SMOTE and Random Forest. *Journal of Applied Informatics and Computing (JAIC)*, 8(2), 272–279. Retrieved from <http://jurnal.polibatam.ac.id/index.php/JAIC>
- Orue-saiz, I., Kazarez, M., & Mendez-zorrilla, A. (2021). Systematic review of nutritional recommendation systems. *Applied Sciences (Switzerland)*, 11(24), 1–13. MDPI.
- Prasetya, J., & Abdurakhman, A. (2023). COMPARISON OF SMOTE RANDOM FOREST AND SMOTE K-NEAREST NEIGHBORS CLASSIFICATION ANALYSIS ON IMBALANCED DATA. *MEDIA STATISTIKA*, 15(2), 198–208. Institute of Research and Community Services Diponegoro University (LPPM UNDIP).
- Putri, V. M., Masjkur, M., & Suhaeni, C. (2021). Performance of SMOTE in a random forest and naive Bayes classifier for imbalanced Hepatitis-B vaccination status. *Journal of Physics: Conference Series* (Vol. 1863, pp. 1–8). IOP Publishing Ltd.
- Rachmatullah, M. I. C. (2022). The Application of Repeated SMOTE for Multi Class Classification on Imbalanced Data. *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, 22(1), 13–24. Universitas Bumigora.
- Salman, H. A., Kalakech, A., & Steiti, A. (2024). Random Forest Algorithm Overview. *Babylonian Journal of Machine Learning*,

2024, 69–79. Mesopotamian Academic Press.

Yu, Y., Wang, L., Huang, H., & Yang, W. (2020). An Improved Random Forest Algorithm. *Journal of Physics: Conference Series* (Vol. 1646, pp. 1–6). IOP Publishing Ltd.

Zhang, X., & Wang, M. (2021). Weighted Random Forest Algorithm Based on Bayesian Algorithm. *Journal of Physics: Conference Series* (Vol. 1924, pp. 1–6). IOP Publishing Ltd.