
ANALISIS PERBANDINGAN RANDOM FOREST DAN KNN PADA KLASIFIKASI PENERIMA MANFAAT PROGRAM MAKAN BERGIZI GRATIS

Muhammad Imam Zarkasyi¹, Peter Jaya Negara Simanjuntak², Junerdi Nababan³
Universitas Satya Terra Bhinneka, Medan

e-mail: ¹m.imamzarkasyi96@gmail.com, ²pejayra@satyaterrabhinneka.ac.id,
³junerdin@satyaterrabhinneka

Abstract: *The Free Nutritional Meal Program (PMBG) is a government initiative to improve students' nutrition and school attendance. This study evaluates and compares the performance of Random Forest (RF) and K-Nearest Neighbors (KNN) algorithms in classifying students most eligible for PMBG based on socio-economic criteria. The dataset comprises 205 public elementary school students in Medan City, collected via questionnaires. Features include parental income, number of dependents, housing status, asset ownership, and participation in other social aid programs. The data was clustered into three priority groups using K-Means. To address class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) was applied. Modeling used three data split scenarios (70:30, 80:20, 90:10) and was evaluated with accuracy, precision, recall, F1-score, and cross-validation. Results show that RF consistently outperformed KNN across all scenarios. After SMOTE, both models improved, with Balanced-RF achieving the highest accuracy and F1-score (94%) in the 70:30 split. The combination of RF and SMOTE proves effective for building an objective and accurate priority classification system for PMBG beneficiaries.*

Keyword: *Free Nutritious Meal Program; Random Forest; K-Nearest Neighbors; SMOTE.*

Abstrak: Program Makan Bergizi Gratis (PMBG) merupakan inisiatif pemerintah yang bertujuan meningkatkan asupan gizi dan mendorong kehadiran siswa di sekolah. Penelitian ini bertujuan untuk mengevaluasi dan membandingkan kinerja dua algoritma machine learning, yaitu Random Forest dan K-Nearest Neighbors, dalam mengklasifikasikan siswa yang paling berhak menerima manfaat PMBG berdasarkan kriteria sosial-ekonomi. Dataset yang digunakan terdiri dari 205 siswa Sekolah Dasar Negeri di Kota Medan yang dikumpulkan melalui kuesioner. Fitur yang digunakan meliputi pendapatan orang tua, jumlah tanggungan, status tempat tinggal, kepemilikan aset, dan partisipasi dalam program bantuan sosial lainnya. Dataset yang telah dikumpulkan kemudian dikelompokkan menggunakan algoritma K-Means menjadi tiga kluster prioritas. Untuk mengatasi ketidakseimbangan distribusi data, digunakan metode Synthetic Minority Over-sampling Technique (SMOTE). Pemodelan dilakukan dalam tiga skenario pembagian data (70:30, 80:20, 90:10) dan dievaluasi menggunakan metrik akurasi, presisi, recall, f1-score, dan cross-validation. Hasil penelitian menunjukkan bahwa algoritma Random Forest secara konsisten memberikan kinerja yang lebih unggul dibandingkan KNN pada semua skenario. Setelah penerapan SMOTE, kedua algoritma mengalami peningkatan performa, dengan Random Forest-Balanced mencatat akurasi dan f1-score tertinggi sebesar 94% pada skenario 70:30. Temuan ini menunjukkan bahwa kombinasi Random Forest dan SMOTE merupakan pendekatan yang efektif dan efisien untuk membangun sistem klasifikasi prioritas penerima manfaat PMBG yang objektif dan akurat.

Kata kunci: Program Makan Bergizi Gratis; Random Forest; K-Nearest Neighbors; SMOTE;.

PENDAHULUAN

Pemerintah Indonesia telah meluncurkan Program Makan Bergizi Gratis (PMBG) sebagai salah satu strategi untuk menanggulangi masalah stunting, meningkatkan asupan gizi anak sekolah, dan sekaligus mendorong peningkatan angka partisipasi sekolah (Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia [Kemendikbudristek], 2023). Program ini ditargetkan untuk menjangkau jutaan siswa di seluruh Indonesia, dengan prioritas pada daerah yang memiliki prevalensi stunting tinggi dan keluarga dengan kondisi sosial-ekonomi rendah (World Bank, 2022). Efektivitas program ini sangat bergantung pada ketepatan dalam mengidentifikasi dan menargetkan penerima manfaat agar bantuan dapat tepat sasaran (Hardi & Prasetyo, 2024).

Kota Medan, sebagai salah satu kota metropolitan terbesar di Indonesia, menghadapi tantangan dalam pendistribusian bantuan sosial secara merata dan tepat sasaran. Data dari Dinas Pendidikan Kota Medan menunjukkan terdapat ratusan SD Negeri yang menjadi sasaran program ini (Dinas Pendidikan Kota Medan, 2024). Seleksi penerima manfaat yang selama ini dilakukan seringkali masih bersifat manual berdasarkan rekomendasi guru atau wali kelas, yang berpotensi menyebabkan subjektivitas dan ketidakefisienan (Febrianto & Sari, 2023). Oleh karena itu, diperlukan sebuah sistem yang dapat membantu proses klasifikasi dan prioritas penerima manfaat secara lebih objektif, akurat, dan terukur.

Kecerdasan buatan (Artificial Intelligence) dan machine learning menawarkan solusi potensial untuk permasalahan ini. Dua algoritma yang sering digunakan untuk tugas klasifikasi adalah Random Forest (RF) dan K-Nearest Neighbors (KNN). Random Forest dikenal memiliki kinerja yang kuat, stabil, dan tahan terhadap overfitting, sehingga banyak diaplikasikan dalam berbagai bidang

klasifikasi fields (Breiman, 2001), (Rahmasari, Rifany, Piharyanto, & Kurniawan, 2024). Sementara itu, KNN merupakan algoritma sederhana yang efektif untuk dataset dengan dimensi tidak terlalu tinggi dan memiliki intuisi yang mudah dipahami (Cover & Hart, 1967).

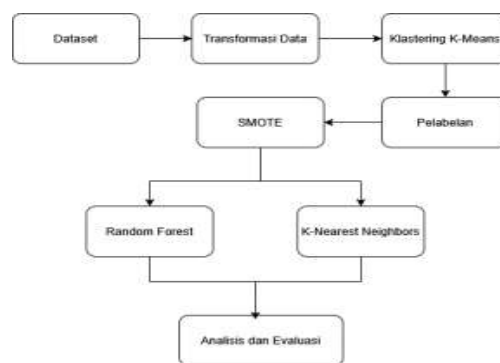
Beberapa penelitian sebelumnya telah menerapkan algoritma machine learning untuk klasifikasi bantuan sosial. Penelitian oleh Wijaya dan D. Puspita berhasil menerapkan Decision Tree untuk menentukan penerima bantuan tunai dengan akurasi 85% (Wijaya & Puspita, 2024). Studi lain oleh B. Santoso dan A. Nurjanah membandingkan beberapa algoritma untuk prediksi penerima bantuan pangan dan menemukan bahwa Support Vector Machine (SVM) memberikan performa terbaik (Santoso & Nurjanah, 2024). Penelitian oleh M. A. Firdaus dan S. K. Wibowo menggunakan Naïve Bayes untuk klasifikasi penerima bantuan sosial dengan akurasi 82% (Firdaus & Wibowo, 2023). Sementara itu D. R. Putra dan E. S. Ningrum, mengaplikasikan Random Forest untuk identifikasi keluarga miskin dengan hasil akurasi 89% (Putra & Ningrum, 2022). Meskipun beberapa penelitian sebelumnya telah menggunakan machine learning untuk klasifikasi penerima bantuan sosial, sebagian besar masih berfokus pada perbandingan performa algoritma dalam konteks aplikatif tanpa mengaitkannya dengan pengembangan teori dalam social targeting atau policy analytics. Penelitian ini tidak hanya menguji performa algoritma, tetapi juga mengintegrasikan kerangka data-driven decision-making dalam konteks penargetan bantuan sosial, yang masih jarang dieksplorasi secara teoretis dalam literatur terkait. Dengan demikian, penelitian ini tidak hanya memberikan solusi teknis, tetapi juga berkontribusi pada pengembangan teori mengenai bagaimana model machine learning dapat meningkatkan efisiensi dan objektivitas dalam penyaluran program sosial, khususnya dalam konteks negara berkembang seperti Indonesia.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk membangun model klasifikasi untuk memprioritaskan penerima manfaat PMBG di SD Negeri Kota Medan menggunakan algoritma Random Forest dan K-Nearest Neighbors serta mengevaluasi dan membandingkan kinerja kedua algoritma tersebut menggunakan metrik akurasi, presisi, recall, f1-score, dan cross-validation.

Dengan adanya penelitian ini, diharapkan dapat dihasilkan sebuah model klasifikasi yang dapat membantu pihak sekolah dan dinas terkait dalam mengambil keputusan yang lebih objektif, efektif, dan efisien dalam penyaluran Program Makan Bergizi Gratis

METODE

Penelitian ini mengusulkan sebuah pipeline analitik untuk mengatasi masalah klasifikasi pada dataset tidak seimbang. Teknik yang dijalankan pada penelitian ini dimulai dengan preparasi dan transformasi data, dilanjutkan dengan klastering menggunakan K-Means untuk memahami pengelompokan data. Teknik oversampling SMOTE kemudian diterapkan untuk menyeimbangkan distribusi kelas. Hasil cluster yang telah di-oversample dijadikan dasar dalam proses pelabelan. Model klasifikasi akhir dibangun dengan algoritma Random Forest dan K-Nearest Neighbors, yang selanjutnya dibandingkan performanya melalui analisis evaluasi menyeluruh. Alur tahapan penelitian seperti pada Gambar 1 berikut.



Gambar 1 Tahapan Penelitian HASIL DAN PEMBAHASAN

Dalam memperoleh informasi dari hasil pada penelitian ini, dilakukan tahapan pengujian yang sebelumnya telah dipaparkan pada bagian sebelumnya. Tahapan pengujian yang dilakukan yaitu transformasi data pada dataset terlebih dahulu sebelum pengujian clustering, klastering menggunakan K-Means, dilakukan pelabelan pada data, SMOTE, pengujian clustering pada dataset dengan RF atau KNN, dan menganalisis hasil clustering dari pengujian dataset.

Transformasi Dataset

Adapun pada penelitian ini data yang digunakan sebagai instrument pengujian dari metode yang diusulkan yaitu pengumpulan 205 data respon lengkap dari orang tua/wali murid SD Negeri di Medan. Setiap responden memberikan jawaban lengkap untuk seluruh item dalam instrumen, tanpa ada data yang hilang atau tidak terisi dalam keseluruhan dataset. Data yang dikumpulkan terdiri dari 11 kolom, yang meliputi informasi identifikasi responden serta serangkaian pernyataan terkait penerimaan program makan bergizi gratis dari pemerintah Indonesia. Pada tahapan ini, dataset telah tersusun lengkap dan telah ditransformasi sehingga dapat diolah. Tabel 2 sampel menampilkan data sebagai representasi dari keseluruhan pengumpulan data.

Tabel 1 Dataset Setelah Ditransformasi

No	Email	Nama	Q1	Q2-Q6	Q7	Total Score

1	User1	U1	1.2	...	0.5	38.3
2	User2	U2	3.5	...	8.2	19.3
3	User3	U3	0.9	...	0.3	43.8
...
204	User204	U204	4.3	...	2	7.2
205	User205	U205	1.5	...	0	39.7

Preprocessing Data

Pada dataset yang telah ditransformasi, terlihat adanya variasi skor total tingkat stres yang cukup signifikan antar responden. Variasi ini menunjukkan adanya perbedaan tingkat penerimaan program makan bergizi gratis yang diteliti. Skor total ini kemudian digunakan sebagai dasar utama untuk mengelompokkan secara lebih sistematis. Proses klusterisasi dengan metode K-Means membagi responden ke dalam tiga kluster berdasarkan tingkat penerimaan program makan bergizi gratis.

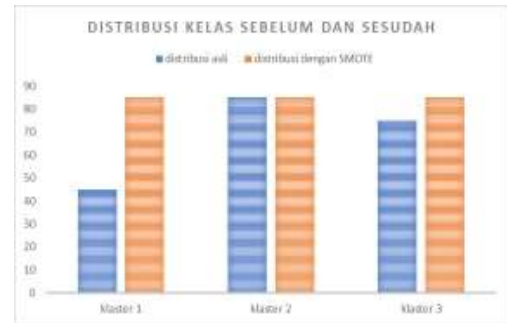
Tabel 2 Hasil Klusterisasi K-Means

Klaster	Jumlah Data	Q1	Q2	Q3 - Q7	Rentang Score	Label Prioritas
0	45	0.8 - 1.5	4-5	...	10-18	Prioritas Tinggi
1	85	1.6 - 2.5	3-4	...	19-25	Prioritas Sedang
2	75	2.6 - 4.5	2-3	...	26-35	Tidak Prioritas

Dari tabel tersebut terlihat bahwa kluster 0 (Prioritas Tinggi) memiliki pendapatan terendah (≤ 1.5 juta/bulan) dan tanggungan terbanyak (4-5 orang). Sebagian besar tinggal menumpang dan tidak memiliki kendaraan. Kluster 1 (Prioritas Sedang) memiliki pendapatan menengah, tanggungan 3-4 orang, sebagian besar tinggal kontrak dan memiliki sepeda motor. Sementara kluster 2 (Tidak Prioritas) memiliki pendapatan tertinggi (≥ 2.6 juta/bulan), tanggungan sedikit (2-3 orang), mayoritas tinggal di rumah sendiri dan memiliki kendaraan.

Ketidakeimbangan pada jumlah data pada tiap-tiap kluster berpotensi menimbulkan bias dalam proses klasifikasi, karena model cenderung mengoptimalkan performa pada kategori dengan jumlah data terbanyak, sehingga mengurangi akurasi prediksi untuk kategori yang lebih sedikit datanya (Budiarto, 2023).

Untuk mengatasi permasalahan ke distribusi data antar kategori tingkat stres, diterapkan teknik Synthetic Minority Over-sampling Technique (SMOTE). Distribusi data awal menunjukkan ketidakseimbangan yang signifikan, dengan kelas "Prioritas Sedang" sebagai mayoritas (85 sampel) sementara "Prioritas Tinggi" hanya 45 sampel. Setelah penerapan SMOTE, jumlah data pada setiap kelas disamakan menjadi 85 sampel per kelas, sehingga total dataset menjadi 255 sampel. Perbandingan distribusi data sebelum dan sesudah penerapan SMOTE ditampilkan pada Gambar 2.



Gambar 2 Perbandingan Distribusi Kelas Sebelum dan Sesudah SMOTE

Gambar 2 memperlihatkan bahwa penerapan metode SMOTE berhasil menyamakan jumlah data pada setiap kategori tiap kluster menjadi 85 data per kluster. Akibatnya, total data mengalami peningkatan dari semula 205 menjadi 255. Temuan ini mengindikasikan bahwa SMOTE mampu menangani masalah ketidakseimbangan kelas secara efektif, sehingga menghasilkan dataset yang lebih merata dan siap digunakan untuk proses pelatihan model klasifikasi (He & Garcia, 2009).

Analisis dan Evaluasi

Pada tahap ini, model klasifikasi Random Forest dan K-Nearest Neighbors (KNN) diuji menggunakan data asli serta data yang telah seimbang dengan metode SMOTE. Data dibagi menjadi set pelatihan dan pengujian dengan perbandingan 70%:30%, 80%:20%, dan 90%:10%. Selain itu, dilakukan juga validasi silang (Cross-validation) 5-Fold Cross-Validation untuk menguji stabilitas dan generalisasi model serta untuk menilai seberapa efektif model dalam menggeneralisasi ke data baru.

Perbandingan Akurasi dan F1-Score

Evaluasi kinerja algoritma dilakukan dengan membandingkan metrik akurasi dan F1-Score pada tiga skenario pembagian data, yaitu 70:30, 80:20, dan 90:10. Perbandingan ini bertujuan untuk menganalisis konsistensi dan ketahanan model baik pada data asli maupun data yang telah diseimbangkan menggunakan SMOTE. Hasil perbandingan lengkap ditampilkan pada Tabel 4 berikut.

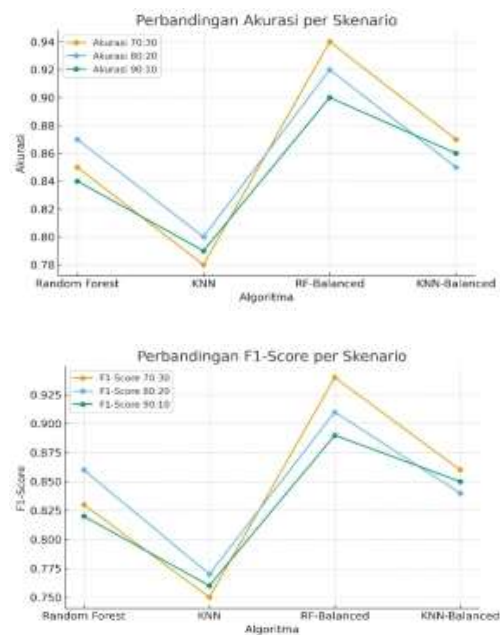
Tabel 4 Hasil Perbandingan

Algoritma	Skenario 70:30		Skenario 80:20		Skenario 90:10	
	A	F1-Score	A	F1-Score	A	F1-Score
RF	0.85	0.83	0.84	0.82	0.87	0.86
KNN	0.78	0.75	0.79	0.76	0.80	0.77
RF-Balanced	0.94	0.94	0.90	0.89	0.92	0.91
KNN-Balanced	0.87	0.86	0.86	0.85	0.85	0.84

Berdasarkan Tabel 4, dapat dianalisis bahwa pada data asli, Random Forest secara konsisten menunjukkan performa yang lebih unggul dibandingkan K-Nearest Neighbors di seluruh skenario pembagian data. Akurasi tertinggi yang dicapai oleh Random Forest pada data asli adalah 87% dalam skenario 80:20.

Gambar 3 Perbandingan Akurasi

Penerapan teknik SMOTE terbukti memberikan peningkatan signifikan terhadap kinerja kedua algoritma (Batista, Prati, & Monard, 2004). Pada data yang telah diseimbangkan, KNN-Balanced mengalami peningkatan akurasi yang cukup besar, meskipun masih berada di bawah RF-Balanced. Sementara itu, Random Forest-Balanced mencapai performa terbaik dengan akurasi dan F1-Score puncak sebesar 94% pada skenario 70:30. Hasil ini mengindikasikan bahwa dengan data yang telah diseimbangkan dan proporsi data uji yang memadai (30%), algoritma Random Forest mampu menghasilkan generalisasi yang sangat baik dan stabil (Rodríguez, Pérez, & Lozano, 2010).



Evaluasi Presisi, Recall, dan F1-Score (Skenario 70:30)

Untuk mengevaluasi kinerja model secara lebih mendalam, khususnya dalam menangani ketidakseimbangan kelas, dilakukan analisis terhadap metrik Presisi, Recall, dan F1-Score pada skenario 70:30.

Skenario ini dipilih karena menunjukkan hasil kinerja tertinggi berdasarkan evaluasi sebelumnya, sekaligus memberikan porsi data uji yang memadai untuk pengujian yang lebih representatif. Hasil perbandingan metrik evaluasi tersebut disajikan pada Tabel 5 di bawah ini

Tabel 5 Evaluasi Presisi, Recall dan F1-Score

Algoritma	Presisi	Recall	F1-Score
Random Forest	0.84	0.82	0.83
K-Nearest Neighbors	0.79	0.77	0.75
Random Forest-Balanced	0.95	0.94	0.94
K-Nearest Neighbors-Balanced	0.88	0.86	0.86

Tabel 5 mengonfirmasi keunggulan Random Forest-Balanced. Nilai Presisi yang tinggi (95%) menunjukkan bahwa ketika model ini memprediksi seorang siswa sebagai "Prioritas Tinggi", prediksinya hampir selalu benar. Nilai Recall yang tinggi (94%) menunjukkan model ini juga sangat efektif dalam menemukan hampir semua siswa yang seharusnya masuk kategori "Prioritas Tinggi". Keseimbangan antara Presisi dan Recall tercermin dari F1-Score yang tinggi (Goutte & Gaussier, 2005).

Confusion Matrix (RF-Balanced Skenario 70:30)

Untuk menganalisis secara lebih rinci jenis kesalahan klasifikasi yang dilakukan oleh model terbaik, yaitu Random Forest-Balanced, pada skenario 70:30, digunakan Confusion Matrix. Analisis ini bertujuan untuk melihat distribusi prediksi model terhadap kelas aktual, khususnya dalam mengidentifikasi pola kesalahan yang terjadi antar kelas. Hasil Confusion Matrix disajikan pada Tabel 6 berikut.

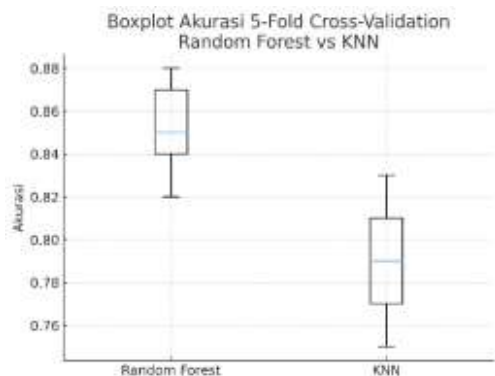
Tabel 6 Confusion Matrix RF-Balanced

Algoritma	Presisi	Recall	F1-Score
Random Forest	0.84	0.82	0.83
K-Nearest Neighbors	0.79	0.77	0.75
Random Forest-Balanced	0.95	0.94	0.94
K-Nearest Neighbors-Balanced	0.88	0.86	0.86

Confusion Matrix pada Tabel 6 menunjukkan bahwa model Random Forest-Balanced memiliki kinerja yang sangat solid. Sebagian besar prediksi terkonsentrasi pada diagonal utama. Sebagian kecil kesalahan terjadi antara kelas yang berdekatan (misalnya, "Prioritas Sedang" diprediksi sebagai "Prioritas Tinggi" atau sebaliknya), yang dapat dimaklumi mengingat kemiripan karakteristik di antara kluster yang berdekatan.

Cross-Validation

Untuk menilai stabilitas dan kemampuan generalisasi model dalam menangani variasi data, dilakukan evaluasi menggunakan metode 5-Fold Cross-Validation pada data asli. Teknik ini membagi dataset menjadi lima bagian yang sama besar, di mana empat bagian digunakan untuk pelatihan dan satu bagian untuk pengujian, dengan proses yang diulang secara bergiliran untuk setiap bagian. Hasil evaluasi menunjukkan bahwa algoritma Random Forest mencapai akurasi rata-rata sebesar 0,85 dengan deviasi standar $\pm 0,03$, sedangkan algoritma K-Nearest Neighbors memperoleh akurasi rata-rata 0,79 dengan deviasi standar $\pm 0,04$.



Gambar 4 Cross-Validation RF dan KNN

Temuan ini memperkuat kesimpulan sebelumnya bahwa Random Forest tidak hanya lebih akurat, tetapi juga lebih stabil dan memiliki kemampuan generalisasi yang lebih baik dibandingkan K-Nearest Neighbors pada dataset ini (Witten, Frank, & Hall, 2011). Deviasi standar yang lebih rendah pada Random Forest mengindikasikan konsistensi performa yang lebih tinggi di berbagai pembagian data, sehingga model ini lebih dapat diandalkan untuk diterapkan pada data baru (Refaelizadeh, Tang, & Liu, 2009).

SIMPULAN

Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan, maka dapat disimpulkan bahwa Algoritma Random Forest (RF) secara konsisten menunjukkan kinerja klasifikasi yang lebih unggul dibandingkan algoritma K-Nearest Neighbors (KNN) dalam mengidentifikasi prioritas penerima manfaat Program Makan Bergizi Gratis di SD Negeri Kota Medan, baik pada data asli maupun data yang telah diseimbangkan. Penerapan teknik Synthetic Minority Over-sampling Technique (SMOTE) terbukti efektif dalam meningkatkan kinerja kedua algoritma, dengan meningkatkan akurasi, presisi, recall, dan f1-score. Peningkatan paling signifikan terjadi pada model KNN, meskipun kinerjanya tetap di bawah RF. Kombinasi Random Forest

dan SMOTE menghasilkan model terbaik (Random Forest-Balanced) dengan akurasi dan f1-score tertinggi sebesar 94% pada skenario pembagian data 70:30. Model ini tidak hanya akurat tetapi juga memiliki keseimbangan yang baik antara presisi dan recall, sehingga sangat direkomendasikan untuk diimplementasikan. Dengan demikian, penggunaan algoritma Random Forest yang didukung oleh teknik penyeimbangan SMOTE dapat menjadi solusi berbasis teknologi yang efektif untuk pendekatan serta efisien untuk membangun sistem klasifikasi prioritas penerima manfaat PMBG yang objektif dan akurat.

DAFTAR PUSTAKA

- Anisa, S., Komarudin, A., & Ramadhan, E. (2024). Sistem klasifikasi untuk menentukan tingkat stres mahasiswa secara online menggunakan metode K-Nearest Neighbors. *Jurnal Informatika Teknologi dan Sains (JINTEKS)*, *6*(3), 568–578.
- Astari, D. F., Chrisandro, Y. H., & Melina. (2023). Klasifikasi tingkat stres saat tidur menggunakan algoritma Random Forest. *Jurnal Mahasiswa Teknik Informatika*, *7*(5), 3676–3684.
- Batista, G. E. A. P. A., Prati, R. C., & Monard, M. C. (2004). A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD Explorations Newsletter*, *6*(1), 20–29.
- Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32.
- Budiarto, M. (2023). Pengaruh ketidakseimbangan data pada kinerja algoritma klasifikasi. *Jurnal Sistem Informasi*, *8*(2), 77–85.
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357.

- Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, *13*(1), 21–27.
- Cutler, A., Cutler, D. R., & Stevens, J. R. (2012). Random forests. Dalam *Ensemble machine learning* (hlm. 157–176). Springer.
- Dinas Pendidikan Kota Medan. (2024). *Data jumlah Sekolah Dasar Negeri per kecamatan*. Medan.
- Febrianto, A., & Sari, L. (2023). Tantangan penyaluran bantuan sosial pendidikan di daerah perkotaan. *Seminar Nasional Administrasi Pendidikan*.
- Firdaus, M. A., & Wibowo, S. K. (2023). Klasifikasi penerima bantuan sosial menggunakan algoritma Naïve Bayes Classifier. *Jurnal Ilmiah Teknologi Informasi*, *12*(3), 145–156.
- Firmansyach, W. A., Hayati, U., & Wijaya, Y. A. (2023). Analisa terjadinya overfitting dan underfitting pada algoritma Naive Bayes dan Decision Tree dengan teknik cross validation. *Jurnal Mahasiswa Teknik Informatika*, *7*(1), 262–269.
- Goutte, C., & Gaussier, E. (2005). A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. Dalam *European Conference on Information Retrieval* (hlm. 345–359).
- Hardi, S., & Prasetyo, T. W. (2024). Analisis ketercapaian program bantuan sosial di Indonesia. *Jurnal Kebijakan Publik*, *15*(2), 112–125.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.
- He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, *21*(9), 1263–1284.
- Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia. (2023). *Panduan pelaksanaan Program Makan Bergizi Gratis*. Jakarta.
- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. Dalam *Proceedings of the 14th International Joint Conference on Artificial Intelligence* (Vol. 2, hlm. 1137–1143).
- Powers, D. M. W. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation. *Journal of Machine Learning Technologies*, *2*(1), 37
- Putra, D. R., & Ningrum, E. S. (2022). Identifikasi keluarga miskin dengan Random Forest berbasis data sensus. *Seminar Nasional Teknologi Informasi dan Komunikasi*, 89–95.
- Rahmasari, F., Rifany, M., Piharyanto, T., & Kurniawan, R. (2024). Perbandingan algoritma machine learning dalam klasifikasi status banjir di Sumatera Utara. *Seminar Nasional Sains Data*, *4*(1), 307–318.
- Rafaelizadeh, P., Tang, L., & Liu, H. (2009). Cross-validation. Dalam *Encyclopedia of database systems* (hlm. 532–538). Springer.
- Rodríguez, J. D., Pérez, A., & Lozano, J. A. (2010). Sensitivity analysis of k-fold cross validation in prediction error estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*(3), 569–575.
- Rofi, M. M., Setiawan, F. A., & Riana, F. (2024). Perbandingan metode K-NN dan Random Forest pada klasifikasi mahasiswa berpotensi dropout. *INFOTECH Journal*, *10*(1), 84–89.
- Santoso, B., & Nurjanah, A. (2024). Komparasi algoritma klasifikasi untuk menentukan penerima bantuan pangan. *Building of Informatics, Technology and Science (BITS)*, *6*(4), 1501–1510.
- Tangkelobo, E., Mayaut, W., Listanto, H., Binanto, I., & Sianipar, N. F. (2023). Perbandingan algoritma klasifikasi Random Forest, Gaussian Naive

-
- Bayes, dan K-Nearest untuk data tidak seimbang. *Seminar Nasional Teknik Elektro, Informatika dan Sistem Informasi (SINTaKS)*.
- Wijaya, R., & Puspita, D. (2024). Penerapan algoritma Decision Tree C4.5 untuk klasifikasi penerima bantuan tunai. *Jurnal Teknologi Informasi*, *10*(1), 45–56.
- Wijiyano, W., Pradana, A. I., Sopingi, S., & Atina, V. (2024). Teknik K-Fold cross validation untuk mengevaluasi kinerja mahasiswa. *Jurnal Algoritma*, *21*(1).
- Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data mining: Practical machine learning tools and techniques* (3rd ed.). Morgan Kaufmann.
- World Bank. (2022). *Global report on food security and nutrition*. Washington, D.C