Universitas Ahmad Dahlan

Jurnal Ilmiah Matematika
Vol. 10, No. 1, Mei 2023, pp. 43-52
ISSN 2774-3241
http://journal.uad.ac.id/index.php/Konvergensi

Analisis Performa Algoritma Smote-Bagging Dalam Klasifikasi Data Tidak Seimbang Dengan Metode Chi-Square Automatic Interaction Detection (CHAID)

Tyas Kusuma Argani [a,1,*]

[*a] Universitas Ahmad Dahlan, Yogyakarta, Indonesia;
[1] tyas2000015032@webmail.uad.ac.id
*Correspondent Author

Received: *Revised:* *Accepted:*

KATAKUNCI

CHAID
Data tidak seimbang
Klasifikasi
SMOTE-Bagging
Stunting

ABSTRAK

Klasifikasi data tidak seimbang sering menghadapi tantangan dalam mencapai keseimbangan antara sensitivitas dan spesifisitas. Penelitian ini menganalisis performa algoritma *SMOTE-Bagging* pada klasifikasi data tidak seimbang menggunakan metode *Chi-Square Automatic Interaction Detection* (CHAID), dengan studi kasus stunting pada balita tahun 2022 di Bojongsoang. SMOTE (*Synthetic Minority Over-sampling Technique*) digunakan untuk meningkatkan representasi kelas minoritas dalam dataset, kemudian digabungkan dengan teknik *Bagging* untuk meningkatkan kinerja klasifikasi. Hasil penelitian menunjukkan bahwa algoritma *SMOTE-Bagging* CHAID meningkatkan performa dalam klasifikasi data tidak seimbang, dengan peningkatan sensitivitas sebesar 65%, *Area Under Curve* (AUC) sebesar 42%, dan keseimbangan antara sensitivitas dan spesifisitas (*G-Mean*) sebesar 71%. Implementasi *SMOTE-Bagging* meningkatkan sensitivitas dan memberikan keseimbangan yang lebih baik antara sensitivitas dan spesifisitas.

KEYWORDS

CHAID
Classification
Imbalanced Data
SMOTE-Bagging
Stunting

ABSTRACT

Classification of imbalanced data often faces challenges in achieving a balance between sensitivity and specificity. This study analyzes the performance of the SMOTE-Bagging algorithm in classifying imbalanced data using the Chi-Square Automatic Interaction Detection (CHAID) method, with a case study of stunting in toddlers in 2022 in Bojongsoang. SMOTE (Synthetic Minority Over-sampling Technique) is used to enhance the representation of minority classes in the dataset, subsequently combined with Bagging to improve classification performance. The research findings demonstrate that the SMOTE-Bagging CHAID algorithm enhances performance in imbalanced data classification, with a 65% increase in sensitivity, an Area Under Curve (AUC) of 42%, and a balanced sensitivity-specificity trade-off (G-Mean) of 71%. The implementation of SMOTE-Bagging improves sensitivity and provides better balance between sensitivity and specificity.

This is an open-access article under the CC–BY-SA license.

 ISSN 2774-3241

Pendahuluan

Klasifikasi merupakan proses analisis data untuk memprediksi label kelas pada sampel yang akan dikategorikan [1]. Tantangan utama dalam proses klasifikasi adalah bagaimana menghasilkan model dengan akurasi tinggi dan kesalahan prediksi yang rendah. Untuk mengatasi tantangan ini, salah satu pendekatan yang banyak digunakan adalah metode ensemble, seperti Bagging atau Bootstrap Aggregating, yang bertujuan untuk meningkatkan akurasi dan mengurangi error prediksi [2] [3]. Bagging bekerja dengan cara menghasilkan beberapa model dari variasi data pelatihan yang dibentuk melalui proses bootstrap, sehingga tiap model belajar dari sampel berbeda dan memberikan hasil agregasi yang lebih stabil.

Algoritma klasifikasi seperti Decision Tree, Naïve Bayes, dan Logistic Regression telah banyak dimanfaatkan dalam berbagai kasus di dunia nyata [3]. Salah satu metode Decision Tree yang cukup populer adalah CHAID (Chi-square Automatic Interaction Detection) yang mampu mengelompokkan data berdasarkan hasil uji chi-square terhadap variabel dependen [4]. Penelitian sebelumnya menunjukkan bahwa metode Bagging yang dikombinasikan dengan CHAID mampu meningkatkan akurasi prediksi [5] [6]. Namun demikian, metode CHAID yang berbasis data kategorikal cenderung sensitif terhadap ketidakseimbangan kelas, di mana jumlah data dalam satu kelas jauh lebih banyak dibandingkan kelas lainnya [7].

Permasalahan data tidak seimbang ini dapat diatasi melalui pendekatan berbasis data seperti SMOTE (Synthetic Minority Oversampling Technique) yang menciptakan data sintetis pada kelas minoritas agar seimbang dengan kelas mayoritas [8]. Penerapan SMOTE terbukti efektif dalam meningkatkan performa berbagai algoritma klasifikasi seperti K-Nearest Neighbor, Naïve Bayes, CHAID, dan Random Tree [9] [10] [11]. Penggabungan SMOTE dengan metode ensemble seperti Bagging (SMOTE-Bagging) telah terbukti lebih optimal dalam menangani ketidakseimbangan data dan meningkatkan performa klasifikasi, khususnya dalam memaksimalkan perhatian terhadap kelas minoritas [12] [13].

Penelitian ini menggunakan varian SMOTE yang lebih lanjut, yaitu SMOTE-NC (Synthetic Minority Oversampling Technique – Nominal Continuous), yang mampu memperlakukan variabel kategorikal dan numerik secara terpisah, serta mempertahankan label asli dari fitur

 ISSN 2774-3241

kategorikal [14]. SMOTE-NC telah terbukti meningkatkan sensitivitas prediksi pada data tidak seimbang [15] [16].

Masalah stunting pada anak balita merupakan salah satu tantangan besar di bidang kesehatan di Indonesia, di mana pertumbuhan anak terganggu akibat kekurangan gizi kronis dalam 1000 hari pertama kehidupan [17]. Oleh karena itu, pemanfaatan machine learning sebagai sistem prediksi stunting sangat penting untuk mendukung upaya pencegahan yang lebih tepat sasaran [18].

Berdasarkan kajian literatur tersebut, penelitian ini akan menganalisis performa algoritma SMOTE-Bagging dalam mengklasifikasikan data tidak seimbang menggunakan metode CHAID, khususnya dalam kasus prediksi stunting pada balita.

Metode

Penelitian ini diawali dengan melakukan tinjauan literatur untuk memperoleh pemahaman yang mendalam terkait metode yang akan digunakan dan permasalahan yang diangkat dalam penelitian. Selanjutnya dilakukan pengambilan data, kemudian data tersebut dianalisis dan dipahami untuk mengenali karakteristik serta permasalahan yang ada, khususnya terkait ketidakseimbangan kelas. Data yang telah dipahami kemudian melalui tahap pre-processing agar siap digunakan dalam proses analisis. Setelah itu, dilakukan dua pendekatan implementasi, yaitu implementasi algoritma CHAID secara langsung dan implementasi kombinasi CHAID dengan metode SMOTE-Bagging. Kedua hasil implementasi tersebut kemudian dievaluasi dan dibandingkan untuk menilai performa klasifikasi yang dihasilkan. Tahapan ini berakhir dengan penyusunan hasil evaluasi sebagai dasar pengambilan kesimpulan

Hasil dan Pembahasan

1. Preprocessing Data

Langkah-langkah preprocessing meliputi pemilihan fitur, deteksi missing value, penggantian nilai unik pada variabel Usia Saat Ukur, Jenis Kelamin, dan TB/U, serta pemisahan data training dan data testing.

a. Pemilihan fitur

Proses pemilihan fitur pada penelitian ini didasarkan pada pedoman tatalaksana gizi yang berjudul “Petunjuk Teknis Pemberian Makanan Tambahan (PMT) Berbahan Pangan Lokal untuk Balita dan Ibu Hamil” oleh kemenkes RI dan website World Health Organization (WHO)

 ISSN 2774-3241

dengan alamat website https://www.who.int/tools/child-growth-standards/standards sehingga dipilih fitur Usia Saat Ukur, Berat, Tinggi, BB/U, BB/TB, dan TB/U sebagai variabel independen.

b. Deteksi missing value

Dalam penelitian ini, variabel 41 dengan missing value dihapus, sehingga digunakan total 6672 data setelah deteksi dan penanganan missing value.

c. Mengganti nilai unik

Pada tahap ini, nilai unik dalam variabel Usia Saat Ukur, Jenis Kelamin, dan TB/U akan diubah untuk memudahkan proses selanjutnya. Nilai unik dari variabel Usia Saat Ukur diubah dari satuan tahun, bulan, dan hari menjadi satuan hari saja untuk mempermudah penggunaan dalam proses SMOTE-NC. Selain itu, perubahan pada variabel Jenis Kelamin dan TB/U dilakukan untuk memberikan gambaran yang lebih jelas. Khususnya, pada variabel TB/U, nilai unik "Stunting" digunakan sebagai istilah umum yang mencakup semua kasus pertumbuhan anak yang terhambat, baik itu pendek maupun sangat pendek.

d. Split data

Rasio pemisahan data dapat bervariasi berdasarkan jumlah data, kompleksitas masalah, dan tujuan model. Dalam penelitian ini, dataset dibagi menjadi 70% data training dan 30% data testing, menghasilkan 4670 data training dan 2002 data testing.

2. Klasifikasi Data dengan Algoritma SMOTE-Bagging CHAID

Langkah awal pada tahap ini adalah memisahkan kelas mayoritas dan minoritas. Kelas mayoritas melalui proses *bootstrapping*, sementara kelas minoritas di-*resampling* menggunakan SMOTE-NC. Hasil dari kedua proses ini kemudian akan digabungkan dan nilai unik varibel kontinu dalam dataset gabungan tersebut akan ditransformasi menjadi variabel nominal untuk mempermudah klasifikasi menggunakan metode CHAID. Tahap akhirnya adalah melakukan *voting* hasil uji dari setiap pohon yang telah dilatih menggunakan data *testing*.

a. Proses SMOTE-NC

Proses SMOTE-NC dilakukan dengan beberapa tahap yaitu perhitungan median, perhitungan tetangga terdekat, dan membuat data sintesis. Hasil dari pembuatan data sintesis berupa suatu nilai dari variabel kontinu yang disajikan pada tabel berikut :

Table 1. Variabel Kontinu Data Sintetis

No. Urut	Usia Saat Ukur	*Berat*	*Tinggi*
Sintetis	179	6,12	61,18

Tabel 1 karena satuan pada variabel Usia Saat Ukur adalah hari, maka dilakukan pembulatan ke atas pada nilai dalam variabel tersebut. Kemudian data sintetis pada variabel nominal diambil dari nilai yang paling sering muncul pada masing-masing variabel nominal k tetangga terdekat, sehingga didapatkan hasil pada Tabel 2 sebagai berikut.

Tabel 2. Variabel Kontinu dan Variabel Nominal pada Data Sintetis

No. *Urut*	Usia Saat *Ukur*	*Berat*	*Tinggi*	Jenis *Kelamin*	*BB/U*	*BB/TB*	*TB/U*
Sintetis	179	6,12	61,18	Laki-Laki	Berat Badan Normal	Gizi Baik	Stunting

Lakukan proses SMOTE-NC tersebut secara berulang hingga mencapai jumlah data yang diinginkan. Pada penelitian ini, data sintetis dibuat sebanyak 2263 data, sehingga jumlah kelas minoritas pada data *training* menjadi sebanyak 2310 data.

b. Perhitungan Bootstraping

Pada tahap ini, kelas mayoritas yang telah dipisahkan akan menjalani proses bootstrapping. Dalam penelitian ini, data training akan di-bootstrapping, yaitu mengambil sampel acak dengan pengembalian sebanyak 20 kali. Setiap sampel 47 bootstrap akan terdiri dari 2310 data dengan variabel target Normal. Hasil dari bootstrapping ini akan digabungkan dengan hasil dari SMOTE-NC untuk membentuk 20 sampel ($\{z_1^1, z_2^1, \ldots, z_{4620}^1\}$, $\{z_1^2, z_2^2, \ldots, z_{4620}^2\}, \ldots, \{z_1^{20}, z_2^{20}, z_{4620}^{20}\}$) yang akan dilatih menggunakan metode CHAID.

c. Transformasi Nilai Unik menjadi Kategorikal

Pada tahap ini, variabel kontinu pada sampel-sampel baru dan data *testing* akan diubah menjadi variabel nominal. Variabel Usia Saat Ukur akan diubah menjadi lima kategori berdasarkan tahun, sedangkan variabel berat dan tinggi akan dikategorikan menjadi tiga kategori berdasarkan quartil atas dan quartil bawah seperti pada Tabel 3 berikut:

Tabel 3. Perubahan Kategori dalam Fitur Usia Saat Ukur, Jenis Kelamin, dan TB/U

	Nilai Unik dalam Fitur	
	Sebelum	***Sebelum***
Usia Saat Ukur	829, 895, 1727, …, 360, 185, 46	'< 1 Tahun', '< 2 Tahun', '< 3 Tahun', '< 4 Tahun', '< 5 Tahun'
Berat	12,7, 13,1, 15, …, 13,95, 14,01	'10.20 - 15.00', '< 10.20', '> 15.00'
Tinggi	89,6, 91,2, 107, 76, …, 108,8, 82,7, 85,3	'< 81.00', '81.00 - 99.00', '> 99.00'

 ISSN 2774-3241

d. Membentuk model Chi-Square Automatic Interaction Detection (CHAID)

Pada proses ini akan dibuat 20 pohon keputusan dari 20 sampel hasil SMOTE dan Bootstrapping melalui tiga tahap yaitu penggabungan, pemisahan dan penghentian. Dengan syarat minimum jumlah pengamatan pada child node sebanyak 50 data dan maksimal kedalaman pohon yaitu 10. Berikut merupakan proses pembuatan pohon pertama:

1) Pemisahan

Proses ini digunakan untuk menentukan root node, yang mana semua prosesnya dilakukan untuk semua variabel independent dan cari nilai p-value. Hasilnya disajikan pada tabel

Table 4. Nilai Statistik Chi-Square Semua Variabel Independen terhadap Variabel Dependen TB/U

Variabel Independen	***Chi-Square***	***p-value***	***Keputusan***
Tinggi	843,555	6.674263e-184	H_0 ditolak
BB/U	629,374	4.318413e-136	H_0 ditolak
Berat	523,342	2.278915e-114	H_0 ditolak
BB/TB	172,86	1.790010e-35	H_0 ditolak
Usia Saat Ukur	134,934	3.426683e-28	H_0 ditolak
Jenis Kelamin	29,103	6.862691e-08	H_0 ditolak

Tabel 4 menunjukkan bahwa semua variabel independen memiliki hubungan dengan variabel dependen. Untuk menentukan root nod, pilih variabel yang memiliki nilai chi-square terbesar dan nilai *p*-value terendah, yaitu variabel Tinggi. Kategori pada variabel Tinggi kemudian digunakan untuk memisahkan dan menentukan cabang atau aturan selanjutnya yang disebut juga child node. Proses pemisahan ini akan berlanjut pada setiap node selama masih ada variabel independen yang signifikan.

2) Penggabungan

Dalam pohon pertama, terdapat beberapa variabel independen yang mengalami proses penggabungan yaitu variabel BB/U pada kedalaman kedua (hasil pemisahan dari kategori < 81.00 pada variabel Tinggi) dan variabel Usia Saat ukur pada kedalaman ketiga dan keempat (hasil pemisahan dari kategori-kategori pada variabel BB/U dan hasil pemisahan dari kategori >15.00 pada variabel Berat). Berikut pegujian statistik *chi- square* pada pasangan kombinasi kategori pada variabel BB/U:

Tabel 5. Nilai statistik *chi-square* TB/U dan BB/U

Kategori	***Chi-Square***	***p-value***	***Keputusan***
Kurang*Risiko Lebih	249.249.387	3.785132e-56	H_0 ditolak

 ISSN 2774-3241

Berat Badan Normal*Kurang	115.000.709	7.866932e-27	H_0 ditolak
Risiko Lebih*Sangat Kurang	86.000.000	1.799356e-20	H_0 ditolak
Berat Badan Normal*Risiko Lebih	62.917648	2.155324e-15	H_0 ditolak
Berat Badan Normal*Sangat Kurang	26.838.922	2.211384e-07	H_0 ditolak
Kurang*Sangat Kurang	1.341.574	2.467565e-01	H_0 diterima

Dari Tabel 5, diperoleh nilai χ^2 hitung untuk uji antara variabel TB/U dan BB/U. Pada kategori Kurang dan Sangat Kurang dimana H_0 diterima, menunjukkan bahwa variabel TB/U dan BB/U pada kategori Kurang dan Sangat Kurang saling bebas dan karenanya digabung menjadi satu kategori campuran. Untuk kategori lainnya, H_0 ditolak, yang menunjukkan bahwa kedua variabel tidak saling bebas. Namun, karena jumlah pengamatan pada kategori Risiko Lebih terhadap kategori Stunting adalah 0, maka kategori Risiko Lebih digabung juga dengan Berat Badan Normal.

Pengali Bonferroni untuk variabel BB/U ini dihitung sebagai berikut :

$$M = \sum_{i=0}^{r-1}(-1)^i \frac{(r-i)^c}{i!(r-i)!} = \sum_{i=0}^{2-1}(-1)^i \frac{(2-i)^4}{i!(2-i)!} = 7 \quad (1)$$

Setelah menghitung pengali Bonferroni, nilai tingkat kesalahan untuk uji signifikansi antara variabel TB/U dan variabel BB/U pada kategori campuran {Risiko Lebih, Berat Badan Normal} dan {Kurang, Sangat Kurang} adalah hasil kali p-value dan M, yaitu (0,000)(7) = 0,000. Oleh karena itu, keputusan yang diambil adalah menolak $H0$ karena nilai p-value terkoreksi tetap lebih kecil dari α = 5%, yang berarti variabel BB/U dengan kategori campuran tersebut tidak saling bebas. Proses ini diulang untuk semua variabel independen lainnya.

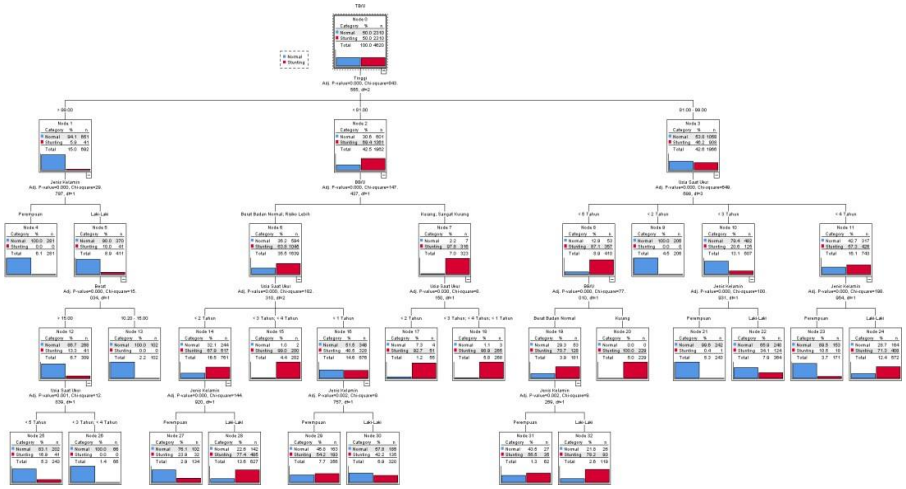
3) Penghentian

Pada pohon pertama, tahap penghentian terjadi pada node 4, 9, 13, 15, 17, 18, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, dan 32 karena tidak ada variabel independen yang signifikan berdasarkan uji chi-square untuk memisahkan variabel dependen di node tersebut. Proses ini juga diterapkan pada 19 pohon lainnya $(T_1(.), T_2(.), \ldots, T_{20}(.))$. Sehingga bentuk pohon CHAID SMOTE-Bagging pertama disajikan pada gambar.

 ISSN 2774-3241

Setelah menghitung pengali Bonferroni, nilai tingkat kesalahan untuk uji signifikansi antara variabel TB/U dan variabel BB/U pada kategori campuran {Risiko Lebih, Berat Badan Normal} dan {Kurang, Sangat Kurang} adalah hasil kali p-value dan M, yaitu (0,000)(7) = 0,000 . Oleh karena itu, keputusan yang diambil adalah menolak H_0 karena nilai p-value terkoreksi tetap lebih kecil dari α = 5%, yang berarti variabel BB/U dengan kategori campuran tersebut tidak saling bebas. Proses ini diulang untuk semua variabel independen lainnya.

Gambar 1. Diagram Pohon CHAID SMOTE-Bagging Pertama



e. Proses Aggregating

Setelah proses bootstrapping menjadi 20 sampel bootstrap dengan ukuran 4620 data dan melatih 20 model CHAID, kemudian 20 pohon tersebut digabungkan menjadi satu sistem klasifikasi, dan uji data testing dilakukan melalui proses penghitungan suara mayoritas berikut:

$$S_{20}(.) = \arg\max_{k}[card(l|T_i(.) = k|l = 1, 2, \dots, 20)] \quad (2)$$

3. Perbandingan Performa Metode Klasifikasi

Setelah menganalisis data menggunakan algoritma SMOTE-Bagging dengan metode klasifikasi CHAID, langkah selanjutnya adalah membandingkan kinerja metode klasifikasi CHAID dengan dan tanpa menggunakan algoritma SMOTE- Bagging. Perbandingan ini diukur menggunakan kinerja metode klasifikasi yang meliputi F_1 -*Macro Average*, F_1 -*Weighted Average*, sensitivitas, spesifitas, G- *Mean*, dan AUC. Hasil perbandingan klasifikasi ini ditampilkan pada Tabel 6.

Table 6. Perbandingan Kinerja Metode Klasifikasi CHAID
dengan dan tanpa Algoritma SMOTE-Bagging

 ISSN 2774-3241

Model	***F1-Macro Average***	***F1-Weigted Average***	***Sensitivitas***	***Spesitivitas***	***G-Mean***	***AUC***
CHAID	0,50	0,99	0	1	0	0,50
SMOTE-Bagging CHAID	0,46	0,87	0,65	0,78	0,71	0,71

Berdasarkan Tabel 6 dapat dilihat bahwa F_1-*Macro Average* sedikit menurun pada SMOTE-Bagging CHAID, menunjukkan bahwa model ini tidak memperbaiki keseimbangan klasifikasi untuk semua kelas. F_1 -*Weighted Average* juga lebih rendah tetapi tetap menunjukkan performa yang baik, menandakan bahwa CHAID lebih akurat untuk kelas dominan, sementara SMOTE-Bagging CHAID mengorbankan sedikit akurasi untuk keseimbangan yang lebih baik. Sensitivitas meningkat signifikan pada SMOTE-Bagging CHAID, menunjukkan peningkatan dalam identifikasi kelas minoritas. Namun, spesifisitas menurun, menandakan 56 peningkatan false positive pada kelas mayoritas. G-Mean lebih tinggi pada SMOTE-Bagging CHAID, menunjukkan keseimbangan yang lebih baik antara sensitivitas dan spesifisitas. AUC juga lebih tinggi, menunjukkan kemampuan yang lebih baik dalam membedakan antara kelas positif dan negatif.

Simpulan

Dalam evaluasi model klasifikasi untuk kasus stunting balita tahun 2022 di Bojongsoang, penerapan algoritma SMOTE-Bagging CHAID menunjukkan hasil yang signifikan. Penerapan teknik ini berhasil meningkatkan sensitivitas sebesar 65%, AUC sebesar 42%, dan G-Mean sebesar 71%, meskipun ada sedikit penurunan pada metrik *F*1-Macro Average dan spesifisitas. Tanpa SMOTE-Bagging, metode CHAID menghasilkan klasifikasi yang sangat bias, dengan spesifisitas mencapai 100% dan sensitivitas mendekati 0%, yang menunjukkan ketidakseimbangan data. Setelah menggunakan SMOTE-Bagging, sensitivitas meningkat secara signifikan, meskipun spesifisitas sedikit menurun sebesar 22%. Namun, keseimbangan antara keduanya (G-Mean) dan AUC justru meningkat, memperlihatkan kemampuan model yang lebih baik dalam membedakan kelas positif dan negatif. Implementasi SMOTE-Bagging terbukti efektif dalam meningkatkan performa dan keseimbangan klasifikasi pada data yang tidak seimbang.

Daftar Pustaka

[1] A. Srirahayu and L. S. Pribadie, "Review Paper Data Mining Klasifikasi Data Mining," *J. Ilm. Inform. Glob.*, vol. 14, no. 1, 2023, doi: 10.36982/jiig.v14i1.2981.

[2] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, 1996, doi: 10.1023/A:1018054314350.

[3] Y. Pristyanto and A. A. Zein, "Model Balanced Bagging Berbasis Decision Tree Pada Dataset Imbalanced Class," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 12, no. 1, pp. 9–15, 2023, doi: 10.32736/sisfokom.v12i1.1399.

[4] C. A. Gallagher, H. M. Monroe, and J. L. Fish, "An Iterative Approach To Classification Analysis," *J. Appl. Stat.*, vol. 29, pp. 237–280, 2000.

[5] A. Fitrianto, W. Z. A. Wan Muhamad, and B. Susetyo, "Development of direct marketing strategy for banking industry: The use of a Chi-squared Automatic Interaction Detector (CHAID) in deposit subscription classification," *J. Socioecon. Dev.*, vol. 5, no. 1, p. 64, 2022, doi: 10.31328/jsed.v5i1.3420.

[6] C. Deng, F. Yi, X. Li, J. Tang, and G. Sun, "Performance Analysis of CHAID Algorithm for Accuracy," *Proc. - 2023 Int. Conf. Pattern Recognition, Mach. Vis. Intell. Algorithms, PRMVIA 2023*, pp. 182–186, 2023, doi: 10.1109/PRMVIA58252.2023.00036.

[7] "View of SMOTE_ Synthetic Minority Over-sampling Technique.pdf."

[8] Fithria Siti Hanifah, Hari Wijayanto, and Anang Kurnia, "SMOTE bagging algorithm for imbalanced dataset in logistic regression analysis (case: Credit of bank X)," *Appl. Math. Sci.*, vol. 9, no. 137–140, pp. 6857–6865, 2015, doi: 10.12988/ams.2015.58562.

[9] V. Rattan, R. Mittal, J. Singh, and V. Malik, "Analyzing the application of SMOTE on machine learning classifiers," *2021 Int. Conf. Emerg. Smart Comput. Informatics, ESCI 2021*, no. March, pp. 692–695, 2021, doi: 10.1109/ESCI50559.2021.9396962.

[10] P. Chittora *et al.*, "Prediction of Chronic Kidney Disease - A Machine Learning Perspective," *IEEE Access*, vol. 9, no. January, pp. 17312–17334, 2021, doi: 10.1109/ACCESS.2021.3053763.

[11] F. Novitasari, E. Haerani, A. Nazir, J. Jasril, and F. Insani, "Sistem Klasifikasi Penyakit Jantung Menggunakan Teknik Pendekatan SMOTE Pada Algoritma Modified K-Nearest Neighbor," *Build. Informatics, Technol. Sci.*, vol. 5, no. 1, pp. 274–284, 2023, doi: 10.47065/bits.v5i1.3610.

[12] D. S. Sisodia and U. Verma, "Distinct multiple learner-based ensemble smotebagging (ML-ESB) method for classification of binary class imbalance problems," *Int. J. Technol.*, vol. 10, no. 4, pp. 721–730, 2019, doi: 10.14716/ijtech.v10i4.1743.

[13] N. Iriawan *et al.*, "On The Comparison: Random Forest, SMOTEBagging, and Bernoulli Mixture to Classify Bidikmisi Dataset in East Java," *2018 Int. Conf. Comput. Eng. Netw. Intell. Multimed.*, 2018, doi: 10.1109/CENIM.2018.8711035.

[14] M. Mukherjee and M. Khushi, "Smote-enc: A novel smote-based method to generate synthetic data for nominal and continuous features," *Appl. Syst. Innov.*, vol. 4, no. 1, 2021, doi: 10.3390/asi4010018.

[15] D. T. Utari, "Integration of Svm and Smote-Nc for Classification of Heart Failure Patients," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 17, no. 4, pp. 2263–2272, 2023, doi: 10.30598/barekengvol17iss4pp2263-2272.

[16] E. C. Gök and M. O. Olgun, "SMOTE-NC and gradient boosting imputation based random forest classifier for predicting severity level of covid-19 patients with blood samples," *Neural Comput. Appl.*, vol. 33, no. 22, pp. 15693–15707, 2021, doi: 10.1007/s00521-021-06189-y.

[17] D. D. Astuti, R. B. Adriani, and T. W. Handayani, "Pemberdayaan Masyarakat Dalam Rangka," vol. 4, no. 2, pp. 2–6, 2020.