Naskah Publikasi-Eko Prasetyo-22.01.85.0007-22022024-TH

by Fahmi Akbar

Submission date: 22-Feb-2024 01:33PM (UTC+0700)
Submission ID: 2301400786
File name: Naskah_Publikasi_Eko_Prasetyo_2201850007.pdf (359.59K)
Word count: 4027
Character count: 23937

Techno.COM, Vol. 23, No. 1, Februari 2024: 1-10

Optimasi Klasifikasi Data Stunting Melalui Ensemble Learning pada Label Multiclass dengan Imbalance Data

Optimizing Stunting Data Classification Through Ensemble Learning on Multiclass Labels with Imbalance Data

Eko Prasetyo[1], Kristiawan Nugroho[2]
[1,2]Fakultas Teknologi Informasi dan Industri, Universitas Stikubank
[1]ekoprasetyo0007@mhs.unisbank.ac.id, [2]kristiawan@edu.unisbank.ac.id

Abstrak

Salah satu permasalahan kesehatan yang sering ditemui di banyak negara termasuk Indonesia adalah stunting. Stunting telah mendapat banyak perhatian di Indonesia, terlihat dari alokasi APBN masing-masing sebesar Rp48,3 triliun dan Rp49,4 triliun pada tahun 2022 dan 2023 untuk bidang ini. Pada tahun 2022, Kementerian Kesehatan merilis temuan dari Survei Status Gizi Indonesia (SSGI) yang menyatakan bahwa angka stunting di Indonesia mencapai 21,6% pada saat Rapat Kerja Nasional BKKBN pada 25 Januari 2023. Hal ini menunjukkan pentingnya untuk mengerti pemahaman mendalam tentang faktor-faktor yang mengidentifikasi anak-anak berisiko tinggi terkena stunting. Banyak penelitian sebelumnya yang membahas faktor resiko stunting, namun masih sedikit penerapannya dalam metode machine learning, dalam data yang kompleks dan tidak seimbang. Penelitian ini mengevaluasi kinerja dari berbagai metode machine learning yang bertujuan dapat memberikan kontribusi penting dalam bidang kesehatan anak dan analisis data. Diantara metode machine learning yang dipilih metode Bagging Decision Tree mendapatkan nilai accuracy yang terbaik sebesar 78,93%, precision 78% dan recall sebesar 77,99%. Dalam penelitian ini menunjukkan bahwa metode ensemble learning mampu bekerja dengan baik dalam atribut multiclass dan data yang tidak seimbang pada dataset pertumbuhan balita.

Kata kunci: Machine Learning, Bagging, Boosting, Stacking, Imbalance

Abstract

A health issue that is frequently seen in many nations including Indonesia is stunting. Stunting has drawn a lot of attention in Indonesia, as seen by the IDR 48.3 trillion and IDR 49.4 trillion APBN allocations in 2022 and 2023, respectively, for this area. In 2022 the Ministry of Health released findings from the Indonesian Nutrition Status Survey (SSGI) stating that the stunting rate, in the country stood at 21.6% during the National Working Meeting of BKKBN on January 25 2023. This demonstrates how crucial it is to have a thorough grasp of the characteristics that set children up for stunting. Stunting risk variables have been the subject of several prior research, but their use in machine learning techniques for complicated and unbalanced data is still rather limited. This study assesses the effectiveness of many machine learning techniques with the goal of significantly advancing data analysis and child health. The Bagging Decision Tree method outperformed the other machine learning techniques, with the highest accuracy (78.93%), precision (78%), and recall (77.99%). This study demonstrates that the toddler growth dataset's multiclass traits and imbalanced data may be effectively handled by the ensemble learning approach.

Keywords: Machine Learning, Bagging, Boosting, Stacking, Imbalance

1. PENDAHULUAN

Masalah gizi yang sering dialami banyak anak di berbagai negara, terutama di Indonesia, adalah stunting [1]. Masalah ini disebabkan oleh kekurangan asupan gizi yang cukup pada masa pertumbuhan, yang mengakibatkan gangguan pertumbuhan fisik dan perkembangan pada anak

[2,3]. Stunting tidak hanya berdampak jangka panjang pada kesehatan fisik, perkembangan kognitif, motoric [4] dan produktifitas anak di kemudian hari, tetapi juga dapat berpotensi mengancam pembangunan sosial dan ekonomi suatu Negara [3]. Oleh karena itu, pemahaman mendalam tentang faktor-faktor yang berkontribusi terhadap stunting dan upaya untuk mengidentifikasi anak-anak yang berisiko tinggi terkena stunting sangat penting.

Dalam era digital yang semakin maju, kita dihadapkan pada jumlah data yang besar, kompleks [5] dan terdistribusi secara tidak merata dengan banyaknya variabel yang saling berkaitan [6]. Hal ini berakibat sulitnya dalam pengolahan data, proses analisis yang memerlukan waktu cukup lama dan ketrampilan analisis data yang tinggi. Oleh sebab itu, sangat penting untuk memiliki metode klasifikasi yang efektif dan akurat untuk mengambil informasi berharga dari data tersebut [7].

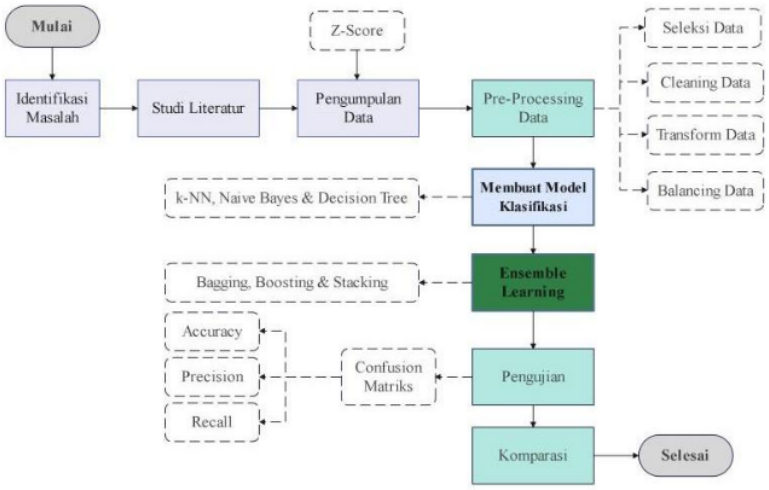
Beberapa teknik machine learning telah diterapkan pada klasifikasi data stunting pada penelitian sebelumnya [6,7,8,9]. Meski begitu, performa model pada penelitian sebelumnya masih dapat ditingkatkan lebih lanjut. Salah satu metode yang efektif untuk meningkatkan akurasi dan keandalan klasifikasi adalah menggunakan metode ensamble learning, dengan cara menggabungkan beberapa model algoritma klasifikasi machine learning [10]. Metode ensemble learning terbukti efektif dalam meningkatkan akurasi seperti dalam pengenalan pola [11], anallisis citra medis [10,12] dan pemprosesan bahasa alami [13,14]. Pada penerapannya, ensemble learning dapat memperkuat kekurangan model tunggal dalam algoritma klasifikasi dan meningkatkan prediksi dengan menggabungkan hasil keluaran dari beberapa model algoritma klasifikasi [10]. Oleh karena itu, untuk meningkatkan akurasi klasifikasi data stunting pada balita, penelitian ini mengkaji penggunaan teknik ensemble learning seperti *Bagging, Boosting* dan *Stacking*. Hal ini dimaksudkan agar permasalahan umum seperti ketidakseimbangan data [15], kompleksitas model [16] dan generalisasi yang buruk dapat diatasi dengan memanfaatkan metode ensemble learning.

Berbagai riset mengenai stunting sudah banyak dilakukan, antara lain pada tahun 2020, Talukder dan Ahammed dalam penelitiaannya menggunakan berbagai macam algoritma machine learning dalam analisis survey demografi dan kesehatan di Bangladesh, dapat dianggap memprediksi secara akurat status gizi pada anak-anak. Algoritma Random Forest memberikan hasil terbesar dengan tingkat akurasi sebesar 68,51%, sensivitas 94,66% dan spesifisitas 69,76%, berdasarkan metric pengukuran kinerja dalam penelitian [17]. Penelitian lainnya dilakukan oleh Khan, dkk pada tahun 2021, mengkaji penerapan metode machine learning untuk memprediksi stunting di Bangladesh menggunakan algoritma klasifikasi *Gradient Boosting, Random Forests*, *SVM, Clasification Tree* dan *Logistic Regression*. Hasil dari penelitian mendapatkan bahwa *Gradient Boosting* memiliki performa yang terbaik sebesar 68,47% dibanding dengan algoritma klasifikasi lainnya. Selain itu dalam penelitian, Khan juga mengungkapkan bahwa karakteristik orang tua, gizi dan rumah tangga merupakan faktor penting dalam memprediksi stunting [18].

Penelitian ini akan menggunakan beberapa teknik klasifikasi machine learning, termasuk Decision Tree, Naive Bayes, dan k-Nearest Neighbor (k-NN). Pemilihan metode k-NN dikarenakan meskipun telah banyak digunakan dalam berbagai aplikasi klasifikasi [19,20], tetapi masih ada tantangan dalam mengoptimalkan kinerjanya, terutama ketika digunakan untuk memprediksi stunting pada anak-anak. Naïve Bayes adalah teknik klasifikasi probabilistik untuk mengklasifikasikan data menggunakan teorema Bayes berdasarkan probabilitas atribut-atribut yang ada dalam data [21]. Selain itu, dari semua algoritma *machine learning* yang telah diujikan dengan teknik normalisasi data, Naïve Bayes merupakan algoritma yang memiliki performa paling stabil [8], karena hal itulah metode naïve bayes digunakan dalam penelitian ini. Metode selanjutnya Decision Tree, merupakan metode klasifikasi berbasis pohon keputusan yang membantu dalam analisis data dimana variabel dataset mempunyai multikolinearitas dan variabel hubungan membentuk algoritma random forest [9] yang termasuk dalam metode *Bagging* ensemble learning [22]. Akurasi yang diberikan oleh ketiga algoritma klasifikasi tersebut akan dievaluasi dengan menggunakan pendekatan ensemble machine learning.

2. METODE PENELITIAN

Untuk mencapai tujuan klasifikasi, terdapat beberapa tahapan yang perlu dilakukan. Berikut adalah tahapan diagram penelitain ini:



Gambar 1 Tahapan Penelitian

2.1 Pengumpulan Data

Setelah melakukan observasi di beberapa posyandu, didapatkan 2.578 data dari 15 desa yang terkumpul di Puskesmas Tlogowungu Kabupaten Pati. Penelitian ini menggunakan data pertumbuhan balita di bulan Oktober 2023 dengan atribut Nomor Induk Kependudukan, Nama Anak, Jenis Kelamin, Tanggal Lahir, Berat Badan lahir, Tinggi Badan lahir, Nama Orang Tua, Provinsi, Kabupaten/Kota, Puskesmas, Desa/Kelurahan, Posyandu, RT, RW, Alamat, Usia Waktu Pengukuran, Umur, Ukur, Tanggal Saat Pengukuran, Berat Badan, Tinggi Badan, Lingkar Kepala, BB/U, TB/U, ZS BB/U, ZS TB/U, dan ZS BB/TB. Penggunaan data tersebut membantu penelitian ini dalam memastikan keandalan dan validitas hasil sesuai dengan tujuan penelitian.

2.2 Pre-Processing Data

Langkah pertama dan terpenting dalam menyiapkan data untuk pemodelan klasifikasi adalah pre-processing. Beberapa tahapan persiapan pada penelitian ini adalah:

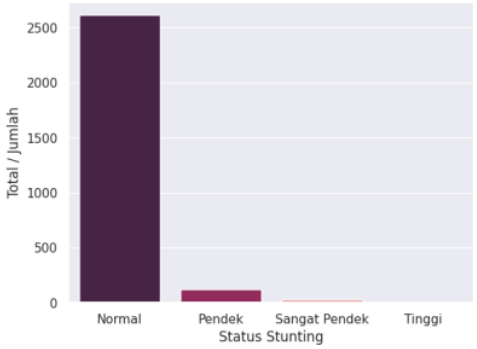
1). Seleksi Data
Pada tahapan ini dilakukan seleksi atau pengambilan atribut penting yang berkaitan dengan penggunaan pengolahan klasifikasi data stunting seperti: JK, Umur, BB Lahir, TB Lahir, Berat, Tinggi, ZS BB/U, ZS TB/U, ZS BB/TB dan Status Stunting sebagai variabel kelasnya.

2). Pembersihan Data
Tahap ini memastikan bahwa tidak ada data yang salah, kosong dan terduplikat, sehingga tahapan klasifikasi menjadi lebih akurat.

3). Transform Data
Pada tahapan ini dilakukan pengkonversian type data nominal ke data numeric, membagi dataset ke variabel X dan Y.

4). Balancing Data
Merupakan tahapan akhir dalam pre-processing yang sangat menentukan nilai pada saat masuk ke dalam tahapan pengujian. Pada Balacing Data, digunakan metode ensemble untuk menyeimbangkan data latih dan data uji menjadi 1000 data sampel pada kelas

klasifikasi yang memiliki multi kelas yaitu Tinggi, Normal, Pendek dan Sangat Pendek dengan teknik RUSBoostClassifier.



Gambar 3 Data Imbalance

Pada grafik di atas, terlihat total atau jumlah ketidakseimbangan data untuk variabel Status Stunting dan dijelaskan pada tabel 1 di bawah ini:

Tabel 1 Total Imbalance

Value	Total/Jumlah	Status Stunting
0	2613	Normal
1	118	Pendek
2	26	Sangat Pendek
3	1	Tinggi

2.3 Model Klasifikasi

Klasifikasi merupakan metode yang dilakukan dalam penelitian ini untuk melakukan proses klasifikasi data stunting. Dalam pemodelan klasifikasi digunakan tiga jenis model tunggal yaitu:

1). *Decision Tree*

Decision tree merupakan salah satu teknik machine learning yang sering digunakan dalam proses kategorisasi atau pengambilan keputusan berdasarkan sekumpulan aturan yang dibuat berdasarkan fitur (atribut) [23]. Formula persamaan yang digunakan dalam metode decision tree dapat dilihat dalam persamaan (1) dibawah ini:

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

Keterangan:
S = Himpunan
A = Atribut data
n = Jumlah partisi atribut dalam A.
$|S_i|$ = Jumlah kasus dalam partisi ke I.
$|S|$ = Jumlah kasus dalam S.

Sedangkan formula dasar *entropy*nya adalah:

$$Entropy(S) = \sum_{i=1}^{n} - pi * log_2 pi \quad (2)$$

Keterangan:
S = Merupakan suatu himpunan.
n = Merupakan jumlah partisi dalam S.
pi = Merupakan proporsi S_i terhadap S.

2). *Naïve Bayes*

Naïve Bayes bekerja dengan melihat frekuensi setiap set pelatihan yang dipilih untuk memutuskan mana yang menawarkan peluang terbaik untuk klasifikasi menggunakan teori probabilitas. Untuk menentukan parameter, metode Naïve Bayes hanya membutuhkan sedikit data latih selama proses klasifikasi. Pada proses Naïve Bayes digunakan formula persamaan (3) [24].

$$P\left(\frac{H}{X}\right) = \frac{P\left(\frac{X}{H}\right)P(H)}{P(X)} \quad (3)$$

Keterangan:

X = Data yang belum diketahui kelasnya.
H = Hipotesis data X merupakan suatu kelas yang spesifik.
P(H|X) = Probabilitas kondisi hipotesis H berdasarkan kondisi.
X P(H) = Probabilitas hipotesis H.
P(X|H) = Probabilitas X berdasarkan kondisi tersebut.
P(X) = Probabilitas dari X.

3). k-NN

Pada metode ini, menentukan nilai k yang tepat merupakan langkah penting dalam melakukan klasifikasi k-NN karena hal ini mempengaruhi sensitivitas sistem dalam bekerja. Formula statistik yang digunakan dalam Algoritma k-NN dapat dilihat pada persamaan (4) [25].

$$d_i = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \quad (4)$$

Keterangan:

d = Jarak.
x = Data latih.
y = Data uji.
n = Dimensi data.
i = Variabel data.

Algoritma k-NN nilai k dapat berupa konstanta yang ditentukan atau variabel acak, dimana metrik jarak yang digunakan dapat berupa jarak *Euclidean, Minkowski, Manhattan, Chebychev, Korelasi,* dan lain-lain [26].

2.4 *Ensemble Learning*

Pada tahapan ensemble learning dalam penelitian ini terdapat beberapa metode yang akan dilakukan, yaitu:

1). Bagging

Teknik Bagging sering juga disebut *bootstrap aggregating*. Bagging merupakan teknik pelatihan beberapa model menggunakan metode algoritma klasifikasi yang sama menggunakan data sampel yang berbeda dari kumpulan dataset aslinya [18,27,28,29].

2). Boosting

Pada metode Boosting, dalam melakukan peningkatan dilakukan pelatihan terhadap model-model dasar atau lemah secara berurutan secara berulang sampai mendapatkan model yang terbaik [27,28,29].

3). Stacking

Fungsi metode Stacking adalah untuk meningkatkan keakuratan algoritma klasifikasi dengan cara menggabungkan model-model klasifikasi yang berbeda untuk meningkatkan kinerja keseluruhan model klasifikasi yang digunakan [27,28,29].

2.5 *Pengujian*

Confusion matrix digunakan untuk pengujian akurasi dengan teknik data mining [19,30]. Skala pengukuran dalam pengujian di penelitian ini menggunakan accuracy, precision dan recall.

1). Accuracy

Merupakan ukuran kerja yang paling intuitif untuk mendapatkan prediksi yang konsisten dari rasio pengamatan di semua set data [19,30].

$$Accuracy = \frac{TP+TN}{TP + FP + TN + FN} \quad (5)$$

2). Precision

Merupakan acuan pada prediksi positif yang secara akurat berasal dari prediksi yang dihitung dari keseluruhan prediksi positif [19,30].

$$Precision = \frac{TP}{TN + FP} \quad (6)$$

3). Recall

Merupakan ukuran efektivitas atau kapasitas sistem untuk menemukan informasi dari sekumpulan data (dataset) [19,30].

$$Recall = \frac{TP}{TP + FP} \quad (7)$$

3. HASIL DAN PEMBAHASAN

Pemrograman python digunakan pada penelitian ini bersama dengan Google Colab yang teritegrasi dengan Google drive secara online. Tahapan dalam pemodelan klasifikasinya adalah sebagai berikut:

3.1 Metode Tunggal

Pada metode tunggal, dilakukan proses klasifikasi dengan metode Decision Tree, Naïve Bayes dan k-NN. Sebelum melakukan analisis dengan metode tunggal perlu dibuat code *estimator = [ ]* untuk mendeklarasikan variabel pada daftar model yang akan dianalisis atau diuji. Hasil pengukuran kinerja metode tunggal dari Accuracy, Precision serta Recall terlihat pada tabel 2 dibawah ini:

Tabel 2 Hasil Kinerja Metode Tunggal

Metode	Accuracy	Precision	Recall
Decision Tree	60,67%	59,24%	58,53%
Naïve Bayes	67,20%	64,38%	64,21%
k-NN	66,67%	65,49%	65,15%

Tabel 2 diatas menunjukkan bahwa dalam analisis metode tunggal Naïve Bayes mendapatkan nilai accuracy tertinggi sebesar 67,20% diikuti oleh metode k-NN sebesar 66,67% dan Decision Tree sebesar 60,67%.

3.2 Ensemble Learning

Ensemble learning berfungsi untuk meningkat kinerja klasifikasi model tunggal yang telah dilakukan. Dalam metode ensemble learning akan dilakukan beberapa model untuk menganalisis metode ensemble mana yang paling baik dalam klasifikasi data pertumbuhan balita.

1). Bagging

Pemodelan Bagging bekerja melakukan pengujian terhadap masing-masing metode tunggal dengan menggunakan data sampel yang berbeda dari dataset aslinya. Pada tabel 3 dibawah ini menunjukkan hasil pengukuran kinerja masing-masing metode tunggal dengan pemodelan Bagging.

Tabel 3 Hasil Kinerja Model Bagging

Metode	Accuracy	Precision	Recall
Model Bagging Decision Tree	78,93%	78%	77,99%
Model Bagging Naïve Bayes	69,60%	67,06%	66,96%
Model Bagging k-NN	72,93%	72,14%	69,65%

Hasil pengujian kinerja pada model Bagging nilai accuracy tertinggi sebesar 78,93% diperoleh oleh model Bagging Decision Tree diikuti model Bagging pada k-NN sebesar 72,93% dan model Bagging Naïve Bayes sebesar 69,60%.

2). Boosting

Boosting dilakukan untuk melakukan pelatihan berulang terhadap metode klasifikasi tunggal sampai mendapatkan model klasifikasi yang terbaik. Penelitian ini dilakukan tiga

macam model Boosting yaitu AdaBoost, Gradient Boosting dan XGBoost dengan hasil pengukuran kinerja sebagai berikut:

Tabel 4 Hasil Kinerja Model Boosting

Metode	Accuracy	Precision	Recall
AdaBoost	53,87%	52,53%	52,14%
Gradient Boosting	70,27%	67,80%	67,49%
XGBoost	67,87%	64,29%	64,15%

Gradient Boosting memperoleh nilai accuracy tertinggi sebesar 70,27% setelahnya XGBoost sebesar 67,87%. XGBoost merupakan pengembangan dari model Gradient Boosting yang bertujuan untuk mencegah overfitting dan optimalisasi sumber daya komputasi yang tersedia [28]. Accuracy model Adaboost merupakan nilai accuracy terendah dalam pemodelan Boosting sebesar 53,87%. Nilai accuracy tersebut dihasilkan dari meningkatkan model performa machine learning dengan menggabungkan model lemah menjadi model yang kuat [31].

3). Stacking

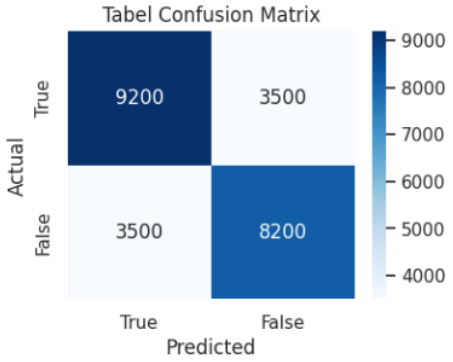
Model Stacking bekerja dalam 2 tahapan dengan mempelajari kumpulan data yang dianalisis. Tahapan pertama Stacking mengumpulkan hasil metode tunggal untuk membuat kumpulan data baru dan selanjutnya dilakukan proses meta-learning untuk memberikan hasil akhir [27]. Tabel 5 dibawah ini menunjukkan pengukuran kinerja dalam model Stacking:

Tabel 5 Hasil Kinerja Model Stacking

Metode	Accuracy	Precision	Recall
Stacking	70,93%	68,85%	68,94%

4). Pengujian

Dengan menggunakan sekumpulan data yang diketahui nilai sebenarnya, dilakukan pengujian confusion matrix dalam bentuk tabel matriks yang merinci kinerja model klasifikasi. Gambar 4 dibawah menujukkan empat kombinasi nilai prediksi dan nilai actual dalam penelitian:



Gambar 4 Tabel Confusion Matrix

Kesimpulan berikut ini adalah penjelasan dari tabel 4 diatas:

1). True Positif (TP) = 9200, artinya model klasifikasi yang dibangun dapat memprediksi dengan tepat dan data memang benar sebanyak 9200

2). True Negatif (TN) = 8200, artinya model klasifikasi yang dibangun dapat memprediksi dengan tepat dan data memang tidak salah sebanyak 8200

3). False Positif (FP) = 3500, artinya model klasifikasi yang dibangun dapat memprediksi dengan salah dan data memang benar sebanyak 3500
4). False Negatif (FN) = 3500, artinya model klasifikasi yang dibangun dapat memprediksi dengan salah dan data memang salah sebanyak 3500.

Hasil analisis penelitian ini memberikan suatu kontribusi pengetahuan pada analisis data, bahwa tidak semua metode ensemble learning dapat membantu meningkatkan akurasi model klasifikasi tunggal seperti yang dijelaskan dalam penelitian-penelitian sebelumnya. Penyebab perbedaan hasil akurasi dikarenakan perbedaan jumlah data (imbalance), atribut dan jumlah multiclass yang tidak sama. Permasalahan ini terbukti dengan tidak semua penggunaan teknik resampling sesuai diterapkan dalam penelitian ini.

4. KESIMPULAN DAN SARAN

Analisis penelitian ini memberikan kesimpulan bahwa tidak semua metode ensemble learning dapat membantu meningkatkan akurasi model klasifikasi tunggal. Teknik RUSBoostClassifier merupakan teknik yang sesuai untuk menangani data imbalance dalam penelitian dikarenakan rentang jarak data yang terlalu jauh. Diantara metode tunggal yang dipilih, metode Naïve Bayes memiliki nilai accuracy yang terbaik sebesar 67,21% dibandingkan metode Decision Tree dan k-NN. Sedangkan diantara metode ensemble, model Bagging Decision Tree memiliki nilai accuracy paling baik sebesar 78,93%, precision 78% dan recall sebesar 77,99%. Pengukuran kinerja klasifikasi ini mungkin dapat lebih optimal bila dilakukan penelitian pengaturan jumlah data sampel, jumlah data latih, jumlah data uji, pengaturan random state dan penggunaan data imbalance dengan rentang yang tidak terlalu jauh.

DAFTAR PUSTAKA

[1] A. Aditianti, I. Raswanti, S. Sudikno, D. Izwardy, and S. E. Irianto, "Prevalensi Dan Faktor Risiko Stunting Pada Balita 24-59 Bulan Di Indonesia: Analisis Data Riset Kesehatan Dasar 2018 [Prevalence and Stunting Risk Factors in Children 24-59 Months in Indonesia: Analysis of Basic Health Research Data 2018]," *Penelit. Gizi dan Makanan (The J. Nutr. Food Res.*, vol. 43, no. 2, pp. 51–64, 2021, doi: 10.22435/pgm.v43i2.3862.

[2] O. Saeful Bachri, R. Mohamad, and H. Bhakti, "PENENTUAN STATUS STUNTING PADA ANAK DENGAN MENGGUNAKAN ALGORITMA KNN Stunting Status Determination in Children using KNN ALgorithm," *J. Ilm. Intech Inf. Technol. J. UMUS*, vol. 3, no. 2, pp. 130–137, 2021.

[3] M. R. Nugroho, R. N. Sasongko, and M. Kristiawan, "Faktor-faktor yang Mempengaruhi Kejadian Stunting pada Anak Usia Dini di Indonesia," *J. Obs. J. Pendidik. Anak Usia Dini*, vol. 5, no. 2, pp. 2269–2276, 2021, doi: 10.31004/obsesi.v5i2.1169.

[4] Y. Permanasari *et al*., "Faktor Determinan Balita Stunting Pada Desa Lokus Dan Non Lokus Di 13 Kabupaten Lokus Stunting Di Indonesia Tahun 2019," *Penelit. Gizi dan Makanan (The J. Nutr. Food Res*., vol. 44, no. 2, pp. 79–92, 2021, doi: 10.22435/pgm.v44i2.5665.

[5] A. Sumiah and N. Mirantika, "Perbandingan Metode K-Nearest Neighbor dan Naive Bayes untuk Rekomendasi Penentuan Mahasiswa Penerima Beasiswa pada Universitas Kuningan," *Buffer Inform.*, vol. 6, no. 1, pp. 1–10, 2020.

[6] J. D. German, A. K. S. Ong, A. A. N. Perwira Redi, and K. P. E. Robas, "Predicting factors affecting the intention to use a 3PL during the COVID-19 pandemic: A machine learning ensemble approach," *Heliyon*, vol. 8, no. 11, p. e11382, 2022, doi: 10.1016/j.heliyon.2022.e11382.

[7] M. Bansal, Prince, R. Yadav, and P. K. Ujjwal, "Palmistry using Machine Learning and OpenCV," *Proc. 4th Int. Conf. Inven. Syst. Control. ICISC 2020*, no. Icisc, pp. 536–539, 2020, doi: 10.1109/ICISC47916.2020.9171158.

[8] M. Çakir, M. Yilmaz, M. A. Oral, H. Ö. Kazanci, and O. Oral, "Accuracy assessment of RFerns, NB, SVM, and kNN machine learning classifiers in aquaculture," *J. KING SAUD Univ. - Sci.*, p. 102754, 2023, doi: 10.1016/j.jksus.2023.102754.

[9] T. Srinath and G. H.S., "Explainable machine learning in identifying credit card defaulters," *Glob. Transitions Proc.*, vol. 3, no. 1, pp. 119–126, 2022, doi: 10.1016/j.gltp.2022.04.025.

[10] Elsevier B.V., "Application of Data Mining Techniques to Predict Breast Cancer," *Procedia Comput. Sci.*, vol. 163, pp. 11–18, 2019, doi: 10.1016/j.procs.2019.12.080.

[11] H. Ramadhan *et al.*, "Impression Determination of Batik Image Cloth By Multilabel Ensemble Classification Using Color Difference Histogram Feature Extraction," *J. Ilm. KURSOR*, vol. 7, no. 4, pp. 173–180, 2014.

[12] S. Bashir, U. Qamar, and F. H. Khan, "Heterogeneous classifiers fusion for dynamic breast cancer diagnosis using weighted vote based ensemble," *Qual. Quant.*, vol. 49, no. 5, pp. 2061–2076, 2015, doi: 10.1007/s11135-014-0090-z.

[13] Normah, B. Rifai, S. Vambudi, and R. Maulana, "Analisa Sentimen Perkembangan Vtuber Dengan Metode Support Vector Machine Berbasis SMOTE," *J. Tek. Komput. AMIK BSI*, vol. 8, no. 2, pp. 174–180, 2022, doi: 10.31294/jtk.v4i2.

[14] K. Nugroho, E. Tjahjaningsih, L. Liana, and R. Mohamad Herdian Bhakti, "Prediksi Ujaran Kebencian Berbasis Text Pada Sosial Media Menggunakan Metode Neural Network," *J. Ilm. Intech Inf. Technol. J. UMUS*, vol. 5, no. 1, pp. 60–68, 2023, doi: 10.46772/intech.v5i1.1063.

[15] Y. Pristyanto, "Penerapan Metode Ensemble Untuk Meningkatkan Kinerja Algoritme Klasifikasi Pada Imbalanced Dataset," *J. Teknoinfo*, vol. 13, no. 1, p. 11, 2019, doi: 10.33365/jti.v13i1.184.

[16] M. Bansal, A. Goyal, and A. Choudhary, "A comparative analysis of K-Nearest Neighbor, Genetic, Support Vector Machine, Decision Tree, and Long Short Term Memory algorithms in machine learning," *Decis. Anal. J.*, vol. 3, no. May, p. 100071, 2022, doi: 10.1016/j.dajour.2022.100071.

[17] A. Talukder and B. Ahammed, "Machine learning algorithms for predicting malnutrition among under-five children in Bangladesh," *Nutrition*, vol. 78, 2020, doi: 10.1016/j.nut.2020.110861.

[18] J. R. Khan, J. H. Tomal, and E. Raheem, "Model and variable selection using machine learning methods with applications to childhood stunting in Bangladesh," *Informatics Heal. Soc. Care*, vol. 46, no. 4, pp. 425–442, 2021, doi: 10.1080/17538157.2021.1904938.

[19] R. Setiawan and A. Triayudi, "Klasifikasi Status Gizi Balita Menggunakan Naïve Bayes dan K-Nearest Neighbor Berbasis Web," *J. Media Inform. Budidarma*, vol. 6, no. 2, p. 777, 2022, doi: 10.30865/mib.v6i2.3566.

[20] N. Zamri *et al.*, "River quality classification using different distances in k-nearest neighbors algorithm," *Procedia Comput. Sci.*, vol. 204, no. 2021, pp. 180–186, 2022, doi: 10.1016/j.procs.2022.08.022.

[21] S. D. Jadhav and H. P. Channe, "Comparative Study of K-NN, Naive Bayes and Decision Tree Classification Techniques," *Int. J. Sci. Res.*, vol. 5, no. 1, pp. 1842–1845, 2016, doi: 10.21275/v5i1.nov153131.

[22] R. I. Arumnisaa and A. W. Wijayanto, "SISTEMASI: Jurnal Sistem Informasi Perbandingan Metode Ensemble Learning: Random Forest, Support Vector Machine, AdaBoost pada Klasifikasi Indeks Pembangunan Manusia (IPM) Comparison of Ensemble Learning Method: Random Forest, Support Vector Machine, AdaB," *Januari*, vol. 12, no. 1, pp. 2540–9719, 2023, [Online]. Available: http://sistemasi.ftik.unisi.ac.id

[23] M. Yunus, M. K. Biddinika, and A. Fadlil, "Classification of Stunting in Children Using the C4.5 Algorithm," *J. Online Inform.*, vol. 8, no. 1, pp. 99–106, 2023, doi: 10.15575/join.v8i1.1062.

[24] S. K. P. Loka and A. Marsal, "Perbandingan Algoritma K-Nearest Neighbor dan Naïve Bayes Classifier untuk Klasifikasi Status Gizi Pada Balita di Kota Solok: Comparison Algorithm of K-Nearest …," *... Indones. J. Mach. Learn. ...*, vol. 3, no. April, pp. 8–14, 2023, [Online]. Available: https://journal.irpi.or.id/index.php/malcom/article/view/474

[25] P. Purwanto and D. Syarif Sihabudin Sahid, "Using KNN Algorithms for Determining the Recipient of Smart Indonesia Scholarship Program," *J. Komput. Terap.*, vol. 7, no. Vol. 7 No. 2 (2021), pp. 163–173, 2021, doi: 10.35143/jkt.v7i2.4962.

[26] N. Hidayati and A. Hermawan, "K-Nearest Neighbor (K-NN) algorithm with Euclidean and Manhattan in classification of student graduation," *J. Eng. Appl. Technol.*, vol. 2, no. 2, pp. 86–91, 2021, doi: 10.21831/jeatech.v2i2.42777.

[27] M. Graczyk, T. Lasota, B. Trawiński, and K. Trawiński, "Comparison of Bagging, Boosting and Stacking," *Asian Conf. Intell. Inf. Database Syst.*, pp. 340–350, 2010.

[28] M. H. D. M. Ribeiro and L. dos Santos Coelho, "Ensemble approach based on Bagging, Boosting and Stacking for short-term prediction in agribusiness time series," *Appl. Soft Comput. J.*, vol. 86, p. 105837, 2020, doi: 10.1016/j.asoc.2019.105837.

[29] R. O. Odegua, "An Empirical Study of Ensemble Techniques (Bagging, Boosting and Stacking) Rising Odegua Nossa Data An Empirical Study of Ensemble Techniques (Bagging, Boosting and Stacking)," *Proc. Conf. Deep Learn*, no. March 2019, 2019, [Online]. Available:

https://www.researchgate.net/publication/338681864

[30] M. Riyyan and H. Firdaus, "PERBANDINGAN ALGORITME NAÃVE BAYES DAN KNN TERHADAP DATA PENERIMAAN BEASISWA (Studi Kasus Lembaga Beasiswa Baznas Jabar)," *J. Inform. dan Rekayasa Elektron.*, vol. 5, no. 1, pp. 1–10, 2022, doi: 10.36595/jire.v5i1.547.

[31] K. Nugroho, T. D. Wismarini, and H. Murti, "Sales Conversion Optimization Analysis Using the Random Forest Method," *Sinkron*, vol. 8, no. 4, pp. 2699–2705, 2023, doi: 10.33395/sinkron.v8i4.12943.

Naskah Publikasi-Eko Prasetyo-22.01.85.0007-22022024-TH

ORIGINALITY REPORT

11%	11%	2%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	doaj.org Internet Source	8%
2	publikasi.dinus.ac.id Internet Source	2%
3	Rhini Fatmasari, Alda Zevana Putri Widodo, Valianda Farradillah Hakim, Windu Gata, Dedi Dwi Saputra. "Pengkategorian Komentar Instagram Terhadap Layanan Akademik dan Non-Akademik Universitas Terbuka", Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi), 2023 Publication	2%

Exclude quotes	On	Exclude matches	< 2%
Exclude bibliography	On		