

Adit Yoga Sisfokom

by Yoga Pristyanto

Submission date: 15-May-2022 11:32PM (UTC+0700)

Submission ID: 1836721353

File name: Adit_Yoga_Sisfokom.docx (405.01K)

Word count: 2829

Character count: 17870

Model Balanced Bagging Berbasis Decision Tree Classification Pada Dataset Imbalanced Class

8 Aditya Ahmad Zein^[1], Yoga Pristyanto^{[2]*}
Program Studi Informatika, Fakultas Ilmu Komputer^[1]
Program Studi Sistem Informasi, Fakultas Ilmu Komputer^[2]
Universitas Amikom Yogyakarta
Yogyakarta, Indonesia

aditya.z@students.amikom.ac.id^[1], yoga.pristyanto@amikom.ac.id^[2]

14
Abstract— A classification algorithm is an algorithm that is very often used in conjunction with human needs, but previous research often encountered obstacles when using classification algorithms. One of the problems that are often encountered is the case of imbalanced datasets. So in this study, an ensemble method is proposed to overcome this. One of the well-known ensemble method algorithm is bagging. The balanced-bagging implementation is used to improve the capabilities of the bagging algorithm. This research involves comparing three different classification models with five datasets that have different imbalance ratios (IR). The model will be evaluated based on balanced accuracy, geometric mean and area under curve (AUC). The first model is a classification process using Decision Tree (without Bagging), which achieves its best performance on the pen_digits dataset with accuracy (97.06%) and g-mean (97.02%). Then the second model is a classification process using a Decision Tree (with Bagging) with the same dataset and an increase in the evaluation results, namely with the results of accuracy (97.86%) and g-mean (97.84%). Then these results are compared with other methods, namely classification using Decision Tree gets an evaluation value with accuracy (96.96%) and g-mean (96.92%). However, in some datasets with large instance values, the second model experienced a decrease in performance, so an increase was made. The third model is a solution to overcome the shortcomings of the second model. The model uses a Decision Tree (with Balanced-Bagging) where there is an increase in the results of the evaluation with the same dataset, namely the results of accuracy (98.56%) and g-mean (98.56%) and this occurs in each dataset.

Keywords— Bagging, Balanced-Bagging, Imbalanced Class, Decision Tree, Classification.

Abstrak— Algoritma klasifikasi merupakan algoritma yang sangat sering digunakan beriringan dengan kebutuhan manusia, namun peneliti sebelumnya sering dijumpai kendala saat menggunakan algoritma klasifikasi. Salah satu permasalahan yang sering sekali dijumpai ialah kasus imbalanced dataset. Sehingga dalam penelitian ini diusulkan ensemble method untuk mengatasinya, salah satu algoritma ensemble method yang terkenal ialah bagging. Implementasi balanced-bagging digunakan untuk meningkatkan kemampuan dari algoritma bagging. Dalam penelitian ini melibatkan perbandingan tiga model klasifikasi berbeda dengan lima dataset yang memiliki imbalanced ratio (IR) yang berbeda. Model akan dievaluasi berdasarkan metrik akurasi (balanced accuracy), geometric

mean dan area under curve (AUC). Model pertama merupakan proses klasifikasi menggunakan Decision Tree (tanpa Bagging) dimana mencapai performa terbaiknya pada dataset pen_digits dengan hasil akurasi (97,06%) dan g-mean (97,02%). Kemudian pada Model kedua merupakan proses klasifikasi menggunakan Decision Tree (dengan Bagging) dengan dataset yang sama terjadi peningkatan hasil evaluasi yaitu dengan hasil akurasi (97,86%) dan g-mean (97,84%) kemudian hasil ini dibandingkan dengan metode lain yaitu klasifikasi menggunakan Decision Tree mendapatkan nilai evaluasi dengan hasil akurasi (96,96%) dan g-mean (96,92%). Namun pada beberapa dataset dengan nilai instance yang besar model kedua mengalami penurunan kinerja sehingga dilakukan peningkatan. Model ketiga merupakan solusi untuk menangani kekurangan model kedua. Pada model ini menggunakan Decision Tree (dengan Balanced-Bagging) dimana terjadi peningkatan yang hasil evaluasi dengan dataset yang sama yaitu hasil akurasi sebesar (98,56%) dan g-mean (98,56%) dan hal ini terjadi pada setiap dataset.

Kata Kunci— Bagging, Balanced-Bagging, Ketidakseimbangan Kelas, Decision Tree, Klasifikasi.

12 I. PENDAHULUAN

Klasifikasi merupakan salah satu metode yang sering digunakan dalam pembelajaran mesin. Berbagai algoritma pembelajaran klasifikasi, seperti Decision Tree, Bayesian Classification, Logistic Regression dll. telah banyak membantu dalam penanganan kasus di dunia nyata. Namun, distribusi kelas yang tidak seimbang dari kumpulan data telah memberikan kesulitan yang serius bagi peneliti, dalam penggunaan algoritma pembelajaran pengklasifikasi yang diasumsikan memiliki distribusi kelas yang relatif seimbang [1]. Pasalnya distribusi kelas yang tidak seimbang akan mengurangi nilai dari hasil evaluasi yang dilakukan saat penelitian, sehingga mengakibatkan kinerja algoritma klasifikasi tidak maksimal [2]. Distribusi kelas yang tidak seimbang pada sebuah dataset terjadi ketika satu kelas dianggap lebih menarik dari kelas yang lain oleh algoritma, kasus ini sering terjadi pada kelas mayoritas, sehingga kelas positif atau minoritas tidak cukup terwakili. Dalam istilah sederhananya, jumlah contoh dari kelas positif (minoritas) jauh lebih kecil daripada jumlah contoh kelas negatif (mayoritas) [3]. Ketika suatu data menunjukkan jumlah nilai yang sedikit mereka akan diabaikan atau dianggap sebagai

Sebagai contoh dalam dunia medis, pasien yang dideteksi menderita kanker ganas (kelas minoritas) lebih sedikit daripada pasien menderita kanker jinak, hal ini berpotensi mendapatkan hasil klasifikasi yang keliru [1].

Mengingat pentingnya permasalahan kelas *imbalance*, telah dilakukan berbagai macam percobaan atau teknik untuk mengatasinya. Teknik tersebut dapat dikategorikan ke dalam beberapa pendekatan, bergantung bagaimana cara mereka mengatasi masalah distribusi kelas tidak seimbang. Pendekatan pada level algoritma dengan membuat atau memodifikasi sebuah algoritma, untuk memperhitungkan signifikansi dari kelas positif. Kemudian yang kedua pendekatan pada level dataset, dimana dilakukan modifikasi pada dataset agar mampu meningkatkan akurasi saat proses klasifikasi dilakukan [4]. Salah satu pendekatan pada level algoritma dalam meningkatkan akurasi kelas *imbalance* adalah dengan menggunakan metode *ensemble* [4]. Metode *ensemble* pada prinsipnya mengkombinasikan sekumpulan classifier yang dilatih dengan tujuan untuk membuat model klasifikasi (*classifier*) campuran yang terimprovisasi sehingga membuat *classifier ensemble* yang terbentuk lebih akurat dari pada *classifier* asalnya dalam melakukan suatu klasifikasi. *Bagging* dan *Boosting* adalah algoritma pada metode *ensemble* yang paling sering digunakan. *Bagging* adalah salah satu metode *ensemble* yang berbasis variasi data *ensemble*, yang terdiri dari memanipulasi data training sedemikian rupa sehingga masing-masing *classifier* dilatih dengan data training yang berbeda [5]. Namun menurut [6] bagging mengalami penurunan saat dihadapkan dengan dataset yang lebih stabil atau tingkat noisenya kecil. Sehingga pada penelitian ini juga akan dilakukan peningkatan terhadap metode bagging dengan memanfaatkan kemampuan dari metode undersampling (*Balanced-Bagging*).

Seperti penelitian yang dilakukan oleh [7], [8], [9], dan [10]. Mereka menunjukkan keberhasilan dari hasil kinerja algoritma bagging saat diplikasikan pada berbagai jenis proses terutama pada saat diaplikasikan pada algoritma klasifikasi. Kemudian pada penelitian yang dilakukan oleh [11] dan [12]. Dimana dalam penelitian [11] menggunakan metode *roughly balanced bagging* yang lebih complex dengan membagi dataset menjadi dataset negative dan dataset positif, kemudian pada penelitian [12] menggunakan *actively balanced bagging* yang lebih berpusat pada pengaturan weightnya. Pada penelitian mereka menunjukkan peningkatan yang terjadi pada hasil evaluasi metode bagging setelah diaplikasikan dengan metode undersampling.

Berdasarkan beberapa permasalahan pada penelitian sebelumnya. Solusi yang diusulkan pada penelitian ini ialah menggunakan Bagging dan *Balanced Bagging* berbasis Decision Tree. Hal ini dilakukan agar dataset yang digunakan tetap natural tanpa proses resampling.

II. METODE

A. Dataset

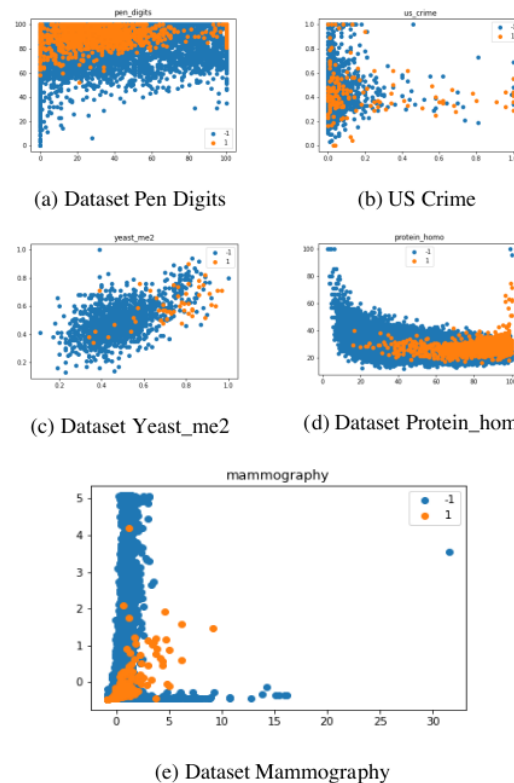
Pada penelitian digunakan data yang merupakan jenis data sekunder yang sudah tersedia di internet dan legal digunakan untuk umum bersumber dari repository KEEL dan UCI. Tabel 1 merupakan keterangan dataset yang digunakan pada

penelitian ini.

TABLE I. KETERANGAN DATASET

Name	Imbalanced Ratio	Instances	Features
pen_digits	9:4:1	10992	16
us_crime	12:1	1994	100
yeast_me2	28:1	1484	8
protein_homo	111:1	145751	74
mammography	42:1	11183	6

Kelima dataset tersebut memiliki kondisi ketidakseimbangan kelas dengan rasio yang berbeda. Berikut gambar 1 merupakan gambaran distribusi kelas beserta persebarannya.

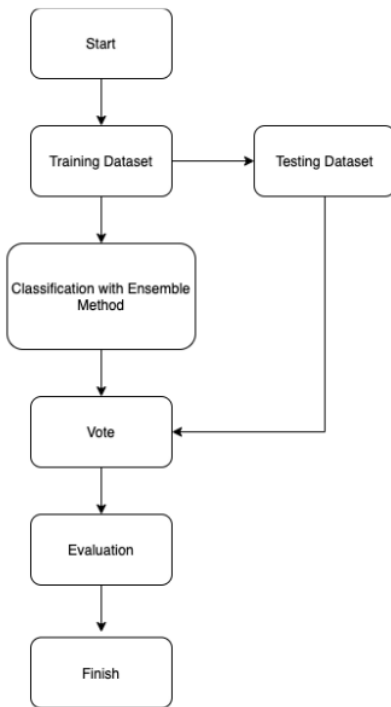


Gambar 1 Distribusi dan Persebaran Kelas Pada Dataset

Berdasarkan gambar 1 dapat dilihat bahwa salah satu kelas cenderung dominan dibandingkan kelas yang lain pada tiap dataset yang digunakan. Oleh karena itu diperlukan sebuah model sebagai solusi ketidakseimbangan kelas pada tiap dataset tersebut.

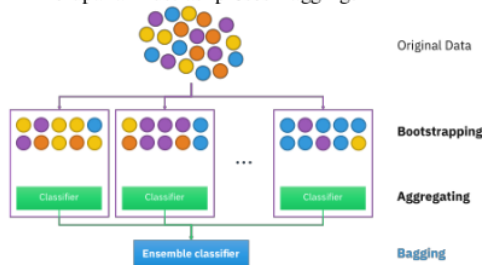
13
B. Alur Penelitian

11 Penelitian ini terdiri dari beberapa langkah yang dilakukan. Gambar 2 merupakan tahapan penelitian yang dilakukan pada penelitian ini.



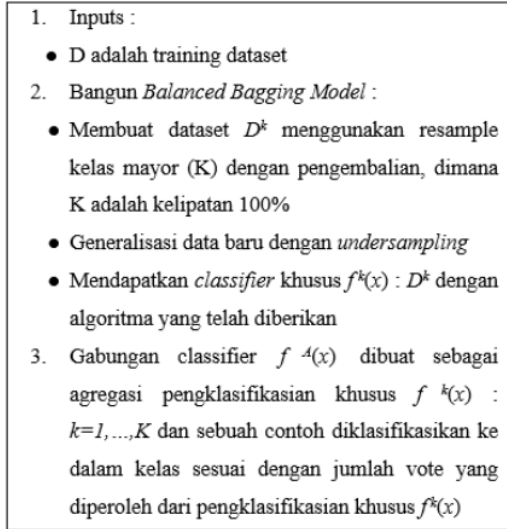
Gambar 2 Alur Penelitian

1. Bagging (Bootstrap Aggregating)
Metode ini diperkenalkan oleh [5] dan menjadi singkatan dari kata “bootstrap aggregating”. Bagging merupakan metode yang memungkinkan pengurangan nilai varians secara signifikan [13]. Teknik ini juga dapat diandalkan saat bertemu kasus overfitting pada teknik pengklasifikasi tunggal [14]. Gambar 3 berikut ini merupakan ilustrasi proses Bagging.



Gambar 3 Ilustrasi Proses Bagging

2. Balanced-Bagging (Undersampling Bagging)
Metode Balanced-Bagging atau under-bagging merupakan metode dari hasil kombinasi metode undersampling dan metode bagging. Dimana kemampuan dari metode undersampling akan dimanfaatkan saat proses bagging sehingga memungkinkan bagi bagging meningkatkan hasil balanced accuracy, geometric mean dan AUC [11][12]. Berikut gambar 4 merupakan pseudocode algoritme balanced bagging.



Gambar 5 Pseudocode Algoritme Balanced Bagging

3. Evaluasi Model
Metode Evaluasi yang digunakan pada penelitian ini adalah confusion matrix. Dengan beberapa indikator untuk menghitungnya yaitu balanced accuracy, geometric mean dan AUC. Balanced Accuracy adalah metrik yang lebih baik dari Accuracy biasa untuk digunakan dengan data yang tidak seimbang. Ini menyumbang kelas hasil positif dan negatif dan tidak mislead dengan data yang tidak seimbang [19].

$$\text{Balanced Accuracy} = \frac{\text{Sensitivity} + \text{Specificity}}{2} \quad (1)$$

Geometric Mean (G-Mean) disarankan dalam penelitian sebagai produk dari akurasi prediksi untuk kedua kelas yaitu sensitivitas: akurasi pada sampel positif, dan Spesifisitas: akurasi pada sampel negative [20].

$$G - \text{Mean} = \sqrt{\text{Sensitivity} \times \text{Specificity}} \quad (2)$$

AUC (Area Under Curve) digunakan sebagai metode evaluasi untuk membuktikan kemampuan model dalam membedakan antar kelas. Semakin tinggi AUC, semakin baik model dalam memprediksi 0 sebagai 0

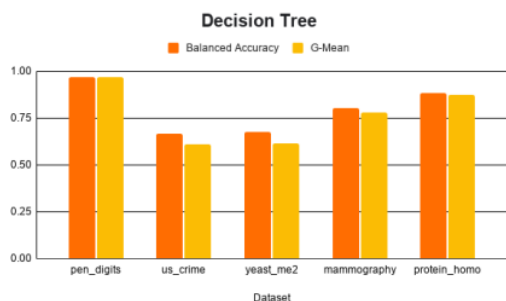
dan 1 sebagai I [21].

III. HASIL DAN PEMBAHASAN

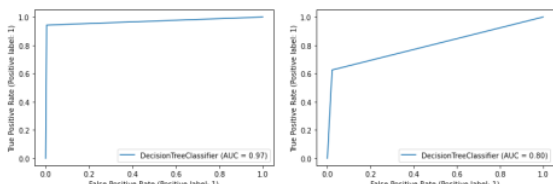
Pada bagian ini difokuskan pada hasil dan pembahasan dari tiga model klasifikasi yang dibandingkan yaitu Decision Tree, Decision Tree+Bagging dan Decision Tree+Balanced Bagging.

A. Decision Tree

Model pertama pada penelitian ini ialah melakukan klasifikasi (Decision Tree) pada dataset tanpa melalui bootstrap aggregating, dimana telah didapatkan hasil yang sudah baik namun hasil ini masih sering berubah-ubah sesuai dengan iterasi yang dilakukan sehingga perlu ditingkatkan hasilnya untuk mendapatkan hasil yang tetap dan baik outputnya. Gambar 6 menunjukkan hasil nilai balanced akurasi dan g-mean algoritme Decision Tree pada setiap dataset yang digunakan. Sedangkan Gambar 7 dan Tabel 2 menunjukkan nilai AUC dari algoritme Decision Tree pada setiap dataset yang digunakan.

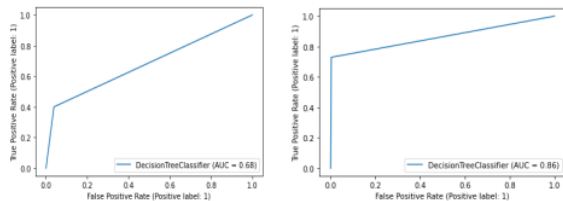


Gambar 6 Grafik Nilai Balanced Akurasi dan G-Mean Algoritma Decision Tree Pada Setiap Dataset



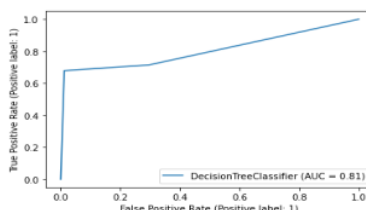
(a) AUC Pen Digits

(b) AUC US_Crime



(c) Dataset Yeast_me2

(d) AUC Protein_homo



(e) AUC Mammography

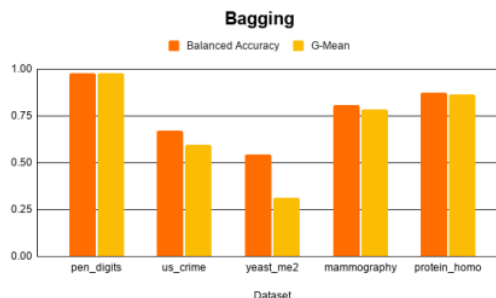
Gambar 7 Grafik Nilai AUC Algoritma Decision Tree Pada Setiap Dataset

TABLE II. NILAI AUC DECISION TREE PADA SETIAP DATASET

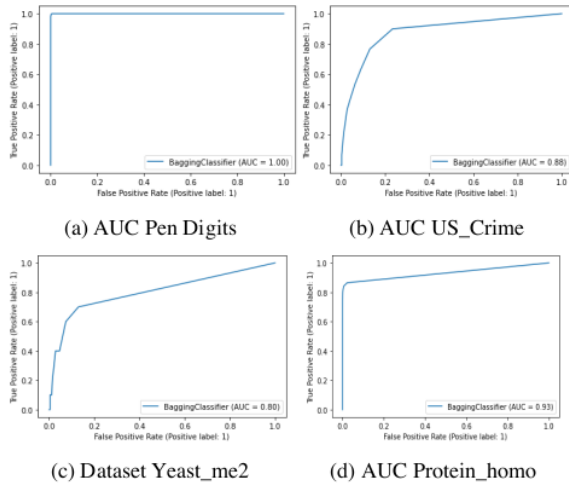
Datasets				
pen_digits	us_crime	yeast_me2	mammography	protein_homo
0.80	0.97	0.68	0.81	0.86

B. Decision Tree + Bagging

Pada model kedua ini akan dilakukan klasifikasi menggunakan Decision Tree pada dataset yang sudah melalui proses bootstrap aggregating. Gambar 8 menunjukkan hasil nilai balanced akurasi dan g-mean algoritme Decision Tree+Bagging pada setiap dataset yang digunakan. Sedangkan Gambar 9 dan Tabel 3 menunjukkan nilai AUC dari algoritme Decision Tree+Bagging pada setiap dataset yang digunakan.



Gambar 8 Grafik Nilai Balanced Akurasi dan G-Mean Algoritma Decision Tree+Bagging Pada Setiap Dataset

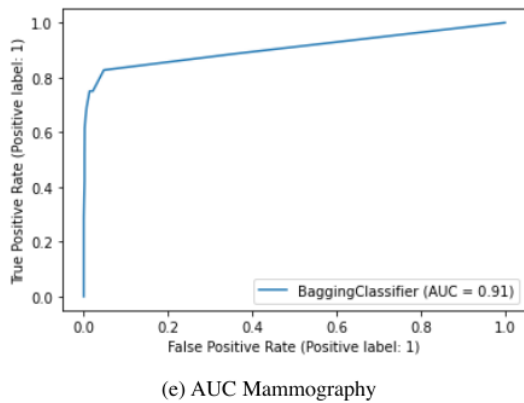


(a) AUC Pen Digits

(b) AUC US_Crime

(c) Dataset Yeast_me2

(d) AUC Protein_homo



(e) AUC Mammography

Gambar 9 Grafik Nilai AUC Algoritma Decision Tree+Bagging Pada Setiap Dataset

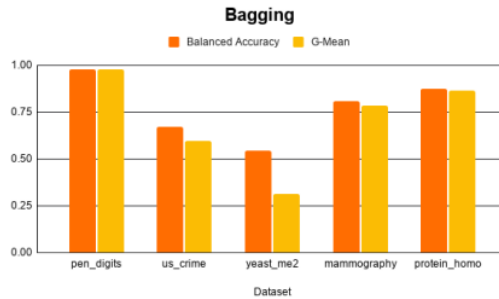
TABLE III. NILAI AUC DECISION TREE+BAGGING PADA SETIAP DATASET

Datasets				
pen_digits	us_crime	yeast_me2	mammography	protein_homo
1.00	0.88	0.80	0.91	0.93

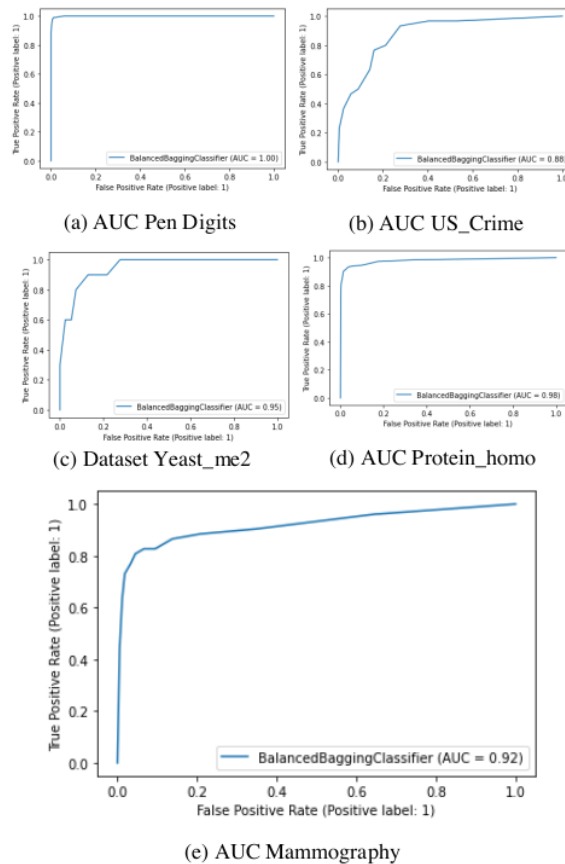
C. Decision Tree + Balanced Bagging

Pada model ini kalsifikasi yang dilakukan ialah hasil proses klasifikasi dimana pada saat bootstrap ag...
 terjadi dikombinasikan dengan undersampling sehingga didapatkan hasil yang lebih baik dari sebelumnya, dimana saat bertemu dengan dataset yang memiliki jumlah noise yang sedikit hasil klasifikasi tetap mencapai pada hasil yang baik. Gambar 10 menunjukkan hasil nilai balanced akurasi dan g-mean algoritme Decision Tree+Balanced Bagging pada setiap dataset yang digunakan. Sedangkan Gambar 11 dan Tabel 4

menunjukkan nilai AUC dari algoritme Decision Tree+Balanced Bagging pada setiap dataset yang digunakan.



Gambar 10 Grafik Nilai Balanced Akurasi dan G-Mean Algoritma Decision Tree+Balanced Bagging Pada Setiap Dataset



(a) AUC Pen Digits

(b) AUC US_Crime

(c) Dataset Yeast_me2

(d) AUC Protein_homo

(e) AUC Mammography

Gambar 9 Grafik Nilai AUC Algoritma Decision Tree+Balanced Bagging Pada Setiap Dataset

TABLE IV. NILAI AUC DECISION TREE+BALANCED BAGGING PADA SETIAP DATASET

Datasets				
pen_digits	us_crime	yeast_me2	mammography	protein_homo
1.00	0.88	0.95	0.92	0.98

D. Perbandingan Kinerja Model Klasifikasi

Perbandingan model dilakukan untuk mengetahui kinerja yang dihasilkan dari ketiga model yang telah diuji. Tabel 5 berikut ini menunjukkan perbandingan keseluruhan dari ketiga model dari sisi balanced akurasi, g-mean, dan AUC.

TABLE V. PERBANDINGAN KINERJA MODEL KLASIFIKASI

Model	Evaluasi	Datasets				
		Pen digits	Us crime	Yeast me2	Mammo graphy	Protei homo
Decision Tree	Balanced Akurasi	0.9674	0.6687	0.6738	0.8033	0.8825
	G-Mean	0.9670	0.6124	0.6156	0.7810	0.8751
	AUC	0.80	0.97	0.68	0.81	0.86
Decision Tree + Bagging	Balanced Akurasi	0.9787	0.6697	0.5448	0.8065	0.8744
	G-Mean	0.9784	0.5972	0.3146	0.7836	0.8654
	AUC	1.00	0.88	0.80	0.91	0.93
Decision Tree + Balanced Bagging	Balanced Akurasi	0.9856	0.7460	0.8853	0.8814	0.9456
	G-Mean	0.9856	0.7375	0.8852	0.8783	0.9451
	AUC	1.00	0.88	0.95	0.92	0.98

Hasil perbandingan kinerja pada tabel 5 menunjukkan terjadinya peningkatan pada setiap dataset baik dari sisi Balanced Akurasi, G-Mean maupun AUC. Hal ini menunjukkan bahwa implementasi metode bagging dapat meningkatkan kemampuan dari model tersebut khususnya pada dataset dalam kondisi imbalanced class.

IV. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan dapat disimpulkan menjadi tiga poin berikut:

1. Implementasi metode bagging dan balanced bagging terhadap algoritma klasifikasi Decision Tree mampu meningkatkan kinerja hasil akurasi (balanced accuracy), geometric mean, dan AUC. Namun metode bagging maupun balanced bagging memiliki kendala dalam mengatasi dataset dengan jumlah instance yang besar, seperti pada kasus dataset protein_homo yang memiliki jumlah instance yang besar dapat menurunkan kinerja dari metode bagging dan balanced bagging yang memiliki dampak juga terhadap nilai dari hasil evaluasi.
2. Dalam hal permasalahan pada ketidakseimbangan kelas metode bagging dan balanced bagging dapat menjadi solusi. Hal ini dapat dilihat dari nilai balanced akurasi,

geometric mean dan AUC pada metode bagging dan balanced bagging lebih baik dibandingkan decision tree original.

3. Saran untuk penelitian selanjutnya ialah dengan menguji model bagging maupun balanced bagging pada dataset multiclass imbalanced dan diimplementasikan pada beberapa algoritma klasifikasi lainnya.

REFERENCES

- [1] Y. Sun, A. K. C. Wong, and M. S. Kamel, "Classification of imbalanced data: A review," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 23, no. 4, pp. 687-719, 2009, doi: 10.1142/S0218001409007326.
- [2] 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.
- [3] A. Ali, S. M. Shamsuddin, and A. L. Ralescu, "Classification with class imbalance problem: A review," *Int. J. Adv. Soft Comput. its Appl.*, vol. 7, no. 3, pp. 176-204, 2015.
- [4] S. S. Pangastuti, K. Fithriasari, N. Iriawan, and W. Suryaningtyas, "Data Mining Approach for Educational Decision Support," *EKSAKTA J. Sci. Data Anal.*, vol. 2, no. February, pp. 33-44, 2021, doi: 10.20885/eksakta.vol2.iss1.art5.
- [5] L. Breimann, "Bagging predictors," *Risks*, vol. 8, no. 3, pp. 1-26, 2020, doi: 10.3390/risks8030083.
- [6] J. R. Quinlan, "Bagging, Boosting, and C4.5," pp. 725-730, 2006.
- [7] E. Yaman and A. Subasi, "Comparison of Bagging and Boosting Ensemble Machine Learning Methods for Automated EMG Signal Classification," *Biomed Res. Int.*, vol. 2019, 2019, doi: 10.1155/2019/9152506.
- [8] M. E. Purbaya, A. F. Nugraha, S. Gustina, and M. K. Azis, "Meta-Algorithms untuk Meningkatkan Kinerja Klasifikasi dalam Keberhasilan Telemarketing Perbankan," *Techno.Com*, vol. 19, no. 4, pp. 385-396, 2020, doi: 10.33633/te.v19i4.3725.
- [9] Y. A. Jatmiko, S. Padmadisastra, and A. Chadidjah, "Analisis Perbandingan Kinerja Cart Konvensional, Bagging Dan Random Forest Pada Klasifikasi Objek: Hasil Dari Dua Simulasi," *Media Stat.*, vol. 12, no. 1, p. 1, 2019, doi: 10.14710/medstat.12.1.1-12.
- [10] C. Makris, G. Pispirigos, and I. O. Rizos, "A distributed bagging ensemble methodology for community prediction in social networks," *Inf.*, vol. 11, no. 4, 2020, doi: 10.3390/INFO11040199.
- [11] S. Hido, H. Kashima, and Y. Takahashi, "Roughly balanced Bagging for Imbalanced data," *Stat. Anal. Data Min.*, vol. 2, no. 5-6, pp. 412-426, 2009, doi: 10.1002/sam.10061.
- [12] J. Blaszczynski and J. Stefanowski, "Actively balanced bagging for imbalanced data," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10352 LNAI, pp. 271-281, 2017, doi: 10.1007/978-3-319-60438-1_27.
- [13] Sutriyono, "PENERAPAN BAGGING UNTUK PENYAKIT JANTUNG KORONER BERBASIS RANDOM FOREST," vol. 3, pp. 536-543, 2020.
- [14] A. Y. Triyanto and R. Kusumaningrum, "Implementasi Teknik Sampling untuk Mengatasi Imbalanced Data pada Penentuan Status Gizi Balita dengan Menggunakan Learning Vector Quantization," *J. IPTEKKOM J. Ilmu Pengetah. Teknol. Inf.*, vol. 19, no. 1, p. 39, 2017, doi: 10.33164/jiptekkom.19.1.2017.39-50.
- [15] I. D. Mienye, Y. Sun, and Z. Wang, "Prediction performance of improved decision tree-based algorithms: A review," *Procedia Manuf.*, vol. 35, pp. 698-703, 2019, doi: 10.1016/j.promfg.2019.06.011.
- [16] R. Garcia Leiva, A. F. Anta, V. Mancuso, and P. Casari, "A novel hyperparameter-free approach to decision tree construction that avoids overfitting by design," *arXiv*, vol. 7, 2019.
- [17] E. C. Abana, "A decision tree approach for predicting student grades in Research Project using Weka," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 7, pp. 285-289, 2019, doi: 10.14569/ijacsa.2019.0100739.
- [18] J. Han, M. Kamber, and J. Pei, *Data mining: Data mining concepts and techniques*. 2014.
- [19] H. Araujo, A. M. Mendonça, A. J. Pinho, and M. I. Torres, "Index

- of balanced accuracy: A performance measure for skewed class distributions," *Maik Nauk. Publ. / Springer SBM*, vol. 5524 LNCS, pp. 441–448, 2009, doi: 10.1007/978-3-642-02172-5_57.
- [20] D. Chakrabarty, "Arithmetic-Geometric Mean: Evaluation of Parameter from Observed Data Containing Itself and Random Error," *Int. J. Electron. Appl. Res.*, vol. 06, no. 02, pp. 98–111, 2019, doi: 10.33665/ijear.2019.v06i02.003.
- [21] D. Brzezinski and J. Stefanowski, "Prequential AUC: properties of the area under the ROC curve for data streams with concept drift," *Knowl. Inf. Syst.*, vol. 52, no. 2, pp. 531–562, 2017, doi: 10.1007/s10115-017-1022-8.

Adit Yoga Sisfokom

ORIGINALITY REPORT

20%

SIMILARITY INDEX

19%

INTERNET SOURCES

9%

PUBLICATIONS

%

STUDENT PAPERS

PRIMARY SOURCES

1	jurnal.atmaluhur.ac.id Internet Source	5%
2	repository.its.ac.id Internet Source	4%
3	www.coursehero.com Internet Source	3%
4	repository.unimus.ac.id Internet Source	2%
5	sistemasi.ftik.unisi.ac.id Internet Source	1%
6	prosiding.uhb.ac.id Internet Source	1%
7	Yoga Pristyanto, Anggit Ferdita Nugraha, Irfan Pratama, Akhmad Dahlan, Lucky Adhikrisna Wirasakti. "Dual Approach to Handling Imbalanced Class in Datasets Using Oversampling and Ensemble Learning Techniques", 2021 15th International Conference on Ubiquitous Information Management and Communication (IMCOM),	1%

2021

Publication

8	www.researchgate.net Internet Source	1 %
9	I Made Artha Agastya, Arief Setyanto, Kusrini, Dini Oktarina Dwi Handayani. "Convolutional Neural Network for Pornographic Images Classification", 2018 Fourth International Conference on Advances in Computing, Communication & Automation (ICACCA), 2018 Publication	1 %
10	obsesi.or.id Internet Source	<1 %
11	repository.uksw.edu Internet Source	<1 %
12	lib.unnes.ac.id Internet Source	<1 %
13	123dok.com Internet Source	<1 %
14	Delpiah Wahyuningsih, Ellya Helmud. "Penerapan Algoritma Genetika Untuk Optimasi Penjadwalan pada MTS Negeri 1 Pangkalpinang", Jurnal Sisfokom (Sistem Informasi dan Komputer), 2020 Publication	<1 %
15	docplayer.info Internet Source	

<1 %

16

ejournal.unib.ac.id

Internet Source

<1 %

17

ojs.stikeslandbouw.ac.id

Internet Source

<1 %

18

www.kci.go.kr

Internet Source

<1 %

Exclude quotes On

Exclude matches Off

Exclude bibliography On