Model Balanced Bagging Berbasis Decision Tree Pada Dataset Imbalanced Class

Yoga Pristyanto[1]*, Aditya Ahmad Zein[2],
Program Studi Sistem Informasi, Fakultas Ilmu Komputer [1]
Program Studi Informatika, Fakultas Ilmu Komputer [2]
Universitas Amikom Yogyakarta
Yogyakarta, Indonesia
yoga.pristyanto@amikom.ac.id [1], aditya.z@students.amikom.ac.id [2]

***Abstract*— The classification algorithm is an algorithm that is very often used in conjunction with human needs, but previous studies often encountered obstacles when using the classification algorithm. One problem that is often encountered is the case of an imbalanced dataset. So in this study, an ensemble method is proposed to overcome this; one of the well-known ensemble method algorithms is bagging. The implementation of balanced bagging is used to improve the capabilities of the bagging algorithm. This study compares three different classification models with five datasets with different imbalanced ratios (IR). The model will be evaluated based on the accuracy (balanced accuracy), geometric mean and area under the curve (AUC). The first model is a classification process using a Decision Tree (without bagging), the second model is a classification process using a Decision Tree (with bagging), and the third model uses a Decision Tree (with Balanced-Bagging). The implementation of the bagging and balanced bagging methods for the Decision Tree classification algorithm was able to improve the performance of the results of the accuracy (balanced accuracy), geometric mean, and AUC. The Decision Tree + Balanced Bagging model generally produces the best performance for all datasets used.**

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

***Abstrak*—Algoritma klasifikasi merupakan algoritma yang sangat sering digunakan beriringan dengan kebutuhan manusia, namun peneliti an 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), Model kedua merupakan proses klasifikasi menggunakan Decision Tree (dengan Bagging) dan model ketiga menggunakan Decision Tree (dengan Balanced-Bagging). Implementasi metode bagging dan balanced bagging terhadap algoritma klasifikasi Decision Tree mampu meningkatkan kinerja hasil akurasi (balanced accuracy), geometric mean, dan AUC. Secara umum model Decision Tree + Balanced Bagging menghasilkan kinerja yang terbaik pada seluruh dataset yang digunakan.**

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

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 membatu 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 *noise*. 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

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v10i3.1399, Copyright ©2023
Submitted : May 15, 2022, Revised : February 11, 2023, Accepted : February 13, 2023, Published : March 09, 2023

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 pengklasifikasian. *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 datset menjadi dataset negative dan datset 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 PENELITIAN

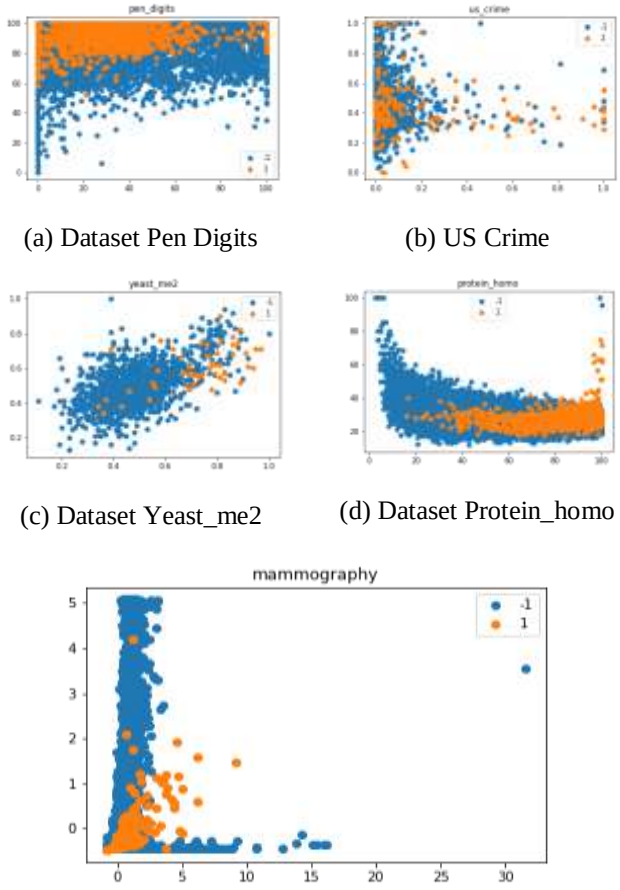
A. *Dataset*

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. Alasan utama menggunakan lima dataset tersebut karena kelimanya memiliki imabalanced ratio yang berbeda dan juga merupakan dataset yang sudah benchmark sehingga komprehensif dan general ketika digunakan untuk melakukan pengujian model klasifikasi. Tabel 1 merupakan keterangan dataset yang digunakan pada penelitian ini.

TABEL 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.



(a) Dataset Pen Digits (b) US Crime

(c) Dataset Yeast_me2 (d) Dataset Protein_homo

(e) Dataset Mammography

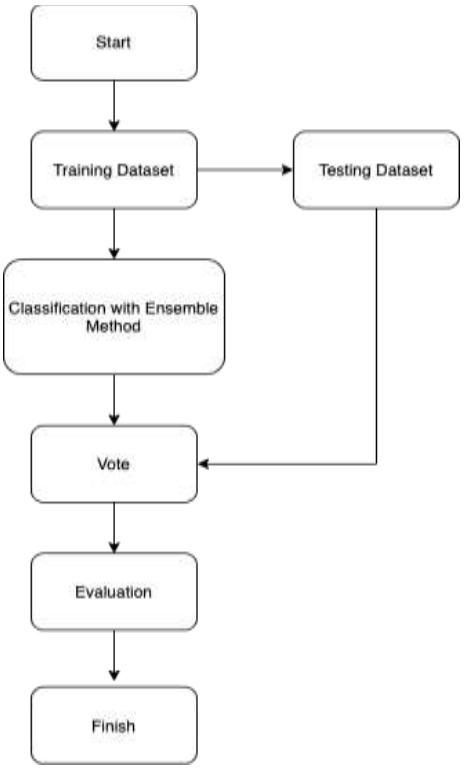
Gambar 1. Distribusi dan Persebaran Kelas Pada Dataset

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

B. *Alur Penelitian*

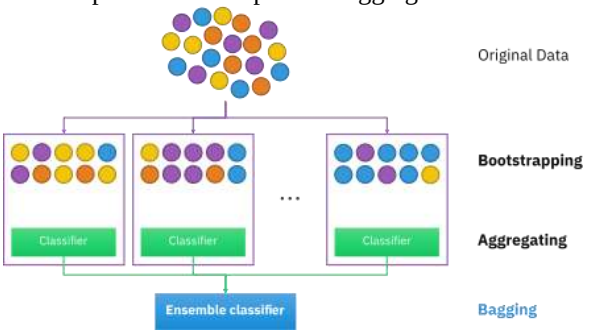
Penelitian ini terdiri dari beberapa langkah yang dilakukan. Gambar 2 merupakan tahapan penelitian yang dilakukan pada penelitian ini.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v10i3.1399, Copyright ©2023
Submitted : May 15, 2022, Revised : February 11, 2023, Accepted : February 13, 2023, Published : March 09, 2023



Gambar 2. Alur Penelitian

1. Pembagian Data
   Penelitian ini diawali dengan melakukan pembagian data menjadi dua bagian yaitu data training sebesar 80% dari keseluruhan data dan data testing sebesar 20% dari keseluruhan data.
2. Model Klasifikasi Ensemble
   Pada penelitian ini ada 3 skenario model yang digunakan yaitu Decision Tree, Decision Tree + Bagging dan Decision Tree + Balanced Bagging.
   a) 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

   b) 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.

1. Inputs :
   - D adalah training dataset
2. Bangun *Balanced Bagging Model* :
   - Membuat dataset D^k menggunakan resample kelas mayor (K) dengan pengembalian, dimana K adalah kelipatan 100%
   - Generalisasi data baru dengan *undersampling*
   - Mendapatkan *classifier* khusus $f^k(x)$: D^k dengan algoritma yang telah diberikan
3. Gabungan classifier $f^A(x)$ dibuat sebagai agregasi pengklasifikasian khusus $f^k(x)$: $k=1,\ldots,K$ dan sebuah contoh diklasifikasikan ke dalam kelas sesuai dengan jumlah vote yang diperoleh dari pengklasifikasian khusus $f^k(x)$

Gambar 5. Pseudocode Algoritme Balanced Bagging

3. Evaluasi Model
   Metode Evaluasi yang digunakan pada penelitian ini adalah confussion matrix. Dengan beberapa indicator 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].

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

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].

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

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 1 [21].

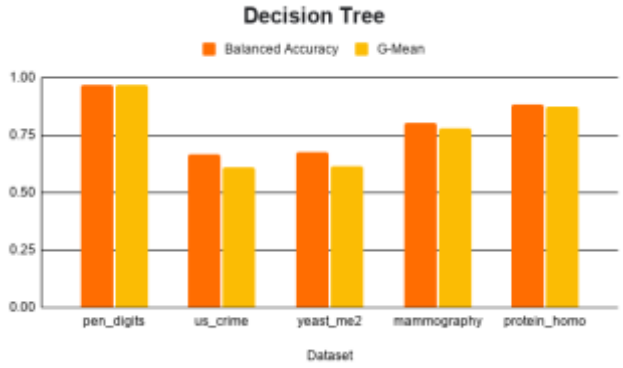
p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v10i3.1399, Copyright ©2023
Submitted : May 15, 2022, Revised : February 11, 2023, Accepted : February 13, 2023, Published : March 09, 2023

III. HASIL DAN PEMBAHASAN

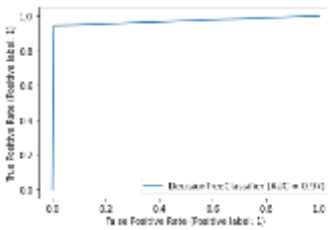
Pada bagian ini difokuskan pada hasil dan pembahasan dari tiga model klasifikasi yang dibandingkan yaitu Decision Tree, Decision Tree+Bagging dan Decision Tree+Balanced Bagging.

A. *Decision Tree*

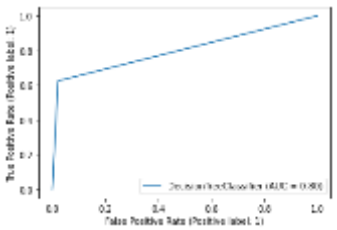
Model pertama pada penelitian ini ialah melakukan kalsifikasi (Decision Tree) pada dataset tanpa melalui bootstrap aggregating, dimana telah didapatkan hasil yang sudah baik namun hasil ini masih sering berubah-ubah sesuai dengan iterasi yang dilakukan sehingga perlu ditingkatkan hasilnya untuk mendapatkan 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.



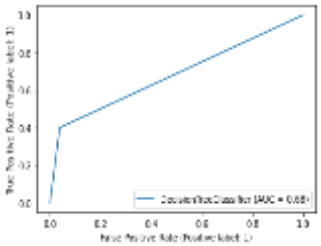
Gambar 6. Grafik Nilai Balanced Akurasi dan G-Mean Algoritma Decision Tree Pada Setiap Dataset



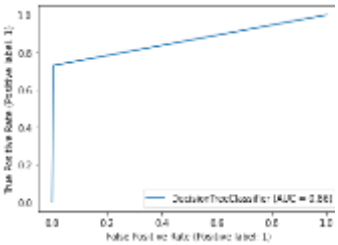
(a) AUC Pen Digits



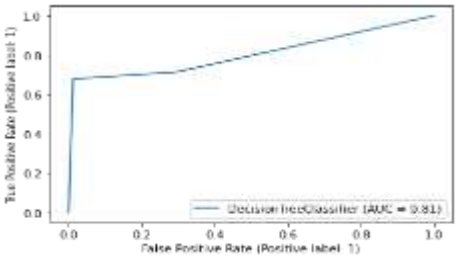
(b) AUC US_Crime



(c) Dataset Yeast_me2



(d) AUC Protein_homo



(e) AUC Mammography

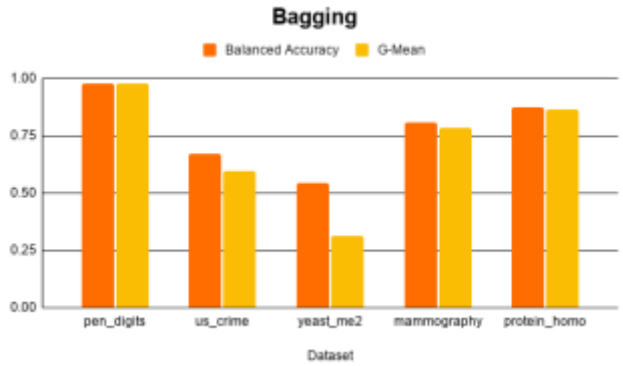
Gambar 7. Grafik Nilai AUC Algoritma Decision Tree Pada Setiap Dataset

TABEL II. NILAI AUC DECISION TREE PADA SETIAP DATASET

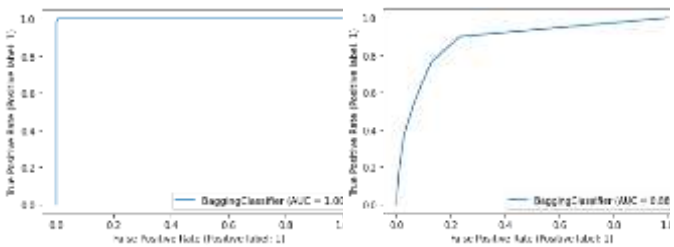
Datasets				
pen_digits	**us_crime**	**yeast_me2**	**mammography**	**protein_homo**
0.80	0.97	0.68	0.81	0.86

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.

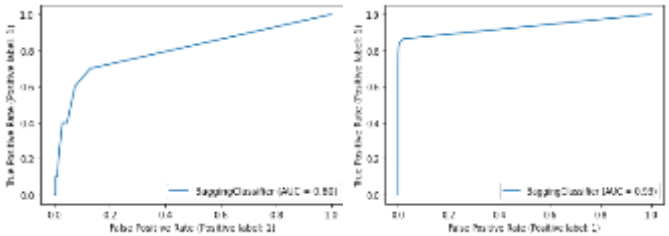


Gambar 8. Grafik Nilai Balanced Akurasi dan G-Mean Algoritma Decision Tree+Bagging Pada Setiap Dataset

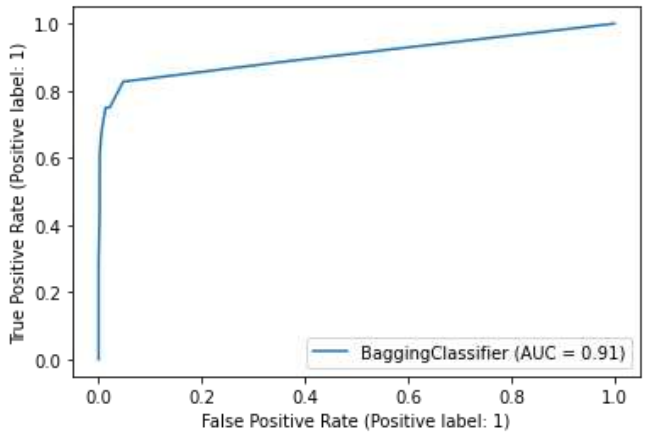


p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v10i3.1399, Copyright ©2023
Submitted : May 15, 2022, Revised : February 11, 2023, Accepted : February 13, 2023, Published : March 09, 2023

(a) AUC Pen Digits (b) AUC US_Crime



(c) Dataset Yeast_me2 (d) AUC Protein_homo



(e) AUC Mammography

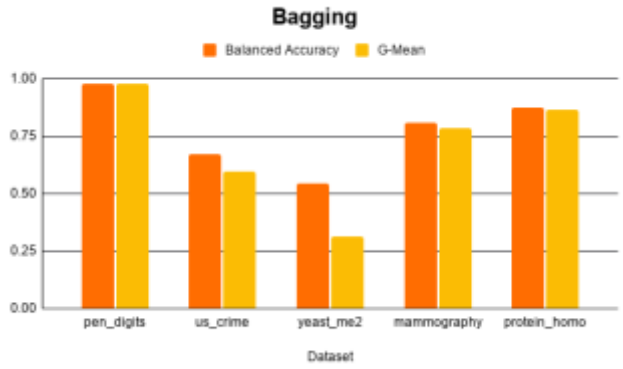
Gambar 9. Grafik Nilai AUC Algoritma Decision Tree+Bagging Pada Setiap Dataset

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

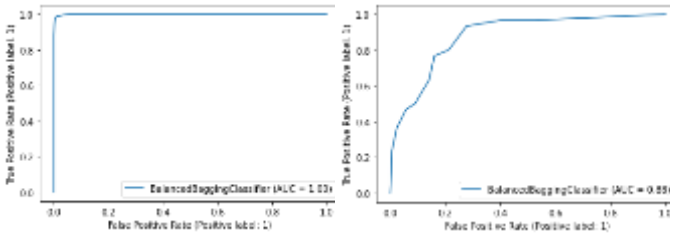
Datasets				
pen_digits	**us_crime**	**yeast_me2**	**mammography**	**protein_homo**
1.00	0.88	0.80	0.91	0.93

C. Decision Tree + Balanced Bagging

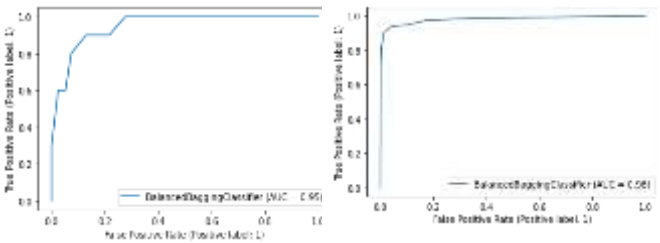
Pada model ini kalsifikasi yang dilakukan ialah hasil proses klasifikasi dimana pada saat bootstrap aggregating terjadi dikombinasikan dengan undersampling sehingga didapatkan hasil yang lebih baik dari sebelumnya, dimana saat bertemu dengan dataset yang memiliki jumlah noise yang sedikit hasil klasifikasi tetap mecapai 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.



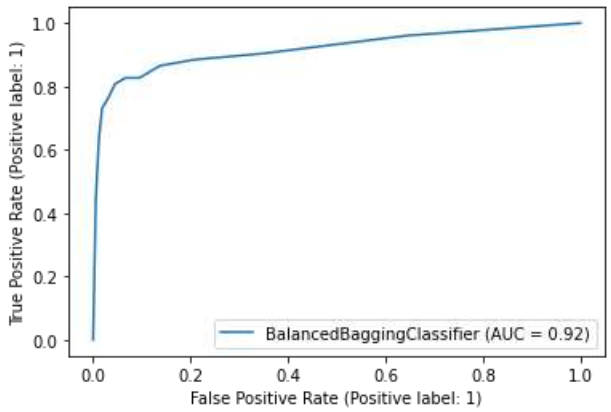
Gambar 10. Grafik Nilai Balanced Akurasi dan G-Mean Algoritma Decision Tree + Balanced Bagging Pada Setiap Dataset



(a) AUC Pen Digits (b) AUC US_Crime



(c) Dataset Yeast_me2 (d) AUC Protein_homo



(e) AUC Mammography

Gambar 11. Grafik Nilai AUC Algoritma Decision Tree+ Balanced Bagging Pada Setiap Dataset

TABEL IV. NILAI AUC DECISION TREE+BALANCED BAGGING PADA SETIAP

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v10i3.1399, Copyright ©2023
Submitted : May 15, 2022, Revised : February 11, 2023, Accepted : February 13, 2023, Published : March 09, 2023

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.

TABEL 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	**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. Secara umum model Decision Tree + Balanced Bagging menghasilkan kinerja yang terbaik pada seluruh dataset yang digunakan.

IV. KESIMPULAN

Berdasarakan 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. Secara umum model Decision Tree + Balanced Bagging menghasilkan kinerja yang terbaik pada seluruh dataset yang digunakan.
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] B. Kim and J. Kim, "Adjusting decision boundary for class imbalanced learning," IEEE Access, vol. 8, pp. 81674–81685, 2020, doi: 10.1109/ACCESS.2020.2991231.

[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] T. Alam, C. F. Ahmed, S. A. Zahin, M. A. H. Khan, and M. T. Islam, "An effective recursive technique for multi-class classification and regression for imbalanced data," IEEE Access, vol. 7, pp. 127615–127630, 2019, doi: 10.1109/ACCESS.2019.2939755.

[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/tc.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. Błaszczyński 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] Z. Yuan and P. Zhao, "An improved ensemble learning for imbalanced data classification," Proc. 2019 IEEE 8th Jt. Int. Inf. Technol. Artif. Intell. Conf. ITAIC 2019, no. Itaic, pp. 408–411, 2019, doi: 10.1109/ITAIC.2019.8785887.

[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. García 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] S. Huda et al., "An Ensemble Oversampling Model for Class Imbalance Problem in Software Defect Prediction," IEEE Access, vol. 6, pp. 24184–24195, 2018, doi: 10.1109/ACCESS.2018.2817572.

[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,"

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v10i3.1399, Copyright ©2023
Submitted : May 15, 2022, Revised : February 11, 2023, Accepted : February 13, 2023, Published : March 09, 2023

Knowl. Inf. Syst., vol. 52, no. 2, pp. 531–562, 2017, doi: 10.1007/s10115-017-1022-8.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v10i3.1399, Copyright ©2023
Submitted : May 15, 2022, Revised : February 11, 2023, Accepted : February 13, 2023, Published : March 09, 2023