

PERBANDINGAN PERFORMA BAGGING DAN ADABOOST UNTUK KLASIFIKASI DATA MULTI-CLASS

Oswaldo Vigo¹⁾, Samuel Lukas^{2*)}, Dion Krisnadi²⁾, Petrus Widjaja¹⁾

¹⁾Mathematics Department, Faculty of Science and Technology, Universitas Pelita Harapan

²⁾Informatics Department, Faculty of Computer Science, Universitas Pelita Harapan

E-mail: OF80010@student.uph.edu, {[samuel.lukas](mailto:samuel.lukas@uph.edu), [dion.krisnadi](mailto:dion.krisnadi@uph.edu), [petrus.widjaja](mailto:petrus.widjaja@uph.edu)}@uph.edu

Abstract – One technique to improve the performance of Machine Learning algorithms is to use Ensemble Learning. The idea of this technique combines several Machine Learning algorithms or commonly referred to as base learners. The purpose of this study is to compare the performance of the two Ensemble Learning algorithms, namely the Bootstrap Aggregating (Bagging) method and the Adaptive Boosting (AdaBoost) method. This study uses eleven datasets with multi-class classifications that are independent of the characteristics (data proportion, number of data, and problems) and the number of different classes of target variables. The results showed that the accuracy and F1 model formed by the Bagging method tended to show better value performance than that of the AdaBoost method on the evaluation metric with an average evaluation value of 72.21% and 61% for Bagging and 66.25% and 53, respectively. 7% for AdaBoost. However, the results of hypothesis testing show that it is not significant enough. In addition, the length of computation time to form the Bagging model and the AdaBoost model is not different.

Keywords: Ensemble Learning, Bagging, Boosting, AdaBoost, multi-class classification

Abstrak – Salah satu teknik untuk meningkatkan performa algoritma *Machine Learning* adalah menggunakan *Ensemble Learning*. Ide teknik ini menggabungkan beberapa algoritma *Machine Learning* atau yang biasa disebut sebagai base learners. Tujuan penelitian ini adalah membandingkan dua performa algoritma Ensemble Learning yaitu metode *Bootstrap Aggregating (Bagging)* dan metode *Adaptive Boosting (AdaBoost)*. Penelitian menggunakan sebelas dataset dengan klasifikasi *multi-class* yang independen terhadap karakteristik (proporsi data, jumlah data, dan masalah) serta jumlah kelas variabel target berbeda. Hasil penelitian menunjukkan bahwa akurasi dan F1 model yang dibentuk oleh metode Bagging cenderung menunjukkan performa nilai yang lebih baik dari metode AdaBoost pada metrik evaluasi dengan rata-rata nilai evaluasi sebesar 72,21% dan 61% untuk Bagging serta 66,25% dan 53,7% untuk AdaBoost. Namun hasil pengujian hipotesis memperlihatkan tidak cukup signifikan. Selain itu lama lama waktu komputasi untuk membentuk model Bagging dan model AdaBoost tidaklah berbeda.

Kata Kunci: Ensemble Learning, Bagging, Boosting, AdaBoost, multi-class classification.

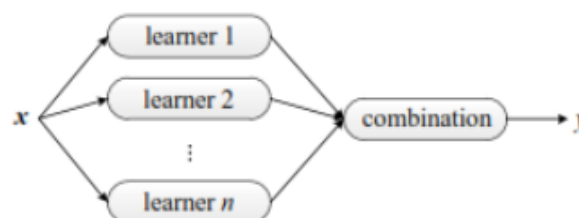
PENDAHULUAN

Segala hal yang dinikmati dan membantu banyak orang adalah hasil dari perkembangan teknologi yang selalu bergerak maju. Algoritma pembelajaran mesin salah satu dari perkembangan teknologi yang saat ini terus berkembang. Pemakainnya jelas berlaku dibanyak bidang diantaranya kesehatan, ekonomi, pendidikan. Seiring berjalannya waktu menciptakan metode yang terbaik untuk dapat mengolah data menjadi semakin efisien dan akurat terus dikembangkan dan diterapkan [1][2].

Pemenang kompetisi meningkatkan performa Netflix classifier sebanyak 10% dengan hadiah sebesar USD 1.000.000 hanya berhasil meningkatkan performa sebesar 8.5% dengan menggunakan algoritma *ensemble learning* pada tanggal 21 September 2009 [1].

Algoritma *ensemble learning* melatih beberapa *learners* untuk memecahkan masalah yang sama dengan

membangun sebuah gabungan dari *learners* dan menyatukan semuanya, gambar 1. [3].



Gambar 1. Penggambaran Ensemble Secara Umum

Sebuah *learner* adalah suatu algoritma pembelajaran mesin yang dapat berupa *decision tree*, *neural network*, lainnya.

Pada penelitian ini akan diuji manakah metoda ensemble learning yang terbaik antara metode Bagging dengan metode Adaboost.

DATA DAN METODOLOGI

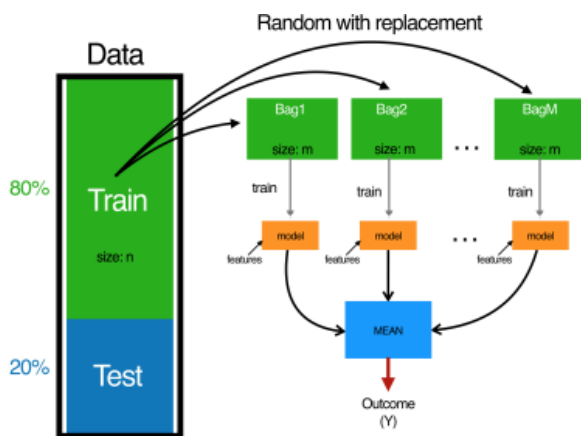
Data yang digunakan dalam penelitian ini terdiri dari 11 data set yang diambil dari situs kaggle.com. Data yang diambil tidak hanya merupakan data dari berbagai sektor seperti medis, keuangan, astronomi, dan lain-lain, tapi juga terdiri dari dua tipe data seimbang dan tidak seimbang. Data yang digunakan telah melewati beberapa tahap yaitu *exploratory data analysis*, *feature engineering*, dan *handling missing values*.

Data kemudian dipisah menjadi dua bagian yakni data pelatihan dan data pengujian dengan proporsi 80% untuk data pelatihan dan 20% untuk data pengujian yang dipilih secara acak.

Metodologi pembentukan model algoritma ensemble dengan basis *logistic regression*, evaluasi model dengan *confusion matrix*, dan kemudian dilakukan pengujian hipotesis parametrik menggunakan uji beda mean.

MODEL BAGGING

Setiap *learner* dilatih dengan sejumlah data yang dipilih secara random dari data pelatihan. Kemudian hasil dari model setiap *learner* digabungkan menjadi luaran dari model bagging. Untuk permasalahan klasifikasi maka luaran model bagging adalah mengambil suara terbanyak dari seluruh luaran model *learner* yang ada (*voting*) sedangkan untuk masalah regresi maka diambil rata-rata luaran dari setiap model *learner*, Gambar 2. [4]



Gambar 2. Proses kerja Model Bagging

MODEL ADABOOST

Perbedaan utama model Adaboost dengan Bagging adalah pada proses pemilihan data pelatihan disetiap *learner*. Pada model bagging setiap data pelatihan mempunyai probabilitas yang sama besar untuk terpilih menjadi data pelatihan pada setiap *learner*. Namun pada model Adaboost tidak [3]. Untuk setiap model *learner* dihitung kinerjanya dengan (1) dimana N adalah jumlah data pelatihan pada model dan T adalah jumlah sampel yang salah prediksi (error) pada data pelatihan itu.

$$K = \frac{1}{2} \ln \left(\frac{1-T/N}{T/N} \right) \dots\dots\dots (1)$$

Semakin besar Error = T/N dari model *learner* maka nilai K semakin kecil. Nilai K inilah yang menentukan

seberapa besar probabilitas sampel data pelatihan terpilih menjadi sampel pada proses pelatihan selanjutnya (2).

$$P(x) = \begin{cases} P(x) \times e^K & x: \text{Error sampel} \\ P(x) \times e^{-K} & x: \text{True sampel} \end{cases} \dots\dots (2)$$

Terlihat kinerja model *learner* semakin kecil maka semakin tinggi probabilitas data pelatihan pada *learner* itu akan terpilih menjadi data pelatihan pada pelatihan selanjutnya.

Setelah seluruh probabilitas data pelatihan diperbaharui maka dilakukan normalisasi data probabilitas setiap data pelatihan.

CONFUSION MATRIX

Confusion matrix adalah salah satu cara untuk mengukur kinerja suatu *learner*. Metode ini dapat diterapkan dalam masalah yang mempunyai prediksi binary maupun multi-class. Output yang dihasilkan dari *confusion matrix* untuk n kelas berupa tabel dengan n² sel diperlihatkan pada Gambar 3.

Empat kondisi yang mungkin terjadi pada *confusion matrix* untuk n = 2, namun Untuk n > 2 maka tidak ada kondisi *True Negatif*

		True Class		
		Apple	Orange	Mango
Predicted Class	Apple	7	8	9
	Orange	1	2	3
	Mango	3	2	1

Gambar 3. Confusion matrik untuk n = 3

- *True Positive (TP)* yang berarti prediksi data positif dan hasilnya benar.
- *True Negative (TN)* yang berarti prediksi data negatif dan hasilnya benar.
- *False Positive (FP)* atau kesalahan tipe pertama, yang berarti prediksi data positif dan hasilnya salah.
- *False Negative (FN)* atau kesalahan tipe dua yang berarti prediksi data negatif dan itu salah.

TIBE DATA SET

Ada dua tipe data set yaitu data seimbang dan data tidak seimbang. Jika distribusi data pada kelas tidak seimbang yang artinya data di salah satu kelas jauh lebih banyak dari data di kelas lainnya maka tipe data set itu adalah tidak seimbang (*imbalance data*) selain itu maka seimbang (*balance data*). Hal ini sangat mungkin terjadi pada banyak bidang seperti medis, penyimpangan data dsb. Pada bidang medis misalnya kelas seseorang sakit jauh lebih kecil dibanding tidak sakit. Pada bidang penyimpangan data, penyimpangan pemakaian kartu kredit lebih sedikit dibanding dengan tidak menyimpang.

Ada empat pengukuran kinerja *learner* yaitu diperlihatkan pada (3)-(6)

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \dots\dots\dots(3)$$

$$Precision = \frac{TP}{TP+FP} \dots\dots\dots(4)$$

$$Recall = \frac{TP}{TP+FN} \dots\dots\dots(5)$$

$$F1 = \frac{2TP}{2TP+FP+FN} \dots\dots\dots(6)$$

Pada penelitian ini akan diuji apakah Kinerja F1 lebih baik dibandingkan dengan kinerja akurasi untuk data yang tidak seimbang

DATA DAN PEMBAHASAN

Data percobaan terdiri dari 11 data set dari berbagai sektor seperti disajikan pada Tabel 1. Setelah melakukan pengolahan data menggunakan python [5], diperoleh hasil akurasi, F1, dan waktu yang dibutuhkan untuk pembentukan model. Hasil pengolahan data diperlihatkan pada Tabel 2.

Tabel 1 : Kesebelas Data Penelitian

Dataset	# record	# attribut	# kelas	Tipe data
Glass	1599	11	6	Imbalance
Sky	1728	6	4	Imbalance
Heart	8068	10	4	Imbalance
Customer	6000	7	9	Balance
Handphone	11678	64	4	Imbalance
Car	214	9	7	Imbalance
Wine	817	9	4	Imbalance
BMI	920	14	5	Balance
Muscle	2111	16	7	Balance
Tech Student	10000	17	3	Balance
Diamond	20000	12	6	Imbalance

Kolom kombinasi pada tabel 2, baik untuk Bagging ataupun Adaboost adalah data Akurasi atau data F1-score dari data yang bersesuaian sesuai dengan tipe datanya. Jika data bertipe Imbalance maka kolom kombinasi adalah data F1-score sedangkan jika data bertipe balance maka data akurasinya.

Tabel 2 : Hasil Pengolahan Data Penelitian

Dataset	Akurasi		F1-Score		Kombinasi		Waktu	
	Bagging	Adaboost	Bagging	Adaboost	Bagging	Adaboost	Bagging	Adaboost
Glass	0,674	0,581	0,623	0,420	0,623	0,420	0,77	0,31
Sky	0,975	0,956	0,972	0,949	0,972	0,949	1,63	12,90
Heart	0,576	0,587	0,328	0,317	0,328	0,317	1,82	0,58
Customer	0,498	0,481	0,456	0,481	0,498	0,481	2,27	3,09
Handphone	0,927	0,927	0,927	0,927	0,927	0,927	3,29	6,15
Car	0,931	0,861	0,858	0,777	0,858	0,777	4,27	12,10
Wine	0,600	0,581	0,293	0,286	0,293	0,286	4,70	3,75
BMI	0,885	0,600	0,885	0,581	0,885	0,6	6,34	2,24
Muscle	0,342	0,291	0,344	0,293	0,342	0,291	7,07	4,29
Tech Student	0,677	0,654	0,675	0,654	0,677	0,654	11,30	4,90
Diamond	0,86	0,768	0,349	0,223	0,349	0,768	7,97	9,64
Rata-rata	0,722	0,662	0,610	0,537	0,614	0,588	4,675	5,450

Kinerja metode Bagging cenderung lebih baik dari metode AdaBoost. Hal ini sesuai hasil penelitian dari salah satu tinjauan pustaka [7] bahwa metode Bagging memberikan hasil yang lebih baik dibandingkan dengan metode ensemble lainnya.

Pengujian normalitas data dilakukan dengan Kalmogorov Smirnov (K1) dan juga Shapiro-Wilk (K2). Hasilnya diperlihatkan pada Tabel 3. Hasil memperlihatkan semua data berdistribusi normal. Pengujian kesamaan variance juga dilakukan dan diperlihatkan pada tabel 4. Hasil memperlihatkan masing-masing memiliki kesamaan variance.

Tabel 3. Hasil Uji Normalitas Data

		p-value	
		K1	K2
Akurasi	Bagging	0,2	0,36
	Adaboost	0,2	0,57
F1-Score	Bagging	0,2	0,08
	Adaboost	0,2	0,25
Kombinasi	Bagging	0,2	0,08
	Adaboost	0,2	0,25
Waktu	Bagging	0,2	0,50
	Adaboost	0,2	0,22

Tabel 4. Hasil Uji Normalitas Data

	p-value
Akurasi Bagging - Adaboost	0,91
F1-Score Bagging - Adaboost	0,94
Kombinasi Bagging - Adaboost	0,99
Waktu Bagging - Adaboost	0,36
Seluruh Bagging	0,70
Seluruh Adaaboost	0,65

Untuk menguji apakah metode Bagging lebih baik dibandingkan dengan metode Adaboost dilakukan uji mean dua populasi. Pengujian dilakukan untuk melihat 4 hipotesa

- apakah akurasi metode Bagging lebih tinggi dari metode Adaboost
- apakah F1-score metode Bagging lebih tinggi dari metode Adaboost
- apakah kombinasi metode Bagging lebih tinggi dari metode Adaboost
- apakah wktu pemodelan metode Bagging lebih cepat dari metode Adaboost

Hasil pengujian diperlihatkan pada tabel 5. Hasil memperlihatkan bahwa keempat hipotesa menolak hipotesa alternatif yang berarti bahwa akurasi, F1-score, kombinasi dan waktu prosesnya metode Bagging tidak lebih baik dibanding dengan Metode Adaboost.

Tabel 5. Hasil Uji Mean dua populasi

	p-value
Akurasi Bagging - Adaboost	0,25
F1-Score Bagging - Adaboost	0,26
Kombinasi Bagging - Adaboost	0,26
Waktu Bagging - Adaboost	0,68

Pengujian lebih lanjut apakah Akurasi, F1-score dan Kombinasinya memiliki nilai mean yang sama baik untuk Metode Bagging maupun Adaboost. Pengujian mean Akurasi, F1-score dan Kombinasinya untuk metode Bagging menghasilkan p-value sebesar 0,48 dan metode Adaboost sebesar 0,39 mengindikasikan bahwa mereka memiliki mean yang sama. Temuan ini memperkuat kesimpulan bahwa tidak cukup data untuk mengatakan metode Bagging lebih baik dibandingkan dengan metode Adaboost baik untuk Akurasi maupun nilai F1-score.

Ketidaksesuaian hasil pengujian ini jika dibandingkan dengan penelitian lainnya [7] bahwa metode Bagging lebih baik dibanding dengan metode Adaboost sangat mungkin dikarenakan data set penelitian menggunakan data beragam baik data balance dan tidak balance.

KESIMPULAN

Berdasarkan analisis yang dilakukan dapat disimpulkan

1. Tidak cukup data untuk mengatakan metode Bagging lebih baik dibandingkan dengan metode Adaboost baik untuk Akurasi maupun nilai F1-score,
2. Tidak cukup data untuk mengatakan waktu proses pembentukan model metode Bagging lebih cepat dibandingkan dengan metode Adaboost.

DAFTAR PUSTAKA

- [1] Robert M Bell and Yehuda Koren. "Lessons from the netflix prize challenge". *ACM Sigkdd Explorations Newsletter*, 9(2):75-79, 2007.
- [2] Haihua Jiang, Bin Hu, Zhenyu Liu, Gang Wang, Lan Zhang, Xiaoyu Li, and Huanyu Kang. "Detecting depression using

an ensemble logistic regression model based on multiple speech features". *Computational and Mathematical Methods in Medicine*, 2018:1-9, 2018.

- [3] Zhi-Hua Zhou. *Ensemble methods: foundations and algorithms*. Chapman and Hall/CRC, 2019.
- [4] Jiawei Han, Jian Pei, and Micheline Kamber. "Data mining: concepts and techniques", *Elsevier*, 2011.
- [5] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. "Scikit-learn: Machine learning in python", *Journal of machine learning research*, 12(Oct):2825-2830, 2011.
- [6] Joydwip Mohajon. *Confusion matrix for your multi-class machine learning model*. Towards Data Science, 2020.
- [7] Yanli Wu, Yutian Ke, Zhuo Chen, Shouyun Liang, Hongliang Zhao, and Haoyuan Hong. "Application of alternating decision tree with adaboost and bagging ensembles for landslide susceptibility mapping". *Catena*, 187:104396, 2020.