

PENERAPAN TEKNIK BAGGING PADA ALGORITMA KLASIFIKASI UNTUK MENGATASI KETIDAKSEIMBANGAN KELAS DATASET MEDIS

Rizki Tri Prasetyo¹, dan Pratiwi²

¹ Universitas BSI Bandung, rizki.rte@bsi.ac.id

² Universitas BSI Bandung, pratiwi.pwi@bsi.ac.id

ABSTRACT – *The class imbalance problems have been reported to severely hinder classification performance of many standard learning algorithms, and have attracted a great deal of attention from researchers of different fields. Therefore, a number of methods, such as sampling methods, cost-sensitive learning methods, and bagging and boosting based ensemble methods, have been proposed to solve these problems. Some medical dataset has two classes or binominal experiencing an imbalance that causes lack of accuracy in classification. This research proposed a combination technique of bagging and algorithms of classification to improve the accuracy of medical datasets. Bagging technique used to solve the problem of imbalanced class. The proposed method is applied on three classifier algorithm i.e., naïve bayes, decision tree and k-nearest neighbor. This research uses five medical datasets obtained from UCI Machine Learning i.e., breast-cancer, liver-disorder, heart-disease, pima-diabetes and vertebral column. Results of this research indicate that the proposed method makes a significant improvement on two algorithms of classification i.e. decision tree with p value of t-Test 0.0184 and k-nearest neighbor with p value of t-Test 0.0292, but not significant in naïve bayes with p value of t-Test 0.9236. After bagging technique applied at five medical datasets, naïve bayes has the highest accuracy for breast-cancer dataset of 96.14% with AUC of 0.984, heart-disease of 84.44% with AUC of 0.911 and pima-diabetes of 74.73% with AUC of 0.806. While the k-nearest neighbor has the best accuracy for dataset liver-disorder of 62.03% with AUC of 0.632 and vertebral-column of 82.26% with the AUC of 0.867.*

Keywords: *ensemble technique, bagging, imbalanced class, medical dataset.*

ABSTRAKSI – Masalah ketidakseimbangan kelas telah dilaporkan sangat menghambat kinerja klasifikasi banyak algoritma klasifikasi dan telah menarik banyak perhatian dari para peneliti dari berbagai bidang. Oleh karena itu, sejumlah metode seperti metode *sampling*, *cost-sensitive learning*, serta *bagging* dan *boosting*, telah diusulkan untuk memecahkan masalah ini. Beberapa dataset medis yang memiliki dua kelas atau *binominal* mengalami ketidakseimbangan kelas yang menyebabkan kurangnya akurasi pada klasifikasi. Pada penelitian ini diusulkan kombinasi teknik *bagging* dan algoritma klasifikasi untuk meningkatkan akurasi dari klasifikasi dataset medis. Teknik *bagging* digunakan untuk menyelesaikan masalah ketidakseimbangan kelas. Metode yang diusulkan diterapkan pada tiga algoritma *classifier* yaitu, *naïve bayes*, *decision tree* dan *k-nearest neighbor*. Penelitian ini menggunakan lima dataset medis yang didapatkan dari UCI Machine Learning yaitu, *breast-cancer*, *liver-disorder*, *heart-disease*, *pima-diabetes* dan *vertebral column*. Hasil penelitian menunjukkan bahwa metode yang diusulkan membuat peningkatan yang signifikan pada dua algoritma klasifikasi yaitu *decision tree* dengan P value of t-Test sebesar 0,0184 dan *k-nearest neighbor* dengan P value of t-Test sebesar 0,0292, akan tetapi tidak signifikan pada *naïve bayes* dengan P value of t-Test sebesar 0,9236. Setelah diterapkan teknik *bagging* pada lima dataset medis, *naïve bayes* memiliki akurasi paling tinggi untuk dataset *breast-cancer* sebesar 96,14% dengan AUC sebesar 0,984, *heart-disease* sebesar 84,44% dengan AUC sebesar 0,911 dan *pima-*

diabetes sebesar 74,73% dengan AUC sebesar 0,806. Sedangkan *k-nearest neighbor* memiliki akurasi yang paling baik untuk dataset *liver-disorder* sebesar 62,03% dengan AUC sebesar 0,632 dan *vertebral column* dengan akurasi sebesar 82,26% dengan AUC sebesar 0,867.

Kata Kunci: teknik *ensemble*, *bagging*, ketidakseimbangan kelas, dataset medis.

I. PENDAHULUAN

Dataset yang dikategorikan memiliki kelas yang tidak seimbang mengacu pada setidaknya salah satu kelas yang memiliki jumlah lebih sedikit jika dibandingkan dengan kelas lain (Chao, Liu, & Ding, 2013). Untuk masalah ketidakseimbangan dua kelas (binomial), data biasanya dikategorikan menjadi kelas mayoritas dan kelas minoritas. Secara umum, kelas minoritas biasanya mewakili sebuah konsep dengan kriteria lebih baik dari kelas mayoritas. Akan tetapi, sering kali kelas minoritas ini kalah jumlah oleh kelas mayoritas dan terkadang skenario ini mungkin sangat parah (Chawla, Japkowicz, & Kotcz, 2004).

Sebagai algoritma klasifikasi yang paling tradisional, seperti *decision tree*, *k-nearest neighbor* dan RIPPER, cenderung menghasilkan model yang memaksimalkan akurasi klasifikasi secara keseluruhan sehingga kelas minoritas diabaikan (Sun, Kamel, & Wong, 2007). Misalnya, untuk dataset yang hanya memiliki 1% untuk kelas minoritas dari total dataset yang tersedia, bahkan jika algoritma klasifikasi tersebut memodelkan semua hal sebagai kelas mayoritas, maka model akan menghasilkan akurasi keseluruhan sekitar 99%. Namun, kasus kelas minoritas yang kita ingin klasifikasikan secara akurat akan mengalami *misclassification* dengan model ini meskipun akurasi keseluruhan akan sangat tinggi. Oleh karena itu, beberapa metode telah diusulkan untuk menangani masalah-masalah klasifikasi binomial yang tidak seimbang.

Salah satu metode yang paling populer untuk memecahkan masalah ketidakseimbangan kelas adalah teknik *sampling* (Barandela, Sanchez, Garcia, & Rangel, 2003). Teknik yang paling sering digunakan adalah teknik *under-sampling* dan *over-sampling* untuk mengubah distribusi kelas asli dari dataset yang tidak seimbang dengan menghilangkan sebagian besar data kelas mayoritas atau

meningkatkan data dari kelas minoritas. Selain itu, *cost-sensitive learning* juga dapat digunakan untuk memecahkan masalah-masalah ketidakseimbangan kelas (Zhou & Liu, 2006). Selain itu, *bagging* dan *boosting* dapat digunakan untuk menangani masalah-masalah yang tidak seimbang (Hido, Kashima, & Takahashi, 2009).

Secara umum terdapat tiga pendekatan untuk menangani dataset tidak seimbang (unbalanced), yaitu pendekatan pada level data, level algoritmik, dan menggabungkan atau memasang (ensemble) metode (Yap, et al., 2014). Pendekatan pada level data mencakup berbagai teknik resampling dan sintesis data untuk memperbaiki kecondongan distribusi kelas data latih. Pada tingkat algoritmik, metode utamanya adalah menyesuaikan operasi algoritma yang ada untuk membuat pengklasifikasi (classifier) agar lebih konduktif terhadap klasifikasi kelas minoritas (Zhang, Liu, Gong, & Jin, 2011). Sedangkan pada pendekatan menggabungkan atau memasang (ensemble) metode, ada dua algoritma ensemble-learning paling populer, yaitu *boosting* dan *bagging* (Yap, et al., 2014). Algoritma *boosting* telah dilaporkan sebagai meta-teknik untuk mengatasi masalah ketidakseimbangan kelas (class imbalance) (Sun, Kamel, & Wong, 2007). Pada pendekatan algoritma dan ensemble memiliki tujuan yang sama, yaitu memperbaiki algoritma pengklasifikasi tanpa mengubah data, sehingga dapat dianggap ada 2 pendekatan saja, yaitu pendekatan level data dan pendekatan level algoritma (Peng & Yao, 2010). Dengan membagi menjadi 2 pendekatan dapat mempermudah fokus objek perbaikan, pendekatan level data difokuskan pada pengolahan awal data, sedangkan pendekatan level algoritma difokuskan pada perbaikan algoritma atau menggabungkan (ensemble).

Bagging dan Boosting telah berhasil meningkatkan akurasi pengklasifikasi tertentu untuk dataset buatan dan yang sebenarnya. Bagging adalah metode ensemble yang sederhana namun efektif dan telah diterapkan untuk banyak aplikasi di dunia nyata (Liang & Zhang, 2011). Bagging merupakan metode ensemble yang banyak diterapkan pada algoritma klasifikasi, dengan tujuan untuk meningkatkan akurasi pengklasifikasi dengan menggabungkan pengklasifikasi tunggal, dan hasilnya lebih baik daripada random sampling (Alfaro, Gamez, & Garcia, 2013). Secara umum algoritma Boosting lebih baik dari pada Bagging, tetapi tidak merata baik. Boosting telah menunjukkan dapat meningkatkan kinerja pengklasifikasi dalam banyak situasi, termasuk ketika data tidak seimbang (Seiffert, Khoshgoftaar, Hulse, & Napolita, 2008). AdaBoost adalah kependekan dari Adaptive merupakan algoritma machine learning yang dirumuskan oleh Yoav Freund and Robert Schapire. AdaBoost secara teoritis dapat secara signifikan digunakan untuk mengurangi kesalahan dari beberapa algoritma pembelajaran yang secara konsisten menghasilkan kinerja pengklasifikasi yang lebih baik. AdaBoost diterapkan pada Naïve Bayes dapat meningkatkan kinerja sebesar 33,33% dan memberikan hasil yang akurat dengan mengurangi nilai kesalahan klasifikasi dengan meningkatkan iterasi (Korada, Kumar, & Deekshitulu, 2012). Beberapa penelitian tersebut telah menunjukkan bahwa metode ensemble (AdaBoost dan Bagging) dapat memperbaiki kinerja pengklasifikasi.

Penelitian ini mengusulkan kombinasi teknik *bagging* dengan tiga algoritma klasifikasi yaitu, naïve bayes, *decision tree* dan *k-nearest neighbor* yang diterapkan untuk mengklasifikasi dataset medis publik (*breast-cancer*, *liver-disorder*, *heart-disease*, *pima-diabetes* dan *vertebral-column*). Kelima dataset tersebut mengalami masalah klasifikasi yang sama yaitu, ketidakseimbangan kelas. Ketidakseimbangan kelas menyebabkan kurangnya performa akurasi dalam klasifikasi. Maka dengan penerapan teknik *bagging* pada ketiga algoritma klasifikasi

tersebut diharapkan dapat meningkatkan performa akurasi klasifikasi kelima dataset tersebut.

II. KAJIAN LITERATUR

A. *Decision Tree* (C4.5)

Algoritma C4.5 diperkenalkan oleh J. Ross Quinlan yang merupakan pengembangan dari algoritma ID3, algoritma tersebut digunakan untuk membentuk pohon keputusan (Alfisahrin, 2014). Pohon keputusan dianggap sebagai salah satu pendekatan yang paling populer. Dalam klasifikasi pohon keputusan terdiri dari sebuah *node* yang membentuk akar. *Node* akar tidak memiliki *input*. *Node* lain yang bukan sebagai akar tetapi memiliki tepat satu *input* disebut *node internal* atau *test node*, sedangkan *node* lainnya dinamakan daun. Daun mewakili nilai target yang paling tepat dari salah satu *class* (Maimon & Rokach, 2010).

Pohon keputusan dibuat dengan membagi nilai-nilai atribut menjadi cabang untuk setiap kemungkinan. Cara kerja pohon keputusan yaitu dengan melakukan penelusuran dari akar hingga ke cabang sampai *class* suatu objek ditemukan. *Instance* diklasifikasikan dengan mengarahkan dari akar pohon sampai ke daun sesuai dengan hasil tes melalui *node internal* (Alfisahrin, 2014).

Pada dasarnya konsep dari algoritma C4.5 adalah mengubah data menjadi pohon keputusan dan aturan-aturan keputusan (*rule*). C4.5 adalah algoritma yang cocok untuk masalah klasifikasi dan *data mining*. C4.5 memetakan nilai atribut menjadi *class* yang dapat diterapkan untuk klasifikasi baru (Wu & Kumar, 2009). Menurut Kusrini dan Luthfi, sebuah pohon keputusan terdiri dari sekumpulan aturan untuk membagi sejumlah populasi yang heterogen menjadi lebih kecil, lebih homogen dengan memperhatikan pada variabel tujuannya (Alfisahrin, 2014). Variabel tujuan dikelompokkan dengan pasti dan model pohon keputusan lebih mengarah pada perhitungan probabilitas dari tiap-tiap *records* terhadap kategori-kategori tersebut atau untuk mengklasifikasi *record* dengan mengelompokkannya dalam satu *class*.

B. Naïve Bayes

Naive Bayes merupakan metode yang tidak memiliki aturan (Alfisahrin, 2014). Naive Bayes menggunakan cabang matematika yang dikenal dengan teori probabilitas untuk mencari peluang terbesar dari kemungkinan klasifikasi, dengan cara melihat frekuensi tiap klasifikasi pada *data training*. Naive Bayes merupakan metode klasifikasi populer dan masuk dalam sepuluh algoritma terbaik dalam *data mining* (Wu & Kumar, 2009). Algoritma ini juga dikenal dengan nama *Idiot's Bayes*, *Simple Bayes*, dan *Independence Bayes* (Bramer, 2013).

Klasifikasi Naive Bayes adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu *class*. Klasifikasi Bayesian didasarkan pada teorema Bayes, diambil dari nama seorang ahli matematika yang juga menteri *Prebysterian* Inggris, Thomas Bayes (1702-1761) (Bramer, 2013). Menurut Kusri dan Luthfi, klasifikasi bayesian memiliki kemampuan klasifikasi serupa dengan *decision tree* dan *neural network* (Alfisahrin, 2014).

Bayes rule digunakan untuk menghitung probabilitas suatu *class*. Algoritma Naive Bayes memberikan suatu cara mengkombinasikan peluang terdahulu dengan syarat kemungkinan menjadi sebuah formula yang dapat digunakan untuk menghitung peluang dari tiap kemungkinan yang terjadi (Alfisahrin, 2014).

C. K-Nearest Neighbor

Klasifikasi merupakan proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Metode-metode klasifikasi antara lain *C4.5*, *Random Forest*, *Naïve Bayes*, *neural network*, *genetic algorithm*, *fuzzy*, *case-based reasoning*, dan *k-Nearest Neighbor* (Larose, 2005)

Algoritma k-NN adalah suatu metode yang menggunakan algoritma *supervised* (Wu & Kumar, 2009). Perbedaan antara *supervised learning* dengan *unsupervised learning* adalah pada *supervised learning*

bertujuan untuk menemukan pola baru dalam data dengan menghubungkan pola data yang sudah ada dengan data yang baru. Sedangkan pada *unsupervised learning*, data belum memiliki pola apapun, dan tujuan *unsupervised learning* untuk menemukan pola dalam sebuah data (Han & Kamber, 2006). Tujuan dari algoritma k-NN adalah untuk mengklasifikasi objek baru berdasarkan atribut dan training samples (Larose, 2005). Dimana hasil dari sampel uji yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada k-NN. Pada proses pengklasifikasian, algoritma ini tidak menggunakan model apapun untuk dicocokkan dan hanya berdasarkan pada memori. Algoritma k-NN menggunakan klasifikasi ketetangaan sebagai nilai prediksi dari sampel uji yang baru. Pemilihan terhadap teknik perhitungan jarak merupakan hal penting lainnya. Biasanya *Euclidean Distance* dan *Manhattan Distance* digunakan untuk menghitung jarak pada algoritma k-NN (Wu & Kumar, 2009).

D. Teknik Bagging

Bagging adalah singkatan dari *bootstrap aggregating*, menggunakan subdataset (*bootstrap*) untuk menghasilkan set pelatihan L (*learning*), L melatih dasar belajar menggunakan prosedur pembelajaran yang tidak stabil, dan kemudian, selama pengujian, mengambil rata-rata (Breiman, 1996). *Bagging* baik digunakan untuk klasifikasi dan regresi. Dalam kasus regresi, untuk menjadi lebih kuat, seseorang dapat mengambil rata-rata ketika menggabungkan prediksi. *Bagging* adalah (Alpaydin, 2010) sebuah algoritma pembelajaran yang stabil pada perubahan kecil dalam training set menyebabkan perbedaan besar dalam peserta didik yang dihasilkan, yaitu algoritma belajar pada data yang memiliki varians tinggi (*noise*). *Noise* yang tinggi mempengaruhi klasifikasi baru yang dihasilkan, sehingga menyebabkan *miss-classification* (Kim & Kang, 2012). *Noise* dalam hal akurasi prediksi dapat ditingkatkan dengan menggunakan *classifier-ensembles* (Twala, 2010).

Model *ensemble* berlaku dengan penggabungan berbagai teknik pengambilan sampel seperti *bagging*, *boosting*, dan lain-lain untuk menjamin keragaman di kolom *classifier*. Penanganan *noise* data merupakan masalah penting untuk proses pembelajaran klasifikasi, sejak terjadinya *noise* yang tinggi dalam proses pelatihan atau pengujian (klasifikasi) pada dataset, mempengaruhi keakuratan prediksi pengklasifikasi yang dipelajari. Hal ini dapat menimbulkan masalah serius bagi para peneliti. Sebagai contoh, pengenalan bias dapat menghasilkan kesimpulan yang menyesatkan (*miss-classification*) (Twala, 2010).

Data *bootstrap* diciptakan oleh *resampling* contoh seragam dengan penggantian dari data pelatihan asli. Pengklasifikasian dapat dilatih secara paralel dan klasifikasi akhir dihasilkan dengan menggabungkan klasifikasi *ensemble*. *Bagging* telah dianggap sebagai teknik pengurangan varians untuk *classifier* yang diberikan. *Bagging* dikenal sangat efektif ketika pengklasifikasi tidak stabil, yaitu ketika *perturbing set belajar* dapat menyebabkan perubahan signifikan dalam perilaku klasifikasi, karena *bagging* meningkatkan kinerja generalisasi karena pengurangan varians (*noise*) tetap terjaga atau hanya sedikit meningkatkan bias (Kim & Kang, 2012).

E. Confusion Matrix

Menurut Gorunescu, evaluasi kinerja model klasifikasi didasarkan pada pengujian objek yang diprediksi dengan benar dan salah, hitungan ini ditabulasikan *confusion matrix* (Alfisahrin, 2014). *Confusion Matrix* adalah alat yang berguna untuk menganalisis seberapa baik *classifier* dapat mengenali tupel dari kelas yang berbeda (Han & Kamber, 2006). *Confusion matrix* memberikan rincian klasifikasi, kelas yang diprediksi akan ditampilkan dibagian atas *matrix* dan kelas yang diobservasi ditampilkan dibagian kiri.

Sensitivitas dan spesifitas dapat digunakan sebagai ukuran statistik dari kinerja klasifikasi *biner*, sensitivitas dan spesifitas digunakan untuk mengukur model yang paling baik dan untuk memilih

model yang paling efisien. Sensitivitas mengukur proporsi *true positive* yang diidentifikasi dengan benar, spesifitas mengukur proporsi *true negative* yang diidentifikasi dengan benar. *False positif* dikenal sebagai *error* tipe 1, terjadi ketika kasus yang seharusnya diklasifikasikan sebagai negatif diklasifikasikan sebagai positif. *False negatif* dikenal sebagai *error* tipe 2, terjadi ketika kasus yang seharusnya diklasifikasikan sebagai positif diklasifikasi sebagai negatif (Bramer, 2013).

Akurasi dapat dihitung dengan menggunakan persamaan 2.9 berikut:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

TP:Jumlah kasus positif yang diklasifikasi sebagai positif

FP:Jumlah kasus negatif yang diklasifikasi sebagai positif

TN:Jumlah kasus negatif yang diklasifikasi sebagai negatif

FN:Jumlah kasus positif yang diklasifikasi sebagai negatif

F. Kurva ROC

Menurut Gorunescu, kurva ROC banyak digunakan untuk menilai hasil prediksi, kurva ROC adalah teknik untuk memvisualisasikan, mengatur, dan memilih pengklasifikasian berdasarkan kinerja mereka (Alfisahrin, 2014).

Kurva ROC adalah *tool* dua dimensi yang digunakan untuk menilai kinerja klasifikasi yang menggunakan dua *class* keputusan, masing-masing objek dipetakan ke salah satu elemen dari himpunan pasangan, positif atau negatif. Pada kurva ROC, *TP rate* diplot pada sumbu Y dan *FP rate* diplot pada sumbu X.

Untuk klasifikasi *data mining* menurut Gorunescu, nilai AUC dapat dibagi menjadi beberapa kelompok (Alfisahrin, 2014):

1. 0,90 – 1,00= *Excellent Classification*
2. 0,80 – 0,90= *Good Classification*
3. 0,70 – 0,80= *Fair Classification*
4. 0,60 – 0,70= *Poor Classification*
5. 0,50 – 0,60= *Failure*

III. METODE PENELITIAN

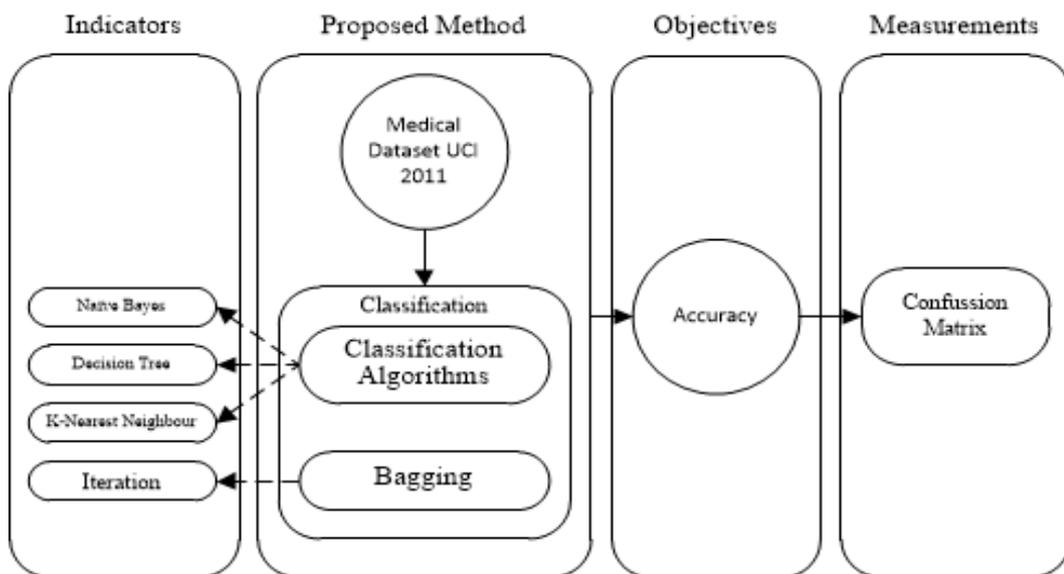
Metode yang diusulkan yaitu menerapkan teknik *bagging* pada beberapa algoritma klasifikasi diantaranya naïve bayes, decision tree (C4.5) dan k-nearest neighbor untuk mengklasifikasikan dataset medikal yang memiliki ketidakseimbangan pada kelasnya (*imbalanced class*). Dataset yang digunakan pada penelitian ini menggunakan dataset sekunder. Data sekunder adalah data yang tidak diperoleh langsung dari obyek penelitian, melainkan telah dikumpulkan oleh pihak lain (Kothari, 2004). Data sekunder yang digunakan pada penelitian ini merupakan kumpulan data biomedis yang diambil langsung dari UCI Machine Learning Repository yang dapat diunduh melalui situs <http://archive.ics.uci.edu>. Dataset yang digunakan pada penelitian ini merupakan dataset medikal seperti *breast-cancer*, *heart-disease*, *liver-disorder*, *pima-diabetes* dan *vertebral column*. Penelitian ini dimulai dari membagi dataset menjadi data *training* dan data *testing* menggunakan *cross validation*. Kemudian menerapkan teknik *bagging*

pada algoritma klasifikasi naïve bayes, *decision tree* dan *k-nearest neighbor* untuk kemudian mengevaluasi hasil klasifikasi menggunakan *confusion matrix* dan kurva ROC. Metode yang diusulkan dapat dilihat pada Gambar 1.

IV. PEMBAHASAN

Penelitian dilakukan dalam dua eksperimen yaitu eksperimen terhadap algoritma *classifier* (naïve bayes, *decision tree* dan *k-nearest neighbor*) tanpa menggunakan optimasi teknik *bagging* (Tabel 1) dan menggunakan optimasi teknik *bagging* (Tabel 2). Kedua eksperimen tersebut diterapkan pada lima dataset medis berbeda yang memiliki kesamaan karakteristik, yaitu *breast-cancer*, *heart-disease*, *liver-disorder*, *pima-diabetes* dan *vertebral column*. *Cross validation* digunakan untuk validasi pada kedua eksperimen tersebut.

Hasil eksperimen yang tertuang pada Tabel I menyatakan bahwa naïve bayes merupakan algoritma paling



Gambar 1. Metode Penelitian

baik untuk klasifikasi *breast-cancer* dengan akurasi 96,14% dengan AUC sebesar 0,978. Sedangkan, klasifikasi *heart-disease* dapat dilakukan dengan baik oleh algoritma naïve bayes dengan akurasi 84,44% dengan AUC sebesar 0,913. Hasil

klasifikasi pada dataset *liver-disorder* diperoleh kurang baik oleh *k-nearest neighbor* dengan akurasi hanya 62,03% dengan AUC hanya sebesar 0,500. Dataset *pima-diabetes*, akurasi terbaik diperoleh oleh naïve bayes dengan akurasi 74,99% dengan AUC sebesar 0,805. Dataset

vertebral-column, akurasi terbaik diperoleh oleh *k-nearest neighbor* dengan akurasi 82,26% akan tetapi memiliki AUC rendah dengan nilai 0,500 jauh dibawah hasil naïve bayes sebesar 0,875 dengan akurasi 78,06%.

Hasil eksperimen yang tertuang pada Tabel 2 menyatakan bahwa setelah penerapan teknik *bagging* pada ketiga algoritma klasifikasi tersebut terdapat kenaikan pada performa akurasi yang dihasilkan oleh setiap algoritma. Algoritma naïve bayes merupakan algoritma paling baik untuk klasifikasi *breast-cancer* dengan akurasi 96,14% dengan AUC sebesar 0,984. Sedangkan, klasifikasi *heart-disease* dapat dilakukan dengan baik

nearest neighbor dengan akurasi hanya 62,03% dengan AUC yang mengalami peningkatan sebesar 0,632. Dataset *pima-diabetes*, akurasi terbaik diperoleh oleh naïve bayes dengan akurasi mengalami penurunan menjadi 74,73% dengan AUC sebesar 0,806. Dataset *vertebral-column*, akurasi terbaik diperoleh oleh *k-nearest neighbor* dengan akurasi 82,26% dan AUC yang meningkat sangat tinggi menjadi 0,867.

Berdasarkan hasil eksperimen pada penelitian ini, diperoleh perbedaan nilai akurasi pada pengujian algoritma naïve bayes, *decision tree* dan *k-nearest neighbour* teknik *bagging* pada kelima dataset medis tersebut, untuk mengetahui

Tabel 1. Hasil Eksperimen Tanpa Menggunakan Teknik Bagging

Hasil Eksperimen Tanpa Menggunakan Teknik Bagging										
Algoritma Klasifikasi	breast-cancer		heart-disease		liver-disorder		diabetes		vertebral-column	
	Akurasi	AUC	Akurasi	AUC	Akurasi	AUC	Akurasi	AUC	Akurasi	AUC
KNN	90,42%	0,5	47,41%	0,5	62,03%	0,5	67,98%	0,5	82,26%	0,5
NB	96,14%	0,978	84,44%	0,913	56,55%	0,649	74,99%	0,805	78,06%	0,875
C4.5	94,70%	0,945	73,70%	0,713	61,73%	0,545	71,86%	0,604	78,71%	0,74

oleh algoritma naïve bayes dengan akurasi 84,44% dengan AUC

sebesar 0,911, akan tetapi hasil ini mengalami penurunan apabila dibandingkan dengan hasil dari eksperimen tanpa menggunakan teknik *bagging*.

Hasil klasifikasi pada dataset *liver-disorder* diperoleh kurang baik oleh k-

apakah metode yang diusulkan dapat meningkatkan performa dalam klasifikasi dataset medis secara signifikan, maka pengujian menggunakan uji beda dilakukan. t-Test Paired Two Sample for Means digunakan pada setiap hasil algoritma klasifikasi antara sebelum dan sesudah menggunakan teknik *bagging*.

Tabel 2. Hasil Eksperimen Dengan Menggunakan Teknik Bagging

Hasil Eksperimen Dengan Menggunakan Teknik Bagging										
Algoritma Klasifikasi	breast-cancer		heart-disease		liver-disorder		Diabetes		vertebral-column	
	Akurasi	AUC	Akurasi	AUC	Akurasi	AUC	Akurasi	AUC	Akurasi	AUC
KNN	90,42%	0,901	57,41%	0,609	62,03%	0,632	67,85%	0,672	82,26%	0,867
NB	96,14%	0,984	84,44%	0,911	57,45%	0,643	74,73%	0,806	78,06%	0,875
C4.5	94,99%	0,96	76,30%	0,817	62,02%	0,569	71,73%	0,706	79,68%	0,856

Hasil pengujian t-Test Paired Two Sample for Means menghasilkan bahwa teknik *bagging* dapat meningkatkan performa dua algoritma klasifikasi secara signifikan pada kelima dataset medis, yaitu *k-nearest neighbor* dengan P value of t-Test < 0,05 (0,0184) dan *decision tree* dengan P value of t-Test < 0,05 (0,0292). Akan tetapi

teknik *bagging* tidak terlalu meningkatkan performa algoritma naïve bayes dengan signifikan, ditandai dengan P value of t-Test > 0,05 (0,9236). Hasil pengujian t-Test dapat dilihat pada Tabel 3.

Tabel 3. Hasil Pengujian Uji Beda

Algoritma	P Value of t-Test	Hasil
KNN	0,0184	Sig. (P < 0,05)
NB	0,9236	Not Sig. (P > 0,05)
C4.5	0,0292	Sig. (P < 0,05)

Hasil dari kedua eksperimen tersebut menjelaskan bahwa naïve bayes dan *k-nearest neighbor* dapat dengan baik mengklasifikasi dataset medis. Dengan menerapkan teknik *bagging* pada algoritma klasifikasi dapat meningkatkan performa akurasi secara signifikan pada algoritma *k-nearest neighbor* dan *decision tree* (C4.5).

V. PENUTUP

Teknik *bagging* diterapkan pada beberapa algoritma classifier (naïve bayes, *decision tree* dan *k-nearest neighbour*) untuk menyelesaikan masalah ketidakseimbangan kelas pada lima dataset medis. Dari dua kali eksperimen menggunakan teknik *bagging* dan tidak menggunakan teknik *bagging*.

Dari dua kali eksperimen, teknik *bagging* terbukti efektif pada dua algoritma klasifikasi (*k-nearest neighbor* dan *decision tree*) dapat meningkatkan hasil akurasi dari klasifikasi lima dataset medis, selain itu hasil uji beda antara tiga algoritma classifier yang dikombinasikan dengan teknik *bagging* dengan tiga algoritma classifier tanpa *bagging* menghasilkan perbedaan yang signifikan pada *k-nearest neighbor* dan *decision tree*, akan tetapi tidak signifikan pada naïve bayes.

Komparasi algoritma klasifikasi diusulkan untuk membandingkan hasil akurasi antara naïve bayes, *decision tree* dan *k-nearest neighbour* yang sudah dikombinasikan dengan teknik *bagging*. Dari lima dataset medis, naïve bayes memiliki akurasi paling tinggi untuk dataset *breast-cancer* sebesar 96,14% dengan AUC sebesar 0,984, *heart-disease* sebesar 84,44% dengan AUC sebesar 0,911 dan *pima-diabetes* sebesar 74,73%

dengan AUC sebesar 0,806. Sedangkan *k-nearest neighbor* memiliki akurasi yang paling baik untuk dataset *liver-disorder* sebesar 62,03% dengan AUC sebesar 0,632 dan *vertebral column* dengan akurasi sebesar 82,26% dengan AUC sebesar 0,867.

Pada penelitian ini, secara umum dengan menerapkan teknik *bagging* untuk mengatasi masalah ketidakseimbangan kelas dapat meningkatkan akurasi pada klasifikasi dataset *breast-cancer*, *heart-disease*, *liver-disorder*, *pima-database* dan *vertebral-column*, namun beberapa hal dapat diterapkan untuk menyempurnakan penelitian ini, yaitu menggunakan algoritma metaheuristik lain untuk fitur seleksi dan penerapan teknik ensemble lain untuk mengatasi masalah ketidakseimbangan kelas serta menambahkan algoritma klasifikasi lain.

DAFTAR PUSTAKA

- Alfaro, E., Gamez, M., & Garcia, N. (2013). adabag: An R Package for Classification with Boosting and Bagging. *Journal of Statistical Software*, 11-35.
- Alfisahrin, S. N. (2014). *Komparasi Algoritma C4.5, Naive Bayes dan Neural Network Untuk Memprediksi Penyakit Jantung*. Jakarta: Pascasarjana Magister Ilmu Komputer STMIK Nusa Mandiri.
- Alpaydin, E. (2010). *Introduction to Machine Learning*. London: The MIT Press.
- Barandela, R., Sanchez, J., Garcia, V., & Rangel, E. (2003). Strategies for learning in class imbalance problems. *Pattern Recognition*, 849-851.
- Bramer, M. (2013). *Principle of Data Mining Second Edition*. London: Springer.
- Breiman, L. (1996). Bagging Predictors. *Machine Learning*, 123-140.
- Chao, W., Liu, J., & Ding, J. (2013). Facial age estimation based on label-sensitive learning and age-oriented regression. *Pattern Recognition*, 628-641.

- Chawla, N., Japkowicz, N., & Kotcz, A. (2004). Editorial: special issue on learning from imbalanced data sets. *ACM SIGKDD Explore*, 1-6.
- Han, J., & Kamber, M. (2006). *Data Mining Concepts and Techniques Second Edition*. San Francisco: Diane Cerra.
- Hido, S., Kashima, H., & Takahashi, Y. (2009). Roughly balanced bagging for imbalanced data. *Statistic. Analysis Data Mining*, 412-426.
- Kim, M.-J., & Kang, D.-K. (2012). Classifier Selection in Ensembles using Genetic Algorithm for Bankruptcy Prediction. *Expert System with Application: An International Journal*, 9308-9314.
- Korada, N., Kumar, N., & Deekshitulu, Y. (2012). Implementation of Naïve Bayesian Classifier and Ada-Boost. *International Journal of Information Sciences and Techniques*, 63-75.
- Larose, D. T. (2005). *Discovering Knowledge in Data: An Introduction to Data Mining*. New Jersey: John Wiley & Sons, Inc.
- Liang, G., & Zhang, C. (2011). Empirical Study of Bagging Predictors on Medical Data. *Proceedings of the 9-th Australasian Data Mining Conference*, 31-40.
- Maimon, O., & Rokach, L. (2010). *Data Mining and Knowledge Discovery Handbook Second Edition*. New York: Springer.
- Mardiana, T. (2011). Rancang Bangun Pada PT. Indoaja Menggunakan Model ITPOSMO. *Paradigma*, XII(02), 34-45.
- Peng, Y., & Yao, J. (2010). AdaOUBoost: Adaptive Over-sampling and Under-sampling to Boost the Concept Learning in Large Scale Imbalanced Data Sets. *Proceedings of the international conference on Multimedia information retrieval*, 111-118.
- Seiffert, C., Khoshgoftaar, T., Hulse, J., & Napolita, A. (2008). Resampling or Reweighting: A Comparison of Boosting Implementations. *20th IEEE International Conference on Tools with Artificial Intelligence*, 445-451.
- Sun, Y., Kamel, M., & Wong, A. (2007). Cost-sensitive boosting for classification of imbalanced data. *Pattern Recognition*, 3358-3378.
- Twala, B. (2010). Multiple Classifier Application to Credit Risk Assessment. *Expert System with Application*, 3326-3336.
- Wu, X., & Kumar, V. (2009). *The Top Ten Algorithms in Data Mining*. New York: CRC Press.
- Yap, B., Rani, K., Rahman, H., Fong, S., Khairudin, Z., & Abdullah, N. N. (2014). An Application of Oversampling, Undersampling, Bagging and Boosting in Handling Imbalanced Datasets. *Proceedings of the First International Conference on Advanced Data and Information Engineering*, 13-22.
- Zhang, D., Liu, W., Gong, X., & Jin, H. (2011). A Novel Improved SMOTE Resampling Algorithm Based on Fractal. *Computational Information Systems*, 2204-2211.
- Zhou, Z., & Liu, X. (2006). Training cost-sensitive neural networks with methods addressing the class imbalance problem. *IEEE Transaction*, 63-77.