

Metode Bagging Untuk *Imbalance Class* Pada Bedah Toraks Menggunakan Naïve Bayes

Fitriyani

Universitas BSI, fitriyani.fyn@bsi.ac.id

ABSTRAK - Kanker paru merupakan masalah yang terjadi di Indonesia dan merokok merupakan salah satu penyebab kanker paru. Bedah toraks dapat mengobati pasien yang terkena kanker paru, akan tetapi usia hidup pasien pasca operasi yang menjadi masalah, jika kanker paru yang diderita pasien sudah parah. *Naïve Bayes* merupakan algoritma *machine learning* yang sederhana dan cepat dalam hal klasifikasi, kinerja yang baik dan mudah dalam penerapannya. Banyak penelitian yang memprediksi hidup atau matinya pasien setelah satu tahun menjalani bedah toraks, salah satunya menggunakan algoritma *Naïve Bayes*. Dataset *thoracic surgery* merupakan data yang berisi tentang pasien yang menjalani bedah toraks dan hidup mati pasien setelah pasca operasi. Dataset bedah toraks memiliki permasalahan *imbalance class* (ketidakseimbangan kelas) karena nilai *false* lebih banyak dibandingkan nilai *true*. Dataset yang tidak seimbang dapat menyebabkan kinerja pada model yang digunakan menurun, sehingga dibutuhkan metode yang dapat menangani permasalahan tersebut. *Bagging* merupakan metode yang tepat dalam menangani permasalahan *imbalance class* dan dapat meningkatkan kinerja dari model yang digunakan. Hasil pada penelitian ini adalah prediksi menggunakan algoritma *Naïve Bayes* dan metode *bagging* untuk *imbalance class* pada dataset bedah toraks.

Kata Kunci: Kanker Paru, *Thoracic Surgery*, *Naïve Bayes*, *Bagging*

ABSTRACT – Lung cancer is a problem that occurs in Indonesia and smoking is one cause of lung cancer. Thoracic surgery can treat patients affected by lung cancer, but the life span of postoperative patients is a problem, if the lung cancer suffered by the patient has been severe. *Naïve Bayes* is a machine learning algorithm that is simple and fast in terms of classification, good performance and easy to implement. Many studies predict the lives or deaths of patients after one year of undergoing thoracic surgery, one of which uses the *Naïve Bayes* algorithm. The thoracic surgery dataset is data about patients undergoing thoracic surgery and life after death postoperatively. The thoracic thoracic dataset has an imbalance class problem (a class imbalance) because the false value is more than the true value. An unbalanced dataset can cause performance in the model used to decrease, so a method that can handle the problem is required. *Bagging* is an appropriate method of dealing with class imbalance issues and can improve the performance of the model used. The results of this study are predicted using *Naïve Bayes* algorithm and *Bagging* method for imbalance class on thoracic surgical dataset.

Keywords: Lung Cancer, *Thoracic Surgery*, *Naïve Bayes*, *Bagging*.

Naskah diterima : 16 Mei 2018, Naskah dipublikasikan : 15 September 2018

PENDAHULUAN

Tingginya angka merokok pada masyarakat akan menjadikan kanker paru sebagai salah satu masalah kesehatan di

Indonesia. Kejadian kanker paru-paru terus meningkat dan menjadi masalah kesehatan global. Kanker paru di Indonesia menduduki peringkat keempat dari semua kanker yang

sering ditemukan di rumah sakit (Supartono & Suryanto, 2012). Pengobatan pada kanker paru dapat dilakukan dengan pembedahan, radioterapi, kemoterapi, imunoterapi, hormonoterapi dan terapi gen (Indonesia, 2003). Operasi toraks adalah spesialisasi bedah yang difokuskan pada prosedur yang melibatkan dada yang lebih dikenal sebagai toraks (Humedad, 2017). Bedah toraks merupakan salah satu operasi paling umum pada pasien kanker paru-paru. Setelah tingkat kelangsungan hidup operasional sangat penting bagi dokter untuk memutuskan dimana operasi pasien dilakukan. Salah satu masalah keputusan utama dalam bedah toraks adalah pemilihan pasien yang tepat untuk operasi dengan resiko dan manfaat bagi pasien baik jangka panjang atau pendek (Koklu, Kahramanli, & Allahverdi, 2013).

Pada penelitian ini menggunakan dataset publik, dataset bedah toraks (*Thoracic Surgery*) dari UCI *Repository*. Penelitian menggunakan dataset publik sangat dianjurkan karena sebanyak 64.79% penelitian menggunakan dataset publik dan sebanyak 35.21% penelitian menggunakan dataset *private* (Wahono, 2015). Selain itu, penggunaan dataset publik dapat membuat penelitian berulang, terbantahkan dan diverifikasi (Catal & Diri, 2009).

Dataset bedah toraks memiliki permasalahan *imbalance class* (ketidakseimbangan kelas) dan hal ini dapat menurunkan kinerja dari model prediksi bedah toraks. Ketidakseimbangan kelas pada dataset bedah toraks karena nilai *false* lebih banyak dibandingkan nilai *true*. *Bagging* (*Bootstrap Aggregating*) merupakan teknik yang dapat meningkatkan klasifikasi dengan kombinasi klasifikasi secara acak pada dataset *training* dan *bagging* juga dapat mengurangi variansi dan menghindari *overfitting* (Wahono & Suryana, 2013).

Banyak penelitian yang telah dilakukan sebelumnya oleh peneliti lain. Penelitian sebelumnya mengenai bedah toraks, akan tetapi menggunakan algoritma lain seperti algoritma PART (*Partial Decision Trees*),

J48 *Decision Tree*, *OneR* (*One Rule*) (Sindhu, Prabha, Veni, & Hemalatha, 2014), *Simple Logistic Regression*, *Multilayer Preceptron* (Harun, 2015), *Fuzzy Rough Nearest Neighbor*, *Fuzzy Nearest Neighbor*, *Classification Via Clustering*, *Sequential Minimum Optimization*, *Naïve Bayes Tree* (El-hasnony, Bakry, & Saleh, 2015), *Conjunctive Rule*, *Decision Table*, *Java Repeated Incremental Pruning* (JRip), *Navajo Nation Gaming Enterprise* (NNge), *Ripple-Down Rule learner* (Ridor), *0-R classifier* (ZeroR) (Koklu et al., 2013), *Artificial Neural Network* (ANN) (Danjuma, 2015). Sedangkan penelitian yang menggunakan algoritma *Naïve Bayes* dan *Bagging*, akan tetapi menggunakan dataset yang berbeda adalah *Support Vector Machine*, *Decision Tree* dan *Naïve Bayes* (Tien Bui, Pradhan, Lofman, & Revhaug, 2012). *Naïve Bayes*, *Bagging* dan *Greedy Forward Selection* (Fitriyani & Wahono, 2015) (Riana, Hidayanto, & Fitriyani, 2017).

Algoritma *Naïve Bayes* termasuk dalam *supervised learning* dan salah satu algoritma pembelajaran tercepat yang dapat menangani sejumlah fitur atau kelas (Lee, 2015). Menurut (Khoshgoftaar, Van Hulse, & Napolitano, 2011) teknik *bagging* lebih baik dibandingkan teknik *boosting* dalam menangani data *noisy* dan *imbalance class*, sehingga pada penelitian ini diusulkan algoritma *Naïve Bayes* dan metode *bagging* untuk model penelitian bedah toraks.

Tujuan penelitian ini adalah untuk mengetahui seberapa bagus atau baik hasil prediksi *Area Under the ROC Curve* (AUC) menggunakan model *Naïve Bayes* dan model *Naïve* dengan *bagging* untuk prediksi bedah toraks.

LANDASAN TEORI

Kanker Paru

Kanker paru adalah salah satu jenis penyakit paru yang memerlukan penanganan dan tindakan yang cepat dan terarah. Penegakan diagnosis penyakit ini membutuhkan ketrampilan dan sarana yang tidak sederhana dan memerlukan pendekatan

multidisiplin kedokteran. Penyakit ini membutuhkan kerja sama yang erat dan terpadu antara ahli paru dengan ahli radiologi diagnostik, ahli patologi anatomi, ahli radiologi terapi dan ahli bedah toraks, ahli rehabilitasi medik dan ahli-ahli lainnya (Indonesia, 2003). Merokok merupakan penyebab utama dari sekitar 90% kasus kanker paru-paru pada pria dan sekitar 70% kasus pada wanita. Semakin banyak rokok yang dihisap, semakin besar resiko untuk menderita kanker paru-paru (Zulkifli, 2011).

Penelitian tentang rokok mengatakan bahwa lebih dari 63 jenis bahan yang dikandung asap rokok itu bersifat karsinogenesis. Secara epidemiologik juga terlihat kaitan kuat antara kebiasaan merokok dengan insidens kanker paru. Keterkaitan rokok dengan kasus kanker paru diperkuat dengan data bahwa risiko seorang perempuan perokok pasif akan terkena kanker paru lebih tinggi daripada mereka yang tidak terpajan kepada asap rokok (Indonesia, 2003). Pengobatan kanker paru adalah *combined modality therapy* (multi-modaliti terapi). Bedah toraks merupakan salah satu pengobatan pasien pada penderita kanker paru selain radiologi dan kemoterapi.

Imbalance Class

Imbalance Class (ketidakseimbangan kelas) memiliki dampak serius pada kinerja klasifikasi (Liu, Wu, & Zhou, 2009) sehingga dibutuhkan metode untuk dapat menangani permasalahan tersebut. Pada ketidakseimbangan kelas, kelas minoritas terhadap mayoritas sering tidak diketahui dengan baik sehingga metode *ensemble* dapat menangani hal tersebut dengan efektif seperti *bagging*, *boosting*, *random forest*, *stacking* dan lain-lain (He & Ma, 2013).

Bagging

Bagging adalah singkatan dari "Bootstrap Aggregating" yang menggunakan dua teknik utama *bagging*, yaitu *bootstrap* dan agregasi. *Bagging* mengadopsi *bootstrap sampling* untuk mengaplikasikan metode *random* kedalam

data *training* untuk menghasilkan *base learners* dengan mengurangi ketergantungan dan perbedaan. (Ying-Liu & Hua Zhou, 2013). Teknik *ensemble* merupakan teknik yang sukses untuk menangani dataset yang tidak seimbang meskipun tidak secara khusus dirancang untuk masalah data yang tidak seimbang (Laradji, Alshayeb, & Ghouti, 2015). Teknik *bagging* merupakan salah satu teknik *ensemble* dan teknik ini pada klasifikasi memisahkan data *training* ke dalam beberapa data *training* baru dengan *random sampling* dan membangun model berbasis data *training* baru (Wahono & Suryana, 2013). Algoritma *bagging* untuk klasifikasi (He & Ma, 2013).

Input: Data set $D = \{(x_i, y_i)\}_{i=1}^n$

Base learning algorithm \mathcal{L}

The number of iterations T

1. for t=1 to T do
2. $h_t = \mathcal{L}(D, \mathcal{D}_{bs})$
/* \mathcal{D}_{bs} is the bootstrap distribution */
3. end for

Output: $H(x) = \max_y \sum_{t=1}^n I(h_t(x) = y)$
/* $I(x)=1$ if x is true, and 0 otherwise */

Naïve Bayes

Bayes merupakan teknik prediksi berbasis probabilistik sederhana yang berdasar pada penerapan teorema *Bayes* (aturan *Bayes* dengan asumsi independen (ketidaktergantungan) yang kuat (naif) (Prasetyo, 2012). Formula teori *Bayes* seperti berikut ini (Sindhu et al., 2014):

$$p(C | F_1 \dots F_n) = \frac{p(C) p(F_1 \dots F_n | C)}{p(F_1 \dots F_n)}$$

Pengklasifikasi *Naïve Bayes* berpendapat bahwa nilai dari atribut pada kelas tertentu tidak bergantung (*independence*) pada nilai atribut lainnya, pendapat ini dapat disebut *class-conditional independence* sehingga perhitungannya

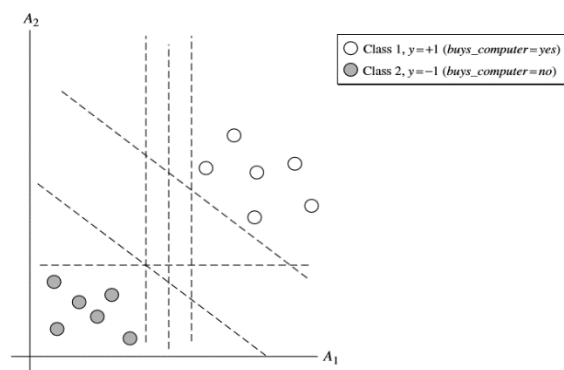
dapat dibuat lebih sederhana dan disebut “naif (*naïve*)” (Han, Kamber, & Pei, 2012). Keuntungan menggunakan algoritma *Naïve Bayes* adalah (Sindhu et al., 2014):

1. Cepat dan model sangat terukur
2. Menyeimbangkan linear dengan jumlah prediktor dan baris
3. Prosedur *Naïve Bayes* adalah paralel
4. *Naïve Bayes* dapat digunakan untuk klasifikasi biner dan multiclass.

Support Vector Machine

Metode *Support Vector Machine* (SVM) berakar dari teori pembelajaran statistik yang memberikan hasil lebih baik dibandingkan metode lain. SVM juga bekerja dengan baik pada dataset berdimensi tinggi (Prasetyo, 2012). Secara konseptual, SVM adalah mesin linear, dilengkapi dengan fitur-fitur khusus, dan berdasarkan metode minimalisasi risiko struktural dan teori belajar statistik sehingga SVM dapat memberikan kinerja generalisasi yang baik dalam masalah pengenalan pola, tanpa memasukkan masalah domain pengetahuan yang memberikan fitur unik di antara mesin belajar lainnya (Gorunescu, 2011).

Contoh SVM pada kasus dua kelas yang terpisah secara linear. Kumpulan data D sebagai $(X_1, y_1), (X_2, y_2), \dots, (X_{|D|}, y_{|D|})$, dimana X_i adalah himpunan tupel pelatihan dengan label kelas terkait y_i . Setiap y_i dapat mengambil salah satu dari dua nilai, $+1$ atau -1 ($y_i \in \{+1, -1\}$). Pada Gambar 1 menunjukkan adanya kelas *buys computer = yes* dan *buys computer = no*. Pada grafik dalam Gambar 1 bahwa data terpisah secara linear dengan garis lurus yang dapat ditarik untuk memisahkan semua tupel kelas $+1$ dan kelas -1 (Han et al., 2012).



Gambar 1. Data *training* yang terpisah secara linear dengan dua Kelas

Sumber: (Han et al., 2012)

C4.5

Algoritma C4.5 dirancang oleh J. Ross Quinlan, dinamakan demikian karena merupakan keturunan dari pendekatan ID3 untuk merangsang pohon keputusan. *Decision tree* merupakan serangkaian pertanyaan yang sistematis dan semuanya diatur sehingga setiap pertanyaan dari *query* atribut (misalnya Outlook) berdasarkan nilai dari atribut (Wu & Kumar, 2009). Contoh dataset C4.5 dapat dilihat pada Tabel 1 dan Gambar 2 merupakan *decision tree* hasil dari algoritma C4.5.

Tabel 1. Input Dataset C4.5

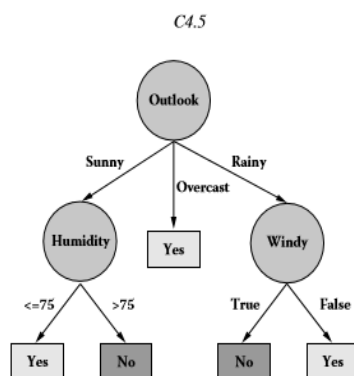
Play	Outlook	Temperature	Humadity	Wind
No	Sunny	85	85	False
No	Sunny	80	90	True
Yes	Overcast	83	78	False
Yes	Rain	70	96	False
Yes	Rain	68	80	False
No	Rain	65	70	True
Yes	Overcast	64	65	True
No	Sunny	72	95	False
Yes	Sunny	69	70	False
Yes	Rain	75	80	False
Yes	Sunny	75	70	True
Yes	Overcast	72	90	True
Yes	Overcast	81	75	False
No	Rain	71	80	True

Algoritma C4.5

Input: Dataset atribut bernilai D

1: $Tree = \{ \}$

- 2: **if** D adalah “murni” OR kriteria lainnya akan terpenuhi **then**
- 3: terminate
- 4: **end if**
- 5: **for all** atribut $a \in D$ **do**
- 6: menghitung kriteria informasi-teori jika kita membagi pada a
- 7: **end for**
- 8: a_{best} = atribut terbaik sesuai dengan kriteria yang dihitung
- 9: Tree = buat simpul keputusan yang menguji a_{best} di root
- 10: D_v = induksi sub dataset dari D berbasis a_{best}
- 11: **for all** D_v **do**
- 12: $Tree_v = C4.5(D_v)$
- 13: melampirkan $Tree_v$ untuk cabang pohon yang sesuai
- 14: **end for**
- 15: **return** Tree



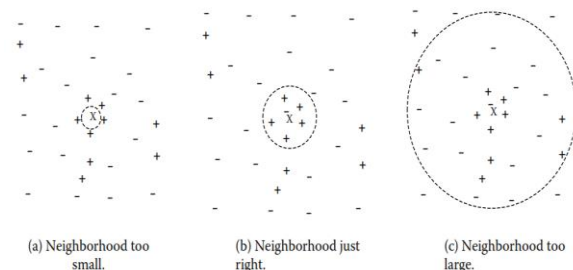
Gambar 2. *Decision Tree* induksi dari dataset C4.5

Sumber: (Wu & Kumar, 2009)

K-Nearest Neighbor

Algoritma *K-Nearest Neighbor* (K-NN) merupakan algoritma yang melakukan klasifikasi berdekatan dengan lokasi dan (jarak) suatu data dengan data yang lain (Prasetyo, 2012). K-NN termasuk kelompok *instance-based learning*. Algoritma ini juga merupakan salah satu teknik *lazy learning*. K-NN bekerja dengan mencari kelompok k objek dalam data *training* yang paling dekat (mirip) dengan obyek pada data baru atau *data testing* (Wu & Kumar, 2009).

Pada Gambar 3 merupakan klasifikasi knn dengan nilai k yang kecil, sedang sampai dengan besar. Pada algoritma K-NN, harus menentukan nilai k yang sesuai karena jika nilai k terlalu kecil hasilnya dapat menyebabkan algoritma sensitif dengan *noise*, akan tetapi jika nilai k terlalu besar akan mengakibatkan tetangga (*neighborhood*) masuk dalam kelas lain (Wu & Kumar, 2009).



Gambar 3. Klasifikasi K-NN small, medium dan large

Sumber: Wu dan Kumar 2009

Dasar algoritma K-NN

Input: D , data training, data testing, z , merupakan vektor dari nilai atribut, dan L , Merupakan kelas dari label pada objek yang digunakan

Output: $c_z \in L$, class dari z

foreach objek $y \in D$ **do** | Dihitung $d(z,y)$, jarak antara z dan y ;

end

pilih $N \subseteq D$, neighborhood dari k dekat data training dari z ;

$$c_z = \underset{v \in L}{\operatorname{argmax}} \sum_{y \in N} I(v = \operatorname{class}(c_y));$$

Dimana $I(.)$ merupakan fungsi indikator dari nilai yang dikembalikan 1 jika argumennya adalah benar dan sebaliknya adalah 0.

Dekat atau jauhnya lokasi (jarak) biasanya dihitung berdasarkan jarak Euclidean atau jarak terdekat dengan rumus sebagai berikut (Han et al., 2012):

$$d(x_1, x_2) = \sqrt{\sum_{i=1}^p (x_{2i} - x_{1i})^2}$$

Cross Validation

Cross validation digunakan ketika tujuan dari machine learning adalah prediksi dan sangat penting untuk mengevaluasi kinerja dari metode machine learning (Dua & Du, 2011).

Pada k-fold cross validation, dataset dibagi secara acak sebanyak K bagian, $X_i = i, \dots, K$. Untuk menghasilkan masing-masing bagian, salah satu bagian K sebagai validasi dan menggabungkan yang tersisa dari bagian $K - 1$ menjadi data training (Alpaydin, 2010). Seperti pada persamaan berikut ini:

$$v_1 = x_i \tau_1 = x_2 \cup x_3 \cup \dots \cup x_k$$

$$v_2 = x_2 \tau_2 = x_1 \cup x_3 \cup \dots \cup x_k$$

$$v_k = x_k \tau_k = x_1 \cup x_2 \cup \dots \cup x_{k-1}$$

Penggunaan 10-fold cross validation merupakan metode standar dalam praktek (Wahono, Suryana, & Ahmad, 2014). Umumnya penggunaan 10-fold cross validation sangat dianjurkan untuk perkiraan akurasi karena bias dan variansi yang relatif rendah (Han, Kamber, & Pei, 2012). 10-fold cross validation diterapkan pada pengukuran evaluasi dari akurasi klasifikasi, dimana dataset dibagi menjadi sepuluh bagian, salah satu dari sepuluh bagian menjadi data testing dan sembilan lainnya menjadi data training untuk membangun model matematika (Shi, Tian, Kou, Peng, & Li, 2011).

Confusion Matrix

Pengukuran kinerja model menggunakan confusion matrix, confusion matrix merupakan alat untuk menganalisa seberapa baik kinerja dari pengklasifikasi dapat mengenali tupel dari kelas yang berbeda (Han et al., 2012). Berikut adalah persamaan model confusion matrix:

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Sensitivitas = recall = \frac{TP_{rate}}{TP + FN}$$

$$Specificity = TN_{rate} = \frac{TN}{TN + FP}$$

$$FP_{rate} = \frac{FP}{FP + TN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$F - Measure = \frac{(1 + \beta^2) \times recall \times precision}{(\beta \times recall + precision)}$$

$$G - Mean = \sqrt{Sensitivitas \times Specificity}$$

Area Under the ROC Curve (AUC)

Pada kelas yang tidak seimbang karena kelas minoritas mendominasi sehingga pengukuran kinerja yang tepat menggunakan Area Under the ROC (Receiver Operating Characteristic) Curve (AUC), f-measure, dan Geometric Mean (G-Mean). Area Under ROC Curve (AUC) digunakan untuk memberikan metrik numerik single untuk dapat membandingkan kinerja dari model, nilai AUC berkisar dari 0 sampai 1 dan model yang lebih baik prediksinya adalah yang mendekati nilai 1 (Gao, Khoshgoftaar, & Wald, 2010). Berikut adalah persamaan model AUC:

$$AUC = \frac{1 + TP_{rate} - FP_{rate}}{2}$$

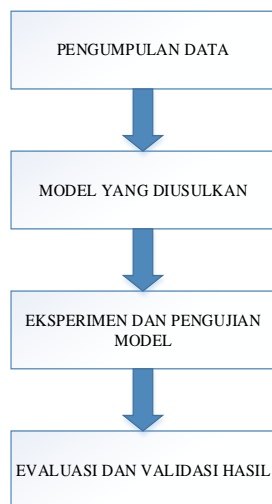
Pedoman umum yang digunakan untuk klasifikasi akurasi sebagai berikut:

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*

METODE PENELITIAN

Metode penelitian yang digunakan adalah pendekatan kuantitatif, di mana tujuan metode ini adalah untuk mengembangkan model, teori dan hipotesis yang berkaitan dengan fenomena alam. Aspek kuantitatif adalah untuk menekankan pengukuran yang mendasar karena memberikan hubungan antara observasi dan formalisasi model, teori dan hipotesis. Terdapat empat metode penelitian yang umum digunakan, salah satunya adalah eksperimen. Pada penelitian ini menggunakan metode eksperimen dengan tahapan pada Gambar 5 (Berndtsson, Hansson, Olsson, & Lundell, 2008).



Gambar 5. Tahapan Penelitian

Pengumpulan Data

Pada penelitian menggunakan data sekunder. Data sekunder merupakan data yang tidak diperoleh langsung pada obyek penelitian, melainkan dikumpulkan oleh pihak lain. Data sekunder dalam penelitian adalah data bedah toraks yang dapat di unduh pada UCI Repository atau dapat juga melalui situs <https://archive.ics.uci.edu/ml/datasets/Thoracic+Surgery+Data>. Data ini berjumlah 470 *record* dan 16 atribut sebagai variabel

prediktor dan 1 atribut sebagai variabel target seperti atribut *risk1yr*. Data operasi kanker ini berisi data pasien yang menjalani pasien kanker dari tahun 2007-2011 (Sindhu et al., 2014). Dataset ini tentang masalah klasifikasi terkait dengan harapan hidup pasca operasi pasien kanker paru-paru, dimana kematian dalam satu tahun setelah operasi. Deskripsi atribut data bedah toraks dapat dilihat pada Tabel 2.

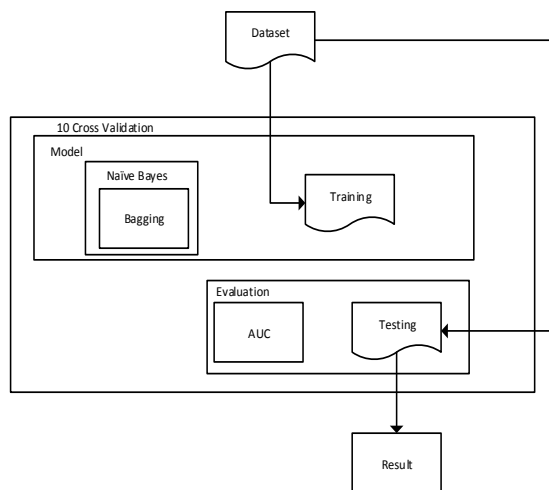
Tabel 2. Deskripsi Atribut Data Bedah Toraks

Nama Atribut	Deskripsi
PRE4	Kapasitas vital paksa-FVC (numeric)
PRE5	Sebuah volume yang telah dihembuskan pada akhir detik pertama ekspirasi paksa - FEV1 (numeric)
PRE6	Status kinerja-skala zubrood
PRE7	Rasa sakit sebelum operasi
PRE8	Hemoptisis sebelum operasi
PRE9	Dyspnoea sebelum operasi
PRE10	Batuk sebelum operasi
PRE11	Kelemahan sebelum operasi
PRE14	T di TNM klinis - ukuran tumor asli, dari OC11 (terkecil) ke OC14 (terbesar)
PRE17	DM tipe 2 - diabetes mellitus
PRE19	MI sampai dengan 6 bulan
PRE25	PAD – penyakit arteri perifer
PRE30	Merokok
PRE32	Asma
AGE	Usia di operasi
RISK1Y	1 tahun masa hidup - (nilai benar jika mati)

Model yang diusulkan

Model yang diusulkan pada penelitian ini dapat dilihat pada Gambar 6. Model yang digunakan adalah Naive Bayes yang diintegrasikan dengan metode Bagging. Dataset akan dibagi menjadi dua bagian yaitu data *training* dan *testing*. Pembagian data training dan testing dilakukan otomatis oleh *10-fold cross validation*. Kemudian data *training* akan diolah dengan metode

Bagging yang akan dilakukan iterasi atau pengulangan sebanyak 10 kali. Selanjutnya diolah dengan algoritma Naïve Bayes. Data *testing* akan menguji model dan di evaluasi menggunakan *Area Under the ROC Curve* (AUC).



Gambar 6. Model yang diusulkan

Eksperimen dan Pengujian Model

Eksperimen dan pengujian model dalam penelitian ini dilakukan dengan beberapa tahapan sebagai berikut:

1. Menyiapkan dataset untuk eksperimen
2. Mendesain arsitektur Naïve Bayes
3. Melakukan *training* dan *testing* terhadap model Naïve Bayes dan mencatat hasil kinerja dari model tersebut
4. Mendesain arsitektur Naïve Bayes dan Bagging
5. Melakukan *training* dan *testing* terhadap model Naïve Bayes dan Bagging dan mencatat hasil kinerja dari model
6. Melakukan komparasi hasil kinerja dengan model lain.

Eksperimen pada penelitian ini menggunakan laptop LENOVO G470 dengan prosesor Intel Celeron CPU B800 @ 1.50 GHz, memori (RAM) 2.00 GB dan sistem operasi Windows 8.1 Pro 32-bit. Aplikasi yang digunakan dalam penelitian ini adalah Rapidminer.

Evaluasi dan Validasi Hasil

Evaluasi dilakukan menggunakan tabel confusion matrix untuk pengukuran kinerja

pada model. Kinerja yang diukur adalah *Area Under the ROC Curve* (AUC). Validasi yang digunakan adalah *10-fold cross validation* dengan membagi data menjadi *testing* dan *training* sebanyak 10 bagian. Kinerja model akan dibandingkan dimana algoritma Naïve Bayes tanpa metode Bagging dengan algoritma Naïve Bayes dengan Bagging. Perbandingan hasil tersebut lebih akurat menggunakan pengukuran AUC.

PEMBAHASAN

Eksperimen yang dilakukan dalam penelitian ini menggunakan laptop Lenovo G470 dengan prosesor Intel Celeron CPU B800 @ 1.50 GHz, memori (RAM) 2.00 GB dan sistem operasi Windows 8.1 Pro 32-bit. Aplikasi yang digunakan Rapidminer v5.3.

Pada Tabel 3 dapat dilihat hasil eksperimen atau pengujian dari model NB (Naïve Bayes). Hasil dari confusion matrix yang muncul menunjukkan tingkat kinerja dari model yang digunakan.

Tabel 3. Confusion Matrix Model NB

	true F	true T	class precision
pred. F	359	58	86.09%
pred. T	41	12	22.64%
class recall	89.75%	17.14%	

Apabila dihitung menggunakan persamaan confusion matrix, maka hasilnya sebagai berikut:

$$Akurasi = 78.94\%$$

$$Sensitivitas = recall = TP_{rate} = 17.14\%$$

$$Specificity = TN_{rate} = 89.75\%$$

$$FP_{rate} = 86.09\%$$

$$Precision = 22.64\%$$

$$F - Measure = 0.20$$

$$G - Mean = 0.39$$

Sedangkan untuk hasil AUC (Area Under the Curve) adalah sebagai berikut:

$$AUC = 0.694$$

Tabel 4 merupakan hasil confusion matrix dari model NB+BG (Naïve Bayes dan Bagging). Tabel ini menunjukkan nilai true positive, true negative, false positive dan false negative.

Tabel 4. Confusion Matrix Model NB+BG

	true F	true T	class precision
pred. F	359	59	85.59%
pred. T	41	11	21.15%
class recall	89.75%	15.71%	

Jika dihitung menggunakan persamaan confusion matrix pada model NB+BG, maka hasilnya sebagai berikut:

$$Akurasi = 83.33\%$$

$$Sensitivitas = recall = TP_{rate} = 15.71\%$$

$$Specificity = TN_{rate} = 89.75\%$$

$$FP_{rate} = 85.89\%$$

$$Precision = 21.15\%$$

$$F - Measure = 0.18$$

$$G - Mean = 0.38$$

Sedangkan untuk hasil Area Under the Curve (AUC) pada model NB+BG hasilnya sebagai berikut:

$$AUC = 0.787$$

Tabel 5 merupakan hasil confusion matrix dari model Support Vector Machine (SVM).

Tabel 5. Confusion Matrix Model SVM

	true F	true T	class precision
pred. F	360	63	85.11%
pred. T	40	7	14.89%
class recall	90.00%	10.00%	

Kemudian dihitung berdasarkan persamaan confusion matrix sehingga hasilnya sebagai berikut:

$$Akurasi = 78.09\%$$

$$Sensitivitas = recall = TP_{rate} = 10.00\%$$

$$Specificity = TN_{rate} = 90.00\%$$

$$FP_{rate} = 85.11\%$$

$$Precision = 14.89\%$$

$$F - Measure = 0.12$$

$$G - Mean = 0.30$$

Sedangkan untuk hasil Area Under the Curve (AUC) model SVM adalah:

$$AUC = 0.517$$

Pada Tabel 6 dapat dilihat hasil confusion matrix dari model C4.5 yang merupakan hasil dari pengujian.

Tabel 6. Confusion Matrix Model C4.5

	true F	true T	class precision
pred. F	400	70	85.11%
pred. T	0	0	0.00%
class recall	100.00%	0.00%	

Apabila dihitung menggunakan persamaan confusion matrix, maka hasilnya adalah:

$$Akurasi = 85.11\%$$

$$Sensitivitas = recall = TP_{rate} = 0.00\%$$

$$Specificity = TN_{rate} = 100.00\%$$

$$FP_{rate} = 85.11\%$$

$$Precision = 0.00\%$$

$$F - Measure = 0$$

$$G - Mean = 0$$

Hasil *Area Under the Curve* (AUC) pada model C4.5 dapat dilihat sebagai berikut:

$$AUC = 0.500$$

Model yang terakhir diuji adalah model K-NN (*K-Nearest Neighbor*). Hasil *confusion matrix* pada model K-NN dapat dilihat pada Tabel 7.

Tabel 7. *Confusion Matrix* Model K-NN

	true F	true T	class precision
pred. F	357	63	85.00%
pred. T	43	7	16.93%
class recall	89.25%	10.00%	

Selanjutnya dihitung menggunakan persamaan *confusion matrix* sehingga hasilnya sebagai berikut:

$$Akurasi = 77.45\%$$

$$Sensitivitas = recall = TP_{rate} = 10.00\%$$

$$Specificity = TN_{rate} = 89.25\%$$

$$FP_{rate} = 85.00\%$$

$$Precision = 16.93\%$$

$$F - Measure = 0.13$$

$$G - Mean = 0.30$$

Hasil *Area Under the Curve* (AUC) untuk model K-NN adalah sebagai berikut:

$$AUC = 0.500$$

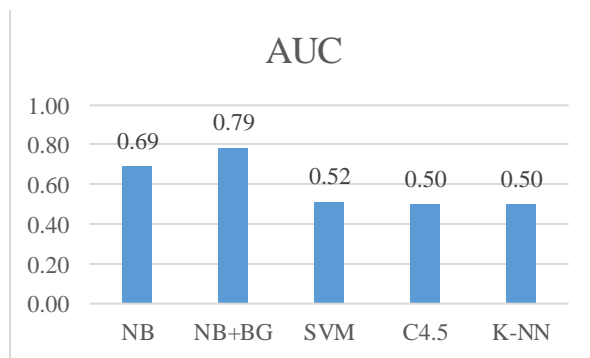
Hasil pengukuran penelitian ini dapat dilihat pada Tabel 8, dimana Tabel 8 adalah hasil *Area Under the ROC Curve* (AUC), Akurasi, sensitifitas, F-Measure dan G-Mean dari model *Naïve Bayes* (NB), model *Naïve Bayes* dan *Bagging* (NB+BG), model *Support Vector Machine* (SVM), model C4.5, dan model K-Nearest Neighbor (K-NN).

Tabel 8 menunjukkan bahwa nilai AUC tertinggi adalah model NB+BG dengan nilai 0.787, sedangkan akurasi tertinggi adalah model C4.5 dengan nilai 85.11%. Pada penelitian ini pengukuran yang digunakan adalah AUC karena untuk penelitian ketidakseimbangan kelas yang paling cocok digunakan adalah pengukuran AUC.

Tabel 8. Hasil Pengukuran Model

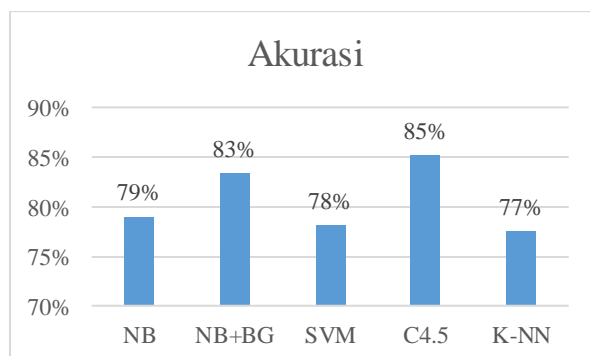
Model	AUC	Akura si	Sens itifit as	F-Measu re	G-Mea n
NB	0.694	78.94 %	17%	0.20	0.39
NB+BG	0.787	83.33 %	16%	0.18	0.38
SVM	0.517	78.09 %	10%	0.12	0.30
C4.5	0.5	85.11 %	0%	0	0
K-NN	0.5	77.45 %	10%	0.13	0.30

Gambar 7 merupakan grafik perbandingan AUC model NB, model NB+BG, model SVM, model C4.5 dan model K-NN.



Gambar 7. Grafik Perbandingan AUC

Pada Gambar 8 merupakan grafik perbandingan akurasi dari model NB, model NB+BG, model SVM, model C4.5 dan model K-NN.



Gambar 8. Grafik Perbandingan Akurasi

Hasil eksperimen pada penelitian ini dapat disimpulkan bahwa model *Naïve Bayes* dengan metode *Bagging* (NB+BG) merupakan hasil terbaik. Metode *Bagging* tidak hanya dapat menyelesaikan permasalahan ketidakseimbangan kelas (*imbalance class*), akan tetapi dapat meningkatkan hasil kinerja dari model. Model NB+BG menghasilkan nilai AUC sebesar 0.787. Selanjutnya kinerja terbaik dihasilkan oleh model *Naïve Bayes* (NB) tanpa *Bagging* dengan nilai AUC sebesar 0.694. Model *Support Vector Machine* (SVM) dengan nilai AUC sebesar 0.517, model C4.5 dengan nilai AUC sebesar 0.5 dan model K-NN mendapatkan nilai AUC sebesar 0.5.

PENUTUP

Kanker paru merupakan masalah yang terus meningkat dan masalah kesehatan global. Ada banyak solusi untuk mengobati kanker paru, salah satunya adalah bedah toraks. Bedah toraks merupakan pengobatan untuk pasien yang menderita kanker paru

Dataset bedah toraks memiliki ketidakseimbangan kelas yang dapat menurunkan kinerja dari model yang digunakan. Metode *bagging* dapat menangani permasalahan ketidakseimbangan kelas sehingga kinerja model menjadi lebih baik. Algoritma *Naïve Bayes* merupakan algoritma pembelajaran mesin yang sederhana tapi bagus dalam hasil kinerja. Hasil eksperimen pada penelitian ini algoritma *Naïve Bayes* dengan metode *Bagging* (NB+BG) memberikan hasil kinerja paling tinggi dibandingkan algoritma lain seperti *Support Vector Machine* (SVM), C4.5 (DT) dan *K-Nearest Neighbor* (K-NN). Pada hasil eksperimen terbukti bahwa metode *Bagging* tidak hanya dapat menyelesaikan permasalahan ketidakseimbangan kelas, akan tetapi meningkatkan model prediksi bedah toraks. Model NB dengan nilai AUC 0.694, model NB+BG mempunyai nilai AUC 0.787, nilai AUC untuk model SVM adalah 0.517, AUC 0.5 merupakan nilai model DT dan model K-NN mempunyai nilai 0.5. Pada penelitian ini dapat disimpulkan bahwa model NB+BG merupakan model terbaik dalam penelitian bedah toraks. Pada penelitian untuk menangani ketidakseimbangan kelas dapat dilakukan dengan menggunakan beberapa metode seperti menggunakan metode *boosting* (*adaboost*), *stacking*, *undersampling*. Metode *undersampling* sangat populer pada permasalahan ketidakseimbangan kelas (Liu et al., 2009).

DAFTAR PUSTAKA

- Alpaydin, E. (2010). *Introduction to Machine Learning*. London: Massachusetts Institute of Technology.
- Berndtsson, M., Hansson, J., Olsson, B., & Lundell, B. (2008). *Thesis Projects A Guide for Students in Computer Science and Information Systems*. London: Springer.
- Catal, C., & Diri, B. (2009). A systematic review of software fault prediction studies. *Expert Systems with Applications*, 36(4), 7346–7354. <https://doi.org/10.1016/j.eswa.2008.10.027>
- Danjuma, K. J. (2015). Performance Evaluation of Machine Learning Algorithms in Post-operative Life Expectancy in the Lung Cancer Patients. *Journal of Computer Science Issues (IJCSI)*.
- El-hasnony, I. M., Bakry, H. M. El, & Saleh, A. A. (2015). Comparative Study among Data Reduction Techniques over Classification Accuracy. *International Journal of Computer Applications*, 122(2), 9–15. <https://doi.org/10.5120/21671-4752>
- Fitriyani, & Wahono, R. S. (2015). Integrasi Bagging dan Greedy Forward Selection pada Prediksi Cacat Software dengan Menggunakan Naïve Bayes. *Journal of Software Engineering*, 1(2), 101–108.
- Gao, K., Khoshgoftaar, T., & Wald, R. (2010). Combining Feature Selection and Ensemble Learning for Software Quality Estimation, 47–52.
- Gorunescu, F. (2011). *Data mining: concepts and techniques*. Chemistry & <https://doi.org/10.1007/978-3-642-19721-5>
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques*. [https://doi.org/10.1002/1521-3773\(20010316\)40:6<9823::AID-ANIE9823>3.3.CO;2-C](https://doi.org/10.1002/1521-3773(20010316)40:6<9823::AID-ANIE9823>3.3.CO;2-C)
- Harun, A. U. (2015). Predicting Outcome of Thoracic Surgery by Data Mining Techniques. *International Journal of Advanced Research in Computer Science and Software Engineering (IJARCSSE)*, 5(1), 7–10.
- Indonesia, P. (2003). Kanker Paru. ... *Diagnosis Dan Penatalaksanaan Di Indonesia*. PDPI, (KANKER PARU PEDOMAN DIAGNOSIS & PENATALAKSANAAN DI INDONESIA), 19. Retrieved from <http://klikpdpi.com/konsensus/konsensus-kankerparu/kankerparu.pdf>
- Khoshgoftaar, T. M., Van Hulse, J., & Napolitano, A. (2011). Comparing boosting and bagging techniques with noisy and imbalanced data. *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans*, 41(3), 552–568. <https://doi.org/10.1109/TSMCA.2010.2084081>
- Koklu, M., Kahramanli, H., & Allahverdi, N. (2013). Applications of Rule Based Classification (pp. 1991–1998).
- Laradji, I. H., Alshayeb, M., & Ghouti, L. (2015). Software defect prediction using ensemble learning on selected features. *Information and Software Technology*, 58, 388–402. <https://doi.org/10.1016/j.infsof.2014.07.005>
- Lee, C.-H. (2015). A gradient approach for value weighted classification learning in naive Bayes. *Knowledge-Based Systems*, (May). <https://doi.org/10.1016/j.knosys.2015.04.020>
- Liu, X.-Y., Wu, J., & Zhou, Z.-H. (2009). Exploratory Undersampling for Class Imbalance Learning. *IEEE Transactions on Systems, Man and Cybernetics*, 39(2), 539–550. <https://doi.org/10.1109/TSMCB.2008.2007853>
- Riana, D., Hidayanto, A. N., & Fitriyani. (2017). Integration of Bagging and greedy forward selection on image pap smear classification using Naïve Bayes. *2017 5th International Conference on Cyber and IT Service Management, CITSM 2017*.

- <https://doi.org/10.1109/CITSM.2017.8089320>
- Sindhu, V., Prabha, S. A. S., Veni, S., & Hemalatha, M. (2014). Thoracic surgery analysis using data mining techniques, *5*(April), 578–586.
- Supartono, & Suryanto, A. (2012). Faktor-Faktor yang Mempengaruhi Ketahanan Hidup Satu Tahun Penderita Kanker Paru Stadium Lanjut di RSUP Dr. Kariadi Semarang. *Medical Hospitalia*, *1*(1), 20–24.
- Tien Bui, D., Pradhan, B., Lofman, O., & Revhaug, I. (2012). Landslide susceptibility assessment in vietnam using support vector machines, decision tree, and nave bayes models. *Mathematical Problems in Engineering*, 2012. <https://doi.org/10.1155/2012/974638>
- Wahono, R. S. (2015). A Systematic Literature Review of Software Defect Prediction : Research Trends , Datasets , Methods and Frameworks, *1*(1).
- Wahono, R. S., & Suryana, N. (2013). Combining particle swarm optimization based feature selection and bagging technique for software defect prediction. *International Journal of Software Engineering and Its Applications*, *7*(5), 153–166. <https://doi.org/10.14257/ijseia.2013.7.5.16>
- Wahono, R. S., Suryana, N., & Ahmad, S. (2014). Metaheuristic Optimization based Feature Selection for Software Defect Prediction. *Journal of Software*, *9*(5), 1324–1333. <https://doi.org/10.4304/jsw.9.5.1324-1333>
- Zulkifli, A. (2011). Kanker Paru. *Buku Ajar Ilmu Penyakit Dalam*, 2254–2261.

RIWAYAT PENULIS

Fitriyani. Memperoleh gelar S.T di bidang Sistem Informasi Universitas BSI dan M.Kom di bidang Ilmu Komputer STMIK Nusa Mandiri. Minat penelitian saat ini meliputi data *mining*, *software engineering*, *machine learning*, sistem pakar dan sistem informasi.