



Perbandingan Kinerja Model Klasifikasi *Decision Tree* , *Bayesian Classifier*, *Instance Base*, *Linear Function Base*, *Rule Base* pada 4 Dataset Berbeda

T. Yudi Hadiwandura

Teknik Informatika, Universitas Riau

tyudihw@lecturer.unri.ac.id

Abstrak

Ada banyak metode klasifikasi yang telah dikembangkan namun kinerjanya selalu berbeda dari satu masalah ke masalah lain. Perlu dilakukan suatu upaya untuk memilih model klasifikasi yang relatif optimal, salah satunya dengan membandingkan kinerja beberapa algoritma dari beberapa kelas model berbeda menggunakan beberapa dataset dengan karakteristik yang berbeda pula sehingga dapat memberikan gambaran tentang kelas model mana yang relatif optimal untuk dipilih sebagai tahap awal dalam memilih algoritma yang akan diterapkan. Pada makalah ini dibandingkan kinerja metode klasifikasi dari model *Decision Tree*, *Bayesian Classifier*, *Instance Base*, *Linear Function Base* dan *Rule Base* menggunakan 4 dataset yang berbeda karakteristiknya. Dari hasil pengujian diperoleh kesimpulan bahwa model *Instance Base* dan *Decision Tree* lebih handal (robust) dibanding kelas model lainnya bila diberi dataset yang mengandung noise dan missing value. Tingkat akurasi (accuracy) model *Bayesian* lebih stabil dibanding kelas model lainnya apabila diberi dataset dengan jumlah atribut yang lebih banyak. Semua kelas model mempunyai skalabilitas (scalability) yang baik dan mampu meningkatkan akurasi bila diberi jumlah record yang lebih besar. Kelas model *Linear Function* dan *Decision Tree* membutuhkan waktu training yang relatif lebih lama dibanding model lainnya, namun membutuhkan waktu testing yang relatif lebih cepat dibanding model lainnya. Secara umum kelas model *Decision Tree* lebih unggul dibanding kelas model lainnya.

Kata Kunci : *Decision Tree*, *Bayesian Classifier*, *Instance Base*, *Linear Function Base*, *Rule Base*.

Abstract

There are many classification methods that have been developed but their performance always differs from one problem to another. An effort is needed to choose a relatively optimal classification model, one of which is by comparing the performance of several algorithms from several different model classes using several datasets with different characteristics so as to provide an overview of which class models are relatively optimal to be chosen as the initial stage in choosing algorithm to be applied. In this paper compared to the performance of the classification method of the *Decision Tree* model, *Bayesian Classifier*, *Instance Base*, *Linear Function Base* and *Rule Base* using 4 datasets with different characteristics. From the test results it can be concluded that the *Instance Base* and *Decision Tree* models are more robust than the other model classes when given a dataset containing noise and missing value. The accuracy of the *Bayesian* model is more stable than other model classes when given a dataset with more attributes. All model classes have good scalability and can improve accuracy when given a larger number of records. The *Linear Function* and *Decision Tree* classes require training time that is relatively longer than other models, but requires testing time which is relatively faster than other models. In general, the *Decision Tree* model class is superior to other model classes.

Keywords: *Decision Tree*, *Bayesian Classifier*, *Instance Base*, *Linear Function Base*, *Rule Base*.

1. Pendahuluan

Pencarian informasi yang berguna dari sejumlah besar tumpukan data dapat dilakukan dengan menggunakan teknik data mining. Data mining merupakan suatu bidang ilmu yang relatif baru dengan berbagai pengaplikasian yang berbeda dan telah menjadi salah satu dari sepuluh bidang ilmu teratas yang mempengaruhi perkembangan teknologi. (Ghorbani & Ghousi, 2019).

Pemanfaatan data mining telah banyak dilakukan diberbagai bidang, misalnya penerapan algoritma K-Nearest Neighbor (KNN) untuk mendiagnosa penyakit Hepatitis (Mahmud & Ilyas, 2017) atau penerapan metode Support Vector Machine (SVM) untuk memprediksi siswa yang berpotensi Drop Out (Fiska, 2017). Berbagai macam metode data mining telah dikembangkan dan diterapkan di banyak big data, namun cakupan dan fokus aplikasinya tidaklah konsisten, dimana kinerja dari metode tersebut selalu berbeda dari satu masalah ke masalah lain. Peneliti harus mencari metode data mining yang relatif optimal untuk dapat segera menyelesaikan masalah yang dihadapi (Zhang et al., 2017).

Ada beberapa model klasifikasi yang sangat populer yaitu *Decision/Classification Trees*, *Bayesian Classifiers*/*Naïve Bayes Classifiers*, *Neural Networks*, *Statistical Analysis*, *Genetic Algorithms*, *Rough Sets*, *K-Nearest Neighbor Classifier*, *Rule-based Methods*, *Memory Based Reasoning*, *Support Vector Machines* (Gorunescu, 2011).

Penelitian-penelitian terkait perbandingan algoritma klasifikasi telah banyak dilakukan, seperti Marcos dan Utomo (2015) yang membandingkan algoritma *Naïve Bayes* dan *C4.5* untuk melakukan klasifikasi penyakit diabetes dengan 9 atribut dan 768 records. Hasil penelitian ini menunjukkan algoritma *Naïve Bayes* dengan akurasi 77.47% lebih baik dibanding *C4.5* dengan akurasi 74.87%.

Hal yang sama juga dilakukan oleh Fatmawati (2016) dan algoritma *Naïve Bayes* lebih baik dibanding *C4.5* dengan akurasi 75.13% dan 73.30%. Kedua penelitian tersebut menggunakan dataset yang sama. Akurasi dari hasil penelitian Marcos dan Utomo lebih baik dari hasil penelitian Fatmawati karena sudah melakukan seleksi fitur sehingga atribut yang digunakan lebih sedikit.

Fitri (2014) juga melakukan perbandingan kinerja algoritma *Naïve Bayesian*, *Lazy-IBK*, *Zero-R* dan *Decision Tree-J48* menggunakan 2 dataset berbeda. Dataset Ecoli dengan 8 atribut dan 336 record dan dataset Yeast dengan 9 atribut dan 1.484 record. Hasil pengujian menggunakan *cross validation* menunjukkan bahwa kinerja model algoritma klasifikasi *Naïve Bayesian* lebih baik dibanding yang lain dengan akurasi 85.12 % untuk dataset Ecoli dan 57.62 % untuk dataset Yeast. Dari hasil penelitian ini terlihat bahwa

kinerja algoritma klasifikasi *Naïve Bayesian* semakin menurun saat jumlah record dalam dataset semakin besar.

Menarianti (2015) menguji beberapa algoritma klasifikasi *Logistic Regression*, *Naive Bayes*, *Decision Tree*, *Neural Network*, *Support Vector Machine*, *K-Nearest Neighbor* dan *Discriminant Analysis* untuk memprediksi resiko pemberian kredit dengan dataset yang memiliki 6 atribut dan 588 record. Penelitian ini menunjukkan algoritma *Logistic Regression* memiliki akurasi tertinggi sebesar 87,41% dibanding algoritma yang lainnya.

Yusa et al., (2016) melakukan perbandingan kinerja algoritma klasifikasi *Decision Tree (C4.5)*, *k-Nearest Neighbor (k-NN)*, dan *Naive Bayes* menggunakan dataset Readmisi Pasien Diabetes dengan 47 atribut dan 49.735 records. Hasil penelitian menunjukkan bahwa kinerja model algoritma klasifikasi *Naive Bayes* adalah yang terbaik dengan nilai akurasi sebesar 57,52%, MAE sebesar 0,512, dan kappa statistic sebesar 0,182.

Hastuti (2012) membandingkan kinerja algoritma *Logistic regression*, *C.45*, *Naïve bayes* dan *Neural network* menggunakan dataset mahasiswa dengan 22 atribut dan 3.861 record, menunjukkan bahwa algoritma klasifikasi *C.45* memiliki kinerja terbaik dengan akurasi 95,29%

Permasalahan yang cukup rumit dalam data mining adalah memilih model yang optimal. Ada 2 cara yang dapat dilakukan, yaitu memilih berdasarkan kelas model tertentu atau memilih secara bebas sesuai keinginan (Gorunescu, 2011). Dari beberapa penelitian yang telah dibahas sebelumnya, pada umumnya model dipilih secara bebas lalu dibandingkan kinerjanya.

Dalam penelitian ini akan dicoba untuk mencari tahu kelas model mana yang relatif optimal untuk dipilih berdasarkan karakteristik dataset yang digunakan. Dari hasil perbandingan ini diharapkan dapat menjawab keraguan dan ketidak pastian dalam menentukan kelas model klasifikasi apa yang relatif optimal untuk diterapkan pada suatu kasus berdasarkan karakteristik dataset yang digunakan.

Makalah ini disusun dalam beberapa bagian. Pada bagian “Landasan Teori” akan dijelaskan beberapa kelas model klasifikasi beserta konsep dasar algoritmanya dan beberapa elemen dari pengukuran kinerja algoritma klasifikasi. Selanjutnya bagian “Hasil dan Pembahasan” akan menjelaskan beberapa tahapan yang dilakukan dan hasil analisis dari pengukuran-pengukuran yang dilakukan. Dan pada bagian “Kesimpulan” akan disampaikan beberapa hal mengenai generalisasi dari bagian-bagian sebelumnya.

2. Landasan Teori

Data mining adalah proses pencarian pola tertentu yang tersembunyi dalam kumpulan data yang besar

untuk mendapatkan suatu pengetahuan baru yang bermanfaat (Han, Kamber, & Pei, 2011).

Pada dasarnya metode data mining di kelompokkan dalam 2 (dua) kategori: *Predictive* dan *Descriptive*. Kategori *Predictive* menggunakan variabel yang sudah ada untuk memprediksi suatu nilai dari variabel lain, sedangkan kategori *Descriptive* mengungkap pola tersembunyi dari data sehingga mudah dipahami oleh pengguna (Gorunescu, 2011).

Kategori *Predictive techniques* dapat dibagi dalam 2 (dua) kelompok utama, operasi *classification* atau *discrimination* dan operasi *prediction* atau *regression*. Klasifikasi (*classification*) merupakan operasi yang menempatkan setiap individu dari suatu populasi kedalam salah satu dari kelas tertentu berdasarkan karakteristik dari individu yang disebut sebagai *independent variables* (Stéphane, 2011)

Proses klasifikasi pada dasarnya terdiri dari 2 tahapan, yaitu tahap membentuk model atau *learning process* dan tahap pengklasifikasian dengan memprediksi kelas berdasarkan data yang diberikan atau disebut sebagai *classification process* (Han et al., 2011).

2.1. Decision Tree

Model *Decision Tree* dibentuk menyerupai struktur *flowchart*, dimana setiap simpul yang bukan simpul daun merupakan atribut pengujian, setiap cabang mewakili *output* dari pengujian, dan setiap simpul daun (*terminal node*) menentukan label *class*. Simpul paling atas dari sebuah pohon adalah node akar (Han et al., 2011).

Decision Tree membedakan antara suatu kelas dengan kelas lainnya berdasarkan tingkat kemurnian kelas (*impurity*) pada suatu simpul. Alat ukur kemurnian kelas (*impurity*) yang umum digunakan adalah *GINI Index*, *Entropy*, *Misclassification measure*, *Chi-square measure*, *G-square measure*. Beberapa algoritma yang termasuk kategori *Decision Tree* antara lain *ID3*, *C4.5*, *C5.0*, *CART*, *CHAID*, dan lain-lain (Gorunescu, 2011).

2.2. Bayesian Classifier

Model klasifikasi Bayesian ini dibangun berdasarkan teorema Bayes. Pengklasifikasian suatu variabel input kedalam suatu *class* target dilakukan dengan membandingkan probabilitas bersyarat setiap *class* target bila diberikan suatu variabel input. *Class* dengan probabilitas terbesar dipilih sebagai *class* target dari variabel input. Beberapa algoritma yang termasuk kategori klasifikasi Bayesian antara lain *Naïve Bayes* dan *Bayesian Belief Networks* (Han et al., 2011).

Algoritma *Naïve Bayes* mengasumsikan bahwa semua atribut adalah variabel independen. Asumsi ini disebut *class-conditional independence* sehingga

komputasi algoritma ini menjadi sangat sederhana. Inilah yang menyebabkan model ini dinamakan dengan "*naïve*".

2.3. Rule-Based Classification

Model klasifikasi *Rule-Based Classifiers* dipresentasikan sebagai kumpulan aturan (*rule*) IF-THEN dengan bagian "IF" (sisi kiri) dari aturan disebut *rule antecedent* dan pada bagian "THEN" (sisi kanan) dari aturan disebut *rule consequent*. Bagian kondisi (*rule antecedent*) dapat berisi satu atau lebih atribut uji dan menggunakan operator *boolean* "AND" sebagai pengait antar kondisi. Bagian *rule consequent* merupakan kelas target atau kelas prediksi. Pengklasifikasian kelas dilakukan berdasarkan *rule consequent* (Han et al., 2011).

Proses pembentukan aturan dapat dilakukan dengan 2 (dua) pendekatan. Pendekatan *Direct method*, mengekstrak aturan langsung dari data yang ada, misalnya pada algoritma *RIPPER*, *CN2*, *Holte's IR*, *Boolean reasoning* dan lain-lain. Pendekatan *Indirect method*, mengekstrak aturan menggunakan model klasifikasi lain seperti *Decision Trees*, *Neural Networks*, dan lain-lain (Gorunescu, 2011).

Jika ada lebih dari satu aturan yang memenuhi kondisi (*satisfied*) maka dipilih aturan dengan prioritas tertinggi. Penentuan prioritas dapat dilakukan dengan 2 (dua) skema. Skema *Size ordering* menentukan prioritas berdasarkan jumlah atribut bernilai true pada bagian kondisi dari aturan. Skema *Rule ordering* menentukan prioritas berdasarkan urutan mayoritas kelas target atau berdasarkan hasil pengukuran rule seperti akurasi (*accuracy*), jangkauan (*coverage*) dan jumlah atribut (*size*). Jangkauan (*coverage*) dihitung berdasar persentase jumlah tupel yang memenuhi semua kondisi dari aturan. Akurasi dihitung berdasarkan persentase jumlah tupel yang diklasifikasikan dengan benar menggunakan aturan tersebut (Han et al., 2011).

2.4. Instance-Based Classification

Model klasifikasi *Instance Base* disebut juga model *Lazy learner*, karena model ini hanya menyimpan data training (*instance*), lalu menunggu datangnya variabel input baru. Klasifikasi dilakukan berdasarkan kelas dalam data training yang variabel-variabelnya mirip dengan variabel input baru. Contoh algoritma yang termasuk kategori ini adalah *K-Nearest-Neighbor (KNN)* dan *Case-Based Reasoning (CBR)* (Han et al., 2011).

Algoritma KNN cukup populer dalam aplikasi data mining. Pada algoritma KNN setiap tupel direpresentasikan sebagai sebuah titik pada ruang *n*-dimensi, dimana *n* adalah jumlah atribut tupel. Untuk sebuah tupel baru, klasifikasi ditentukan berdasarkan

kelas tetangga terbanyak dari sejumlah k tupel tetangga dalam data training, dimana k adalah jumlah tetangga terdekat dari tupel baru. Tetangga dekat dihitung berdasarkan jarak antara titik tupel baru dengan titik tupel data training (misalnya: *Euclidean distance metric*).

2.5. Linear Function Classification

Pada model klasifikasi *Linear function-base*, pemodelan dibuat berdasarkan fungsi-fungsi regresi yang akan menghasilkan suatu nilai tertentu dan menggunakan suatu fungsi aktivasi untuk memetakan nilai tersebut kedalam suatu kelas tertentu. Algoritma yang termasuk kategori ini misalnya *Logistic Regression, Neural Network, Winnow* dan lain-lain (Witten, Frank, & Hall, 2011).

Pada algoritma *Logistic Regression*, independen variabel didefinisikan sebagai persamaan regresi linear $y = \beta_0 + \beta_1 x_1$. Algoritma *Logistic Regression* memprediksi probabilitas keanggotaan variable independen dalam suatu kelas menggunakan model fungsi logit transform atau invers logit dengan persamaan: $\hat{p} = e^{\beta_0 + \beta_1 x_1} / 1 + e^{\beta_0 + \beta_1 x_1}$.

Algoritma *Neural Network* pada dasarnya merupakan kumpulan pasangan *input-output* yang saling terkoneksi. Setiap koneksi mempunyai bobot yang terus dikoreksi selama proses belajar sampai ditemukannya hasil dengan error terkecil. Klasifikasi dilakukan berdasarkan fungsi aktivasi. Tiap pasangan input-output ini disebut *perceptron* dan direpresentasikan dalam bentuk regresi linear $y = \beta_0 + \beta_1 x_1$. Ada banyak sekali jenis *Neural Networks* dan algoritma *Neural Networks*, namun yang paling populer adalah algoritma *Neural Network Backpropagation* (Han et al., 2011)

Support Vector Machines, merupakan campuran dari *linear model* dan *instance-based model*. *Support vector machines* memilih sejumlah tupel sebagai batas pemisah antar kelas yang disebut *support vectors* dari tiap kelas dan membuat sebuah fungsi *linear discriminant* yang dapat memisahkan kelas sejauh mungkin. (Witten et al., 2011)

2.6. Pengukuran Kinerja Model Klasifikasi

Beberapa elemen yang dapat digunakan untuk mengukur kinerja model klasifikasi antara lain (Gorunescu, 2011) :

- a) *Predictive accuracy*: adalah ukuran kemampuan model mengklasifikasikan objek baru secara benar;
- b) *Speed*: adalah ukuran seberapa cepat model dapat memproses data;
- c) *Robustness*: adalah ukuran kemampuan memprediksi secara akurat meskipun terdapat data "noise" dan "missing value";

- d) *Scalability*: adalah ukuran kemampuan model dalam menangani peningkatan jumlah data dan dalam berbagai bidang;
- e) *Interpretability*: adalah ukuran seberapa mudah model dapat dipahami dan diinterpretasikan;
- f) *Simplicity*: adalah ukuran seberapa efektif dan sederhana suatu model (*compactness*).

Pengukuran *Predictive accuracy* suatu model klasifikasi dilakukan berdasarkan pada perhitungan jumlah objek yang dapat di prediksi dengan benar dan jumlah objek yang tidak dapat diprediksi dengan benar. Perhitungan ini di tabulasikan kedalam tabel *Confusion Matrix* sebagaimana disajikan pada Tabel 1 berikut (Gorunescu, 2011).

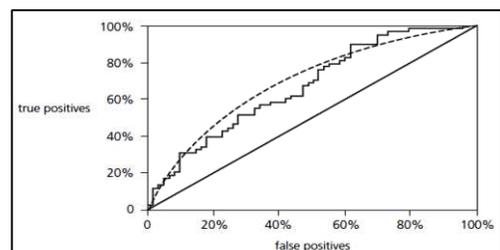
Tabel 1. Confusion matrix

Klasifikasi		Prediksi	
		Kelas=ya	Kelas=tidak
Aktual	Kelas=ya	TP	FN
	Kelas=tidak	FP	TN

dimana:

- TP = *True Positive* (jumlah kelas 'ya' yang diprediksi sebagai 'ya')
- TN = *True Negative* (jumlah kelas 'tidak' yang diprediksi sebagai 'tidak')
- FP = *False Positive* (jumlah kelas 'tidak' yang diprediksi sebagai 'ya')
- FN = *False Negative* (jumlah kelas 'ya' yang diprediksi sebagai 'tidak')
- $Accuracy = (TP+TN) / (TP+TN+FP+FN) \dots\dots\dots(1)$
- $Error\ rate = (FP+FN) / (TP+TN+FP+FN) \dots\dots\dots(2)$
- $TP\text{-rate}/Sensitivity/Recall = TP / (TP+FN) \dots\dots\dots(3)$
- $TN\text{-rate}/Specivicity = TN / (TN+FP) \dots\dots\dots(4)$
- $Precision = TP / (TP+FP) \dots\dots\dots(5)$

Perbandingan kinerja model klasifikasi juga dapat dilakukan secara visual menggunakan grafik kurva ROC (*Receiver Operating Characteristic*). Kurva ROC adalah grafik 2 dimensi dengan *TP-rate* sebagai *Y-axis* dan *FP-rate (1 - TN-rate)* sebagai *X-axis* (Gorunescu, 2011).



Gambar 1. Contoh kurva ROC (Witten et al., 2011)

Pada kurva ROC terdapat garis diagonal yang dapat menunjukkan baik atau buruknya hasil

klasifikasi suatu model. Titik yang berada diatas garis diagonal menunjukkan hasil yang baik dan titik yang berada dibawah menunjukkan hasil yang buruk. Suatu titik pada kurva ROC dikatakan lebih baik dari titik lain apabila mempunyai *TP-rate* yang lebih tinggi dan *FP-rate* yang lebih rendah.

Cara lain untuk membandingkan kinerja model klasifikasi adalah dengan menghitung luas area dibawah kuva ROC atau *Area Under The Curve* (AUC). Semakin besar nilai AUC semakin baik pula kinerja model klasifikasi. Tingkat akurasi dari suatu model klasifikasi dapat ditentukan berdasarkan kriteria nilai AUC sebagai berikut (Gorunescu, 2011):

- 0.90 - 1.00 = sempurna (*excellent classification*);
- 0.80 - 0.90 = baik (*good classification*);
- 0.70 - 0.80 = adil (*fair classification*);
- 0.60 - 0.70 = buruk (*poor classification*);
- 0.50 - 0.60 = gagal (*failure*)

3. Hasil dan Pembahasan

Sebagaimana tujuan dari penelitian ini adalah untuk mengetahui kelas model klasifikasi yang relatif optimal pada suatu dataset dengan karakteristik tertentu, maka dalam penelitian ini dilakukan pengujian dengan membandingkan kinerja algoritma dari beberapa kelas model yang berbeda dan menggunakan beberapa dataset dengan karakteristik yang berbeda sebagaimana disajikan pada Tabel 2.

Pengujian dilakukan menggunakan aplikasi WEKA (*Waikato Environment for Knowledge Analysis*) versi 3.8.3 yang dijalankan diatas platform Ms.Windows 8.1 Enterprise 64bit dengan spesifikasi CPU intel Pentium 2,8GHz dan 6GB RAM (3.5GB *free memory*).

Pada penelitian ini digunakan beberapa algoritma yang dianggap dapat mewakili masing-masing kelas model, yaitu:

- a) Algoritma *C4.5* (mewakili kelas model *Decision Tree*);
- b) Algoritma *Naïve Bayes* (mewakili kelas model *Bayesian Classifier*);
- c) Algoritma *ZeroR* (mewakili kelas model *Rule Based*);
- d) Algoritma *K-Nearest-Neighbor* (mewakili kelas model *Instance Based*);
- e) Algoritma *Logistic Regression* (mewakili kelas model *Linear Function Based*).

Masing-masing algoritma diukur *Predictive accuracy* nya menggunakan *Confusion Matrix*. Selain akurasi akan diukur juga *Sensitivity/Recall*, *Specivicity* dan *Precision* masing-masing algoritma. Untuk pengukuran kecepatan model memproses data (*Speed*) dihitung berdasarkan *CPU time training* dan *CPU time testing*.

Dataset yang digunakan dalam penelitian ini dibatasi hanya untuk masalah klasifikasi 2 kelas

(*binary classification*) dan diambil dari dataset standar aplikasi WEKA atau juga bisa diperoleh dari *UCI Machine Learning Repository* (<https://archive.ics.uci.edu/ml/datasets.html>). Dataset dipilih berdasarkan beberapa karakteristik yang dapat mewakili ukuran-ukuran kinerja yang akan diuji, sebagai berikut:

- a) Untuk mengukur kehandalan (*robustness*), dibandingkan menggunakan dataset dengan missing value dan tanpa missing value;
- b) Untuk mengukur skalabilitas (*scalability*), dibandingkan menggunakan dataset dengan sedikit record dan dengan yang lebih banyak record;
- c) Untuk mengukur kecepatan proses (*speed*), dibandingkan menggunakan dataset dengan sedikit atribut dan dengan yang lebih banyak atribut. Diasumsikan jumlah atribut dapat menggambarkan kompleksitas dari masalah.
- d) Untuk ukuran kemudahan (*interpretability*) dan kesederhanaan (*simplicity*) penerapan, dalam penelitian ini diasumsikan sama untuk semua algoritma yang dibandingkan sehingga tidak diukur.

Tabel 2. Dataset yang digunakan

Dataset	Record	Atribut	Missing (y/n)
Pima diabetes	768	8	n
Breast-cancer	286	9	y
German credit	1.000	20	n
Adult (cencus)	48.842	14	y

Untuk mengurangi efek bias dari hasil pengukuran digunakan teknik validasi *10-fold cross validation*. Untuk mengantisipasi efek dari pembagian *data training* dan *data testing* secara random dalam prosedur *10-fold cross validation*, proses validasi diulangi sebanyak 10 kali sehingga setiap algoritma akan dijalankan sebanyak 100 kali dengan kombinasi *data training* dan *data testing* yang berbeda lalu diambil rata-rata pengukurannya. Ini adalah prosedur standar dalam mengukur akurasi (Witten et al., 2011).

Pengujian dilakukan pada masing-masing dataset dan dihitung komponen-komponen pengukuran untuk tiap algoritma. Hasil pengujian pada *Pima diabetes* dataset disajikan pada Tabel 3, hasil pengujian pada *Breast cancer* dataset disajikan pada Tabel 4, hasil pengujian pada *German credit* dataset disajikan pada Tabel 5 dan hasil pengujian pada *Adult* dataset disajikan pada Tabel 6.

Pada Tabel 3 dapat dilihat bahwa untuk dataset *Pima diabetes* dengan karakteristik sedikit record, sedikit atribut dan tanpa *missing value*, tingkat akurasi algoritma *Logistic Regression* dari kelas model *Linear Function* lebih unggul dibanding algoritma dari kelas model lainnya dengan akurasi sebesar 77,47%.

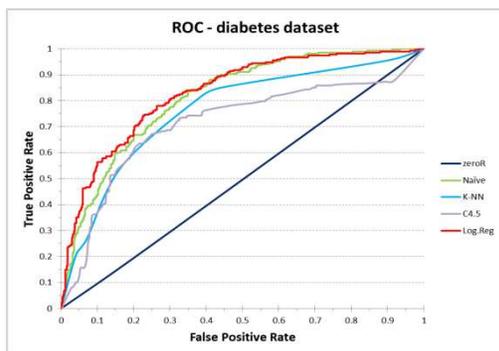
Tabel 3. Pengujian pada dataset Pima diabetes

Jenis pengukuran	ZeroR		Naïve Bayes		Logistic Regression		K-NN		C4.5	
	Mean	Std.Dev	Mean	Std.Dev	Mean	Std.Dev	Mean	Std.Dev	Mean	Std.Dev
Accuracy	65.110	0.343	75.750	5.320	77.470	4.386	73.860	4.550	74.490	5.267
Error rate	34.890	0.343	24.250	5.320	22.530	4.386	26.140	4.550	25.510	5.267
TP-rate/ Sensitivity/ Recall	1.000	0.000	0.840	0.059	0.880	0.048	0.830	0.052	0.820	0.073
TN-rate/ Specivicity	0.000	0.000	0.600	0.097	0.570	0.093	0.560	0.093	0.600	0.111
Precision	0.650	0.003	0.800	0.043	0.790	0.036	0.780	0.038	0.800	0.045
AUC	0.500	0.000	0.820	0.051	0.830	0.049	0.760	0.055	0.750	0.070
CPU-Time training	0.001	0.004	0.001	0.004	0.012	0.008	0.000	0.002	0.006	0.008
CPU-Time testing	0.000	0.002	0.001	0.003	0.000	0.002	0.006	0.009	0.000	0.003

Untuk kasus-kasus klasifikasi bidang kesehatan, selain akurasi juga penting untuk meminimalkan resiko, dimana seseorang diprediksi sakit padahal sehat jauh lebih aman daripada diprediksi sehat padahal sakit. Oleh karena itu perlu membandingkan tingkat *precision* suatu model klasifikasi.

Dari Tabel 3 dapat dilihat bahwa meskipun tingkat *precision* algoritma *Logistic Regression* berada dibawah *Naïve Bayes* dan *C4.5* dengan selisih hanya 0.01(1%), secara umum dapat dikatakan bahwa algoritma *Logistic Regression* lebih unggul dibanding kelas model lainnya.

Pada Gambar 2 juga terlihat bagaimana kurva ROC dari *Logistic Regression* berada diatas kurva ROC kelas model lainnya. Jika dilihat dari nilai AUC nya, algoritma *Logistic Regression* dan *Naïve Bayes* termasuk model klasifikasi yang baik atau *good classification*, algoritma *K-NN* dan *C4.5* termasuk model klasifikasi yang adil (tidak baik dan tidak buruk) atau *fair classification*, sedangkan algoritma *ZeroR* merupakan model klasifikasi yang gagal atau *failure*. Dengan demikian untuk kasus ini algoritma *Logistic Regression* lebih optimal dibanding algoritma dari kelas model lainnya.



Gambar 2. Kurva ROC pada diabetes dataset

Untuk dataset *Breast cancer* dengan karakteristik sedikit record, sedikit atribut dan dengan *missing value*, berdasarkan tingkat akurasi sebagaimana terlihat pada Tabel 4, algoritma *C4.5* dari kelas model *Decision Tree* lebih unggul dibanding algoritma dari

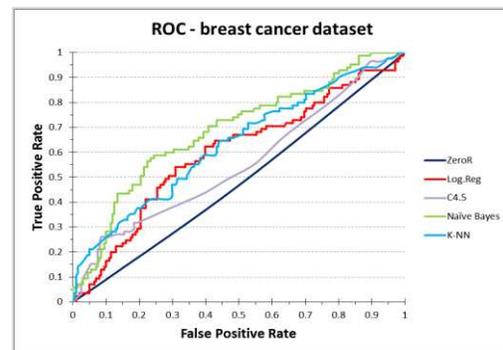
kelas model lainnya dengan akurasi sebesar 74,28%. Pada kasus ini algoritma *Logistic Regression* justru menjadi yang terburuk dengan akurasi hanya sebesar 67,77%. Sebagaimana halnya pada kasus diabetes, pada perbandingan tingkat *precision* algoritma *Naïve Bayes* lebih unggul dari model lainnya dengan *precision* sebesar 78,20%.

Pada Gambar 3 juga terlihat bagaimana kurva ROC dari *Naïve Bayes* berada diatas kurva ROC kelas model lainnya.

Tabel 4. Pengujian pada dataset Breast cancer

Jenis pengukuran	ZeroR		Naïve Bayes		Logistic Regression		K-NN		C4.5	
	Mean	Std.Dev	Mean	Std.Dev	Mean	Std.Dev	Mean	Std.Dev	Mean	Std.Dev
Accuracy	70.296	1.366	72.697	7.737	67.771	6.918	73.126	5.540	74.281	6.051
Error rate	29.704	1.366	27.303	7.737	32.229	6.918	26.874	5.540	25.719	6.051
TP-rate/ Sensitivity/ Recall	1.000	0.000	0.851	0.085	0.833	0.077	0.944	0.056	0.947	0.060
TN-rate/ Specivicity	0.000	0.000	0.434	0.163	0.309	0.149	0.229	0.121	0.260	0.141
Precision	0.703	0.014	0.782	0.054	0.742	0.047	0.744	0.036	0.753	0.042
AUC	0.500	0.000	0.700	0.108	0.630	0.112	0.650	0.106	0.610	0.087
CPU-Time training	0.000	0.002	0.000	0.002	0.016	0.004	0.000	0.002	0.001	0.003
CPU-Time testing	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.004	0.000	0.000

Berdasarkan pengukuran nilai AUC, algoritma *Naïve Bayes* juga lebih unggul dibanding model lainnya dan termasuk model klasifikasi yang adil (tidak baik dan tidak buruk) atau *fair classification* sedangkan kelas model lainnya termasuk kategori klasifikasi yang buruk atau *poor classification*. Sehingga secara umum untuk kasus ini algoritma *Naïve Bayes* lebih optimal dibanding model lainnya meskipun tingkat akurasi-nya berada dibawah algoritma *C4.5*.



Gambar 3. Kurva ROC pada breast cancer dataset

Jika dibandingkan tingkat akurasi model berdasarkan Tabel 3 dan Tabel 4, dapat terlihat bahwa pada algoritma *Logistic Regression* terjadi penurunan akurasi hingga mencapai 10%, sehingga algoritma *Logistic Regression* lebih rentan terhadap keberadaan *noise* dan *missing value* dibanding model lainnya. Pada algoritma *Naïve Bayes* terjadi penurunan akurasi

sebesar 3,05% dan pada algoritma *K-NN* dan *C4.5* terlihat lebih stabil. Sedangkan pada algoritma *ZeroR* terlihat kenaikan akurasi sebesar 5,18%, namun demikian sebagaimana diketahui bahwa algoritma *ZeroR* memprediksi berdasarkan kelas mayoritas maka akurasi -nya sangat bergantung pada persentase dari kelas mayoritas. Sehingga pada penelitian ini algoritma *ZeroR* hanya dijadikan sebagai *baseline* saja.

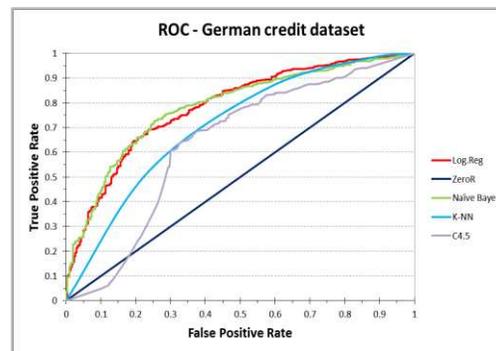
Untuk dataset *German credit* dengan karakteristik lebih banyak record, lebih banyak atribut dan tanpa *missing value*, hasil pengukuran sebagaimana terlihat pada Tabel 5 menunjukkan bahwa tingkat akurasi algoritma *Logistic Regression* kembali lebih unggul dibanding model lainnya yaitu sebesar 75,23% dan hanya berbeda sedikit dengan algoritma *Naïve Bayes* yaitu sebesar 75,16%.

Tabel 5. Pengujian pada dataset German credit

Jenis pengukuran	ZeroR		Naïve Bayes		Logistic Regression		K-NN		C4.5	
	Mean	Std.Dev	Mean	Std.Dev	Mean	Std.Dev	Mean	Std.Dev	Mean	Std.Dev
Accuracy	70.000	0.000	75.160	3.484	75.230	3.530	72.210	3.255	71.250	3.170
Error rate	30.000	0.000	24.840	3.484	24.770	3.530	27.790	3.255	28.750	3.170
TP-rate/ Sensitivity/ Recall	1.000	0.000	0.864	0.040	0.866	0.037	0.861	0.040	0.847	0.036
TN-rate/ Specivicity	0.000	0.000	0.490	0.084	0.487	0.083	0.398	0.084	0.398	0.085
Precision	0.700	0.000	0.799	0.027	0.798	0.027	0.770	0.024	0.767	0.025
AUC	0.500	0.000	0.790	0.043	0.780	0.044	0.690	0.048	0.650	0.062
CPU-Time training	0.000	0.000	0.001	0.004	0.068	0.008	0.000	0.002	0.008	0.008
CPU-Time testing	0.000	0.000	0.001	0.004	0.001	0.003	0.013	0.006	0.000	0.000

Seperti halnya pada bidang kesehatan, untuk kasus ini juga perlu diperhitungkan resiko kesalahan prediksi dimana memprediksi konsumen buruk (*bad*) padahal baik (*good*) jauh lebih aman daripada memprediksi konsumen baik (*good*) padahal buruk (*bad*). Dengan demikian nilai *TP-rate/sensitifity/recall* lebih diperhitungkan daripada nilai akurasi. Berdasarkan Table 5 terlihat bahwa nilai *TP-rate/sensitifity/recall* algoritma *Logistic Regression* juga lebih unggul dibanding model lainnya yaitu sebesar 86,60% dan hanya berbeda sedikit dengan algoritma *Naïve Bayes* yaitu sebesar 86,64% dan algoritma *K-NN* yaitu sebesar 86,61%.

Pada Gambar 4 terlihat bagaimana kurva ROC dari *Logistic Regression* dan *Naïve Bayes* berada diatas kurva ROC kelas model lainnya. Agak sulit menentukan mana yang lebih unggul antara *Logistic Regression* dan *Naïve Bayes* berdasarkan kurva ROC.



Gambar 4. Kurva ROC pada German credit dataset

Dari hasil pengukuran nilai AUC, algoritma *Naïve Bayes* lebih unggul dari pada algoritma *Logistic Regression* dan algoritma kelas model lainnya. Hasil pengukuran AUC menunjukkan bahwa algoritma *Naïve Bayes* dan *Logistic Regression* termasuk model klasifikasi yang adil (tidak baik dan tidak buruk) atau *fair classification*, sedangkan algoritma *ZeroR*, *K-NN* dan *C4.5* merupakan model klasifikasi yang buruk atau *poor classification*. Dengan demikian secara umum untuk kasus ini algoritma *Naïve Bayes* lebih optimal dibanding algoritma dari kelas model lainnya.

Jika dibandingkan tingkat akurasi model berdasarkan Tabel 3 dan Tabel 5, dapat terlihat bahwa semua algoritma pada Tabel 5 mengalami penurunan tingkat akurasi kecuali pada algoritma *ZeroR* (*base line*). Penurunan tingkat akurasi ini lebih disebabkan oleh peningkatan jumlah atribut bukan karena peningkatan jumlah record. Secara umum peningkatan jumlah record akan meningkatkan jumlah *data training* dalam *cross validation* sehingga dapat menaikkan tingkat akurasi, sedangkan peningkatan jumlah atribut dapat menambah kompleksitas.

Namun demikian algoritma *Naïve Bayes* lebih unggul dibanding model yang lain dalam hal skalabilitas dengan penurunan akurasi hanya sebesar 0,59%. Algoritma *C4.5* mengalami penurunan akurasi terbesar (3,24%) sehingga kemampuan skalabilitas-nya lebih rendah dibanding model lainnya.

Untuk dataset *Adult census* dengan karakteristik lebih banyak record, lebih banyak atribut dan dengan *missing value*, hasil pengukuran sebagaimana terlihat pada Tabel 6 menunjukkan bahwa tingkat akurasi algoritma *C4.5* lebih unggul dibanding model lainnya. Hal ini menunjukkan bahwa algoritma *C4.5* mempunyai kehandalan (*robustness*) yang lebih baik dibanding model lainnya.

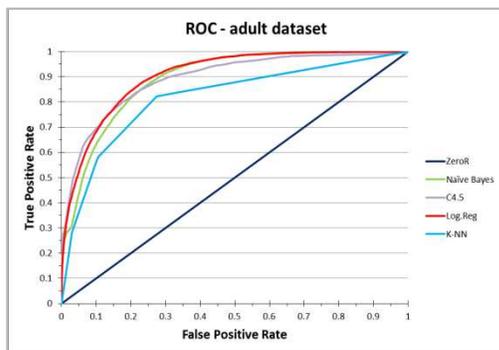
Pada kasus ini tingkat *precision* lebih utama dari pada akurasi, karena memprediksi seseorang sebagai target pajak padahal bukan lebih aman dari pada memprediksi seseorang bukan target padahal target pajak. Sehingga resiko kehilangan target pajak potensial bisa diminimalkan. Dari Tabel 6 dapat dilihat

tingkat *precision* algoritma *C4.5* lebih unggul dibanding model lainnya.

Tabel 6. Pengujian pada dataset Adult cencus

Jenis pengukuran	ZeroR		Naïve Bayes		Logistic Regression		K-NN		C4.5	
	Mean	Std.Dev	Mean	Std.Dev	Mean	Std.Dev	Mean	Std.Dev	Mean	Std.Dev
Accuracy	75.919	0.007	83.420	0.608	85.071	0.584	81.910	0.595	86.268	0.598
Error rate	24.081	0.007	16.580	0.608	14.929	0.584	18.090	0.595	13.732	0.598
TP-rate/ Sensitivity/ Recall	1.000	0.000	0.935	0.005	0.931	0.005	0.895	0.006	0.936	0.007
TN-rate/ Specificity	0.000	0.000	0.518	0.017	0.596	0.017	0.581	0.019	0.631	0.022
Precision	0.759	0.000	0.859	0.004	0.879	0.005	0.871	0.005	0.889	0.006
AUC	0.500	0.000	0.890	0.010	0.910	0.010	0.820	0.010	0.890	0.010
CPU-Time training	0.005	0.007	0.040	0.008	8.485	0.378	0.008	0.008	1.613	0.060
CPU-Time testing	0.001	0.005	0.014	0.005	0.008	0.008	12.431	0.115	0.002	0.005

Pada Gambar 5 terlihat bagaimana kurva ROC dari *Logistic Regression*, *C4.5* dan *Naïve Bayes* berada diatas kurva ROC kelas model lainnya. Agak sulit menentukan mana yang lebih unggul antara *Logistic Regression*, *C4.5* dan *Naïve Bayes* berdasarkan kurva ROC. Untuk itu perlu dibandingkan pula nilai AUC-nya.



Gambar 5. Kurva ROC pada Adult dataset

Berdasarkan Tabel 6 dapat dilihat bahwa nilai AUC algoritma *Logistic Regression* lebih tinggi daripada *C4.5* namun perbedaannya kecil sekali. Dengan demikian secara umum untuk kasus ini algoritma *C4.5* lebih optimal dibanding algoritma dari kelas model lainnya.

4. Simpulan

Berdasarkan hasil dan pembahasan yang telah dilakukan sebelumnya, dapat diambil beberapa kesimpulan sebagai berikut:

1. Kelas model *Instance Base* dan *Decision Tree* lebih handal (*robust*) dibanding kelas model lainnya bila diberi dataset yang mengandung *noise* dan *missing value*.
2. Tingkat akurasi (*accuracy*) kelas model *Bayesian* lebih stabil dibanding kelas model lainnya apabila

diberi dataset dengan jumlah atribut yang lebih banyak.

3. Semua kelas model mempunyai skalabilitas (*scalability*) yang baik dan bahkan mampu meningkatkan akurasi bila diberi jumlah record yang lebih besar.
4. Kelas model *Linear Function* dan *Decision Tree* membutuhkan waktu *training* yang relatif lebih lama dibanding model lainnya, namun membutuhkan waktu *testing* yang relatif lebih cepat dibanding model lainnya
5. Secara umum kelas model *Decision Tree* lebih unggul dibanding kelas model lainnya.

Penelitian ini masih terbatas hanya pada beberapa dataset dari kategori permasalahan *binary classification problem*. Penelitian ini juga masih terbatas pada perbandingan beberapa algoritma sederhana dari beberapa kelas model klasifikasi. Untuk pengembangan selanjutnya dapat menggunakan lebih banyak dataset dengan karakteristik yang lebih beragam dan menggunakan algoritma yang lebih banyak lagi.

5. Referensi

- Fatmawati. (2016). Perbandingan Algoritma Klasifikasi Data Mining Model C4.5 Dan Naive Bayes Untuk Prediksi Penyakit Diabetes. *Jurnal Techno Nusa Mandiri*, XIII(1), 50–59.
- Fiska, R. R. (2017). Penerapan Teknik Data Mining dengan Metode Support Vector Machine (SVM) untuk Memprediksi Siswa yang Berpeluang Drop Out (Studi Kasus di SMKN 1 Sutera). *SATIN - Sains Dan Teknologi Informasi*, 3(1), 15. <https://doi.org/10.33372/stn.v3i1.200>
- Fitri, S. (2014). Perbandingan Kinerja Algoritma Klasifikasi Naïve Bayesian , Lazy-Ibk , Zero-R , Dan Decision Tree- J48. *Dasi*, 15(1), 33–37. Retrieved from <https://ojs.amikom.ac.id/index.php/dasi/article/view/180>
- Ghorbani, R., & Ghousi, R. (2019). Predictive data mining approaches in medical diagnosis: A review of some diseases prediction. *International Journal of Data and Network Science*, 3, 47–70. <https://doi.org/10.5267/j.ijdns.2019.1.003>
- Gorunescu, F. (2011). *Data Mining: Concepts, Models and Techniques* (Vol. 12). Berlin, Heidelberg: Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-19721-5>
- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques* (3rd ed.). Elsevier. <https://doi.org/10.1016/B978-0-12-381479-1.00001-0>
- Hastuti, K. (2012). ANALISIS KOMPARASI ALGORITMA KLASIFIKASI DATA MINING UNTUK PREDIKSI MAHASISWA NON AKTIF. In *Seminar Nasional Teknologi Informasi & Komunikasi Terapan 2012* (Vol. 14, pp. 241–249).
- Mahmud, S. S., & Ilyas. (2017). Penerapan Algoritma Nearest Neighbor Retrieval Untuk Mendiagnosa

- Penyakit Hepatitis. *Sains Dan Teknologi Informasi (SATIN)*, 3(1), 24. <https://doi.org/10.33372/stn.v3i1.202>
- Marcos, H., & Utomo, H. S. (2015). Perbandingan kinerja algoritme c4.5 dan naïve bayes mengklasifikasi penyakit diabetes. *Jurnal Informatika*, 15(2), 141–148. <https://doi.org/http://dx.doi.org/10.30873/ji.v15i2>
- Menarianti, I. (2015). Klasifikasi data mining dalam menentukan pemberian kredit bagi nasabah koperasi. *Jurnal Ilmiah Teknosains*, 1(1), 36–45. Retrieved from <http://journal.upgris.ac.id/index.php/JITEK/article/view/836>
- Stéphane, T. (2011). *Data Mining and Statistics for Decision Making* (1st ed.). West Sussex: John Wiley & Sons Ltd.,Publication. <https://doi.org/10.4018/978-1-5225-2031-3.ch002>
- Witten, I. H., Frank, E., & Hall, M. a. (2011). *Data Mining: Practical Machine Learning Tools and Techniques* (2nd ed.). San Francisco: Morgan Kaufmann Publishers.
- Yusa, M., Utami, E., & Luthfi, E. T. (2016). Analisis Komparatif Evaluasi Performa Algoritma Klasifikasi pada Readmisi Pasien Diabetes. *Jurnal Buana Informatika*, 7(4), 293–302. <https://doi.org/10.24002/jbi.v7i4.770>
- Zhang, Y., Xin, Y., Li, Q., Ma, J., Li, S., Lv, X., & Lv, W. (2017). Empirical study of seven data mining algorithms on different characteristics of datasets for biomedical classification applications. *BioMedical Engineering OnLine*, 16(1), 125. <https://doi.org/10.1186/s12938-017-0416-x>