Terakreditasi SINTA Peringkat 2
Surat Keputusan Dirjen Penguatan Riset dan Pengembangan Ristek Dikti No. 10/E/KPT/2019
masa berlaku mulai Vol. 1 No. 1 tahun 2017 s.d Vol. 5 No. 3 tahun 2021

Terbit online pada laman web jurnal: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 5 No. 3 (2021) 504 - 510 ISSN Media Elektronik: 2580-0760

Analisis Optimasi Algoritma Klasifikasi Naive Bayes menggunakan Genetic Algorithm dan Bagging

Agung Nugroho[1], Yoga Religia[2]
[1,2]Teknik Informatika, Fakultas Teknik, Universitas Pelita Bangsa
[1]agung@pelitabangsa.ac.id, [2]yoga.religia@pelitabangsa.ac.id

Abstract

The increasing demand for credit applications to banks has motivated the banking world to switch to more sophisticated techniques for analyzing the level of credit risk. One technique for analyzing the level of credit risk is the data mining approach. Data mining provides a technique for finding meaningful information from large amounts of data by way of classification. However, bank marketing data is a type of imbalance data so that if the classification is done the results are less than optimal. The classification algorithm that can be used for imbalance data types can use naïve Bayes. Naïve Bayes performs well in terms of classification. However, optimization is needed in order to obtain more optimal classification results. Optimization techniques in handling imbalance data have been developed with several approaches. Bagging and Genetic Algorithms can be used to overcome imbalance data. This study aims to compare the accuracy level of the naïve Bayes algorithm after optimization using the bagging and genetic algorithm. The results showed that the combination of bagging and a genetic algorithm could improve the performance of Naive Bayes by 4.57%.

Keywords: Classification, Bank Marketing, Naïve Bayes, Bagging, Genetic Algotithm

Abstrak

Peningkatan permintaan pengajuan kredit pada perbankan telah memotifasi dunia perbankan untuk beralih pada teknik yang lebih canggih untuk menganalisa tingkat resiko kredit. Salah satu teknik untuk menganalisa tingkat resiko kredit adalah dengan pendekatan data mining. Data mining menyediakan teknik untuk menemukan informasi yang bermakna dari sejumlah data besar dengan cara klasifikasi. Data bank marketing termasuk jenis data *imbalance* sehingga apabila dilakukan klasifikasi hasilnya kurang optimal. Algoritma klasifikasi yang dapat digunakan untuk jenis data *imbalance* dapat menggunakan *naïve bayes*. *Naïve bayes* memiliki kinerja baik dalam hal klasifikasi, namun demikian diperlukan optimasi agar mendapatkan hasil klasifikasi yang lebih optimal. Teknik optimasi dalam menangani data *imbalance* telah banyak dikembangkan dengan beberapa pendekatan. *Bagging* dan *Genetic Algorithm* dapat digunakan dalam mengatasi data *imbalance*. Penelitian ini bertujuan untuk membandingkan tingkat akurasi algoritma *naïve bayes* setelah dilakukan optimasi dengan menggunakan *bagging* dan *genetic algorithm*. Hasil penelitan menunjukkan bahwa kombinasi *bagging* dengan *genetic algorithm* dapat meningkatkan performa naive bayes sebesar 4,57%.

Kata kunci: Klasifikasi, Bank Marketing, Naïve Bayes, Bagging, Genetic Algotithm.

1. Pendahuluan

Pertumbuhan bisnis mendorong banyaknya peluang kredit pada perbankan untuk meningkatkan modal usaha. Kredit merupakan perjanjian pinjam meminjam uang antara bank sebagai penyedia dana dengan pihak nasabah sebagai penerima dana[1]. Pemberian kredit kepada nasabah umumnya dipengaruhi oleh beberapa faktor seperti kepercayaan, jangka waktu, tingkat resiko dan objek kredit[2]. Bank marketing perlu menganalisa terkait hal tersebut untuk menjaga nasabahnya agar tidak terjadi kredit bermasalah karena kredit bermasalah merupakan risiko yang sering terjadi pada setiap pemberian kredit[3].

Tingkat resiko kredit dapat dikurangi dengan melakukan analisa terhadap data nasabah dengan pendekatan data mining[4]. Penelitian terkait evaluasi resiko kredit telah banyak dilakukan, hal ini merupakan masalah yang menarik dalam analisa keuangan[5]. Teknik data mining mampu mengidentifikasi korelasi, pola dan penemuan pengetahuan dari dataset. Data mining juga telah

Diterima Redaksi: 05-04-2021 | Selesai Revisi: 07-05-2021 | Diterbitkan Online: 20-06-2021

berhasil diterapkan di berbagai domain seperti ritel, bisnis, pemasran, kesehatan dan lain sebagainya[6].

Teknik data mining secara umum terdiri dari dua kategori, yaitu bersifat prediktif dan deskriptif. Kedua metode tersebut digunakan untuk mengekstrak pola yang tersembunyi dari sejumlah data besar[7]. Salah satu Teknik data mining adalah klasifikasi. Klasifikasi adalah suatu bentuk analisis data yang mengekstrak model yang mendeskripsikan kelas data[8][7]. Klasifikasi dapat digunakan untuk menganalisa tingkat resiko pinjaman pada bank[8].

Algoritma klasifikasi melakukan prediksi label kelas kategorikal dari sebuah data, sehingga dapat mengklasifikasikannya kedalam salah satu kelas yang ditentukan. *Dataset* Bank Marketing termasuk jenis data *imbalance class*[9], sehingga diperlukan pemilihan algoritma yang tepat dalam mengklasifikasikan data tersebut[10]. Masalah ketidakseimbangan kelas terjadi ketika salah satu kelas dari dataset memiliki sample yang sangat sedikit dibandingkan dengan kelas lainnya sehingga hasil prediksinya akan bias terhadap kelas mayoritas. Sampel minirotas adalah yang jarang muncul tetapi sangat penting dalam klasifikasi [11].

Berbagai jenis algoritma klasifikasi telah diterapkan untuk *imbalance class dataset*, termasuk *naïve bayes*. *Naïve bayes* memiliki kinerja yang baik dan menghasilkan probabilitas rata-rata 71 persen dengan waktu proses yang lebih cepat dibanding algoritma pembelajaran mesin lain serta memiliki reputasi pada keakuratan prediksi[12]. Namun demikian dalam mengatasi ketidakseimbangan kelas, algoritma *naïve bayes* perlu dilakukan optimasi untuk mendapatkan nilai akurasi yang lebih baik[13][14].

Banyak teknik optimasi yang telah dikembangkan untuk mengatasi masalah ketidakseimbangan kelas yang dikelompokkan menjadi tiga jenis pendekatan, yaitu pendekatan level data, pendekatan level algoritma, dan pendekatan *hybrid*[15] dengan teknik *ensemble*[16][17].

Salah satu metode *ensemble* yang telah banyak digunakan adalah *bagging*[16]. *Bagging* atau *Bootstrap Aggregation* merupakan metode *ensemble* yang dapat meningkatkan klasifikasi dengan melakukan kombinasi klasifikasi secara acak pada *dataset training* yang dapat mengurangi variasi dan menghindari terjadinya *overfitting*[18].

Selain *bagging*, dalam mengatasi ketidakseimbangan kelas dapat juga dilakukan dengan *genetic algorithm*[19][20]. Penggunaan *genetic algorithm* sebagai pemilihan fitur pada optimasi klasifikasi naïve bayes telah menujukkan peningkatan akurasi[14].

Berdasarkan uraian sebelumnya penelitian ini bertujuan untuk mengetahui tingkat akurasi algoritma *Naïve Bayes* untuk klasifikasi data bank marketing setelah dilakukan optimasi menggunakan *Bagging* dan *Genetic Algorithm* dalam mengatasi permasalahan ketidakseimbangan kelas.

2. Metode Penelitian

2.1 *Dataset*

Data yang digunakan dalam penelitian ini adalah data *public* yang diambil dari *UCI Repository* tentang data *bank marketing*[21]. *Dataset* dapat diunduh dari laman https://archive.ics.uci.edu/ml/datasets/Bank+Marketing . Jumlah data sebanyak 45.211 *record* dengan 16 atribut dan satu kelas label y. Data tersebut berkaitan dengan kampanye pemasaran langsung dari lembaga perbankan di Postugis berdasarkan panggilan telepon terkait produk deposito bank[9]. Tujuan dari klasifikasi adalah untuk meprediksi apakah klien akan berlangganan deposito berjangka (variable y). Jumlah kelas data minoritas adalah 5289 (yes) dan jumlah kelas mayoritas adalah 39922 (no). Spesifikasi data dapat dilihat pada tabel 1.

Tabel 1. Data

Data	Kelas	Atribut	Record
Bank Marketing	2	16	45.211

Adapun atributnya terdiri dari *variable input* dan *output* dengan spesifikasi dapat dilihat padat tabel 2.

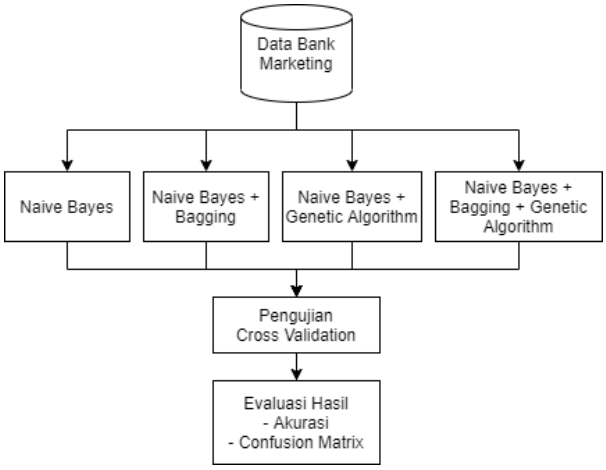
Tabel 2. Atribut Data *Bank Marketing*

Variable	Atribut	Tipe Data
input	age	numeric
	job	categorical
	marital	categorical
	education	categorical
	default	categorical
	balance	numeric
	housing	categorical
	loan	categorical
	contact	categorical
	month	categorical
	day_of_week	categorical
	duration	numeric
	campaign	numeric
	pdays	numeric
	previous	numeric
	poutcome	numeric
output	y (accepted)	binary

2.2 Skema Penelitian

Skema penelitian yang digunakan adalah melakukan perbandingan model algoritma klasifikasi *naïve bayes* dengan *naïve bayes* yang dioptimasi dengan algoritma *bagging* dan *genetic algorithm* untuk klasifikasi data *bank marketing*. Gambar 1 menujukkan skema penelitian yang digunakan.

DOI: https://doi.org/10.29207/resti.v5i3.3067



Gambar 1. Skema Penelitian

Berdasarkan gambar 1 dapat dilihat bahwa proses pengujian dilakukan sebanyak 4 kali yaitu: 1. Pengujian klasifikasi menggunakan algoritma *naïve bayes*, 2. Pengujian klasifikasi menggunakan algoritma *naïve bayes* yang dioptimasi menggunakan *bagging*, 3. Pengujian klasifikasi menggunakan *algoritma naïve bayes* yang dioptimasi menggunakan *genetic algorithm*, dan 4. Pengujian klasifikasi menggunakan algoritma *naïve bayes* dengan optimasi kombinasi *bagging* dengan *genetic algorithm*.

Teknik pengujian menggunakan *cross validation* dan untuk evaluasi hasil pengujian menggunakan *confusion matrix* untuk melihat nilai akurasi, presisi dan *recall*.

2.3 Klasifikasi *Naïve Bayes*

Naïve bayes classifier merupakan model pembelajaran mesin probabilistik yang digunakan untuk melakukan klasifikasi berdasarkan teorema bayes[9]. *Naive bayes* bekerja dengan memprediksi probabilitas bahwa suatu sample data termasuk pada kelas tertentu. Vektor data yang diberikan berada pada kelas C, disebut sebagai probabilitas posterior dan dilambangkan dengan P(C|X)[22]. Tahapan algoritma *naïve bayes* dijabarkan dalam persamaan 1 berikut[8].

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)} \quad (1)$$

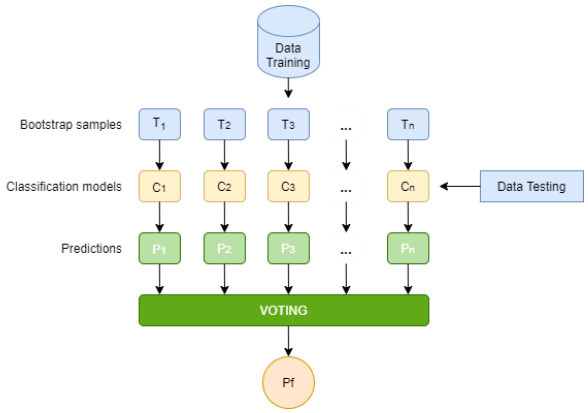
Persamaan 1 menjelaskan nilai P(X) adalah konstan untuk semua kelas, dimana nilai X merupakan *record training* yang dinyatakan dengan n atribut X=(X_1, X_2, ..., X_n). C_i menujukkan kelas dari data, yang dinyatakan dengan C_1, C_2, ..., C_n.

Penggunaan naïve bayes sebagai algoritma klasifikasi telah menunjukkan kinerja yang baik dalam masalah dunia nyata yang komplek[15]. Keuntungan menggunakan Naïve Bayes adalah hanya membutuhkan sedikit data pelatihan untuk menentukan estimasi parameter yang dibutuhkan dalam proses klasifikasi. Naive Bayes bekerja sangat baik di sebagian besar situasi dunia nyata yang kompleks dalam menangani nilai yang hilang dalam kumpulan data yang homogen atau heterogen[23].

2.4 *Bagging*

Bagging merupakan salah satu metode *ensamble learning* yang paling efektif dan populer dalam mengoptimalkan proses klaisifikasi dan telah banyak diterakpan di dunia nyata[24]. *Bagging* juga disebut sebagai *Bootstrap Aggregating* yang merupakanm metode untuk memperbaiki kinerja algoritma klasifikasi pada *machine learning*[25]. *Bagging* bertujuan untuk meningkatkan akurasi klasifikasi dengan menggabungkan klasifikasi tunggal dan hasilnya lebih baik dari pada *random sampling*[26].

Bagging bekerja dengan cara membagi data *training* menjadi beberapa bagian, kemudian dibuat model klasifikasi dari masing-masing bagian untuk mendapatkan hasil prediksi dari model tersebut. Hasil akhir didapatkan dengan cara voting. Voting yang dilakukan dengan cara mengambil suara terbanyak[26]. Cara kerja bagging dapat dilihat pada ilistrasi gambar 2.



Gambar 2. Cara kerja *Bagging*

2.5 *Genetic Algorithm*

Genetic Algorithm (GA) merupakan algoritma yang bekerja dengan pendekatan heruistik yang dapat diterapkan pada berbagai masalah optimasi[27]. *Genetic Algorithm* melakukan proses pencarian nilai optimal pada beberapa titik secara bersamaan dalam satu generasi dengan pendekatan evolusioner pada iterasinya. Pada proses evolusi, sejumlah gen penyusun kromosom akan mengalami proses persilangan dan mutasi. *Genetic Algorithm* menggunakan transisi probabilistik untuk memilih kromosom terbaik untuk mendapatkan solusi yang optimal[23].

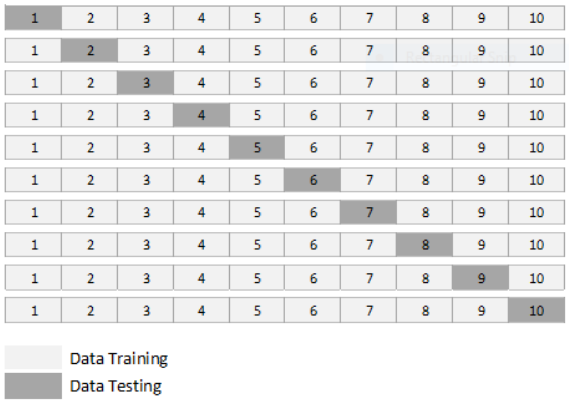
Genetic Algorithm pada dasarnya bekerja dengan melakukan pencarian untuk solusi potensial diarea pencarian yang spesifik yang dapat menghemat waktu karena tidak perlu melakukan pengujian semua kemungkinan kombinasi[28].

DOI: https://doi.org/10.29207/resti.v5i3.3067

2.6 Pengujian menggunakan *Cross Validation*

Cross validation merupakan metode statistik untuk mengevaluasi kinerja dari model atau algoritma. *Cross validation* membagi data menjadi 2 subset data yaitu *dataset training* dan *dataset testing*. Ada beberapa model *cross validation*, umumnya digunakan model *k-fold validation*. *K-fold validation* digunakan karena dapat mengurangi waktu komputasi[29]. Nilai k adalah banyaknya iterasi yang digunakan. 10 *fold validation* adalah salah satu *k-fold validation* yang direkomendasikan untuk pemilihan model terbaik, karena dapat memberikan estimasi akurasi yang maksimal.

Pada 10 *fold validation*, data dibagi menjadi 10 *fold* dengan ukuran yang sama sehingga menajdi 10 *subset* data. Dari 10 *fold* diambil 9 *fold* menjadi data *training* dan 1 *fold* sebagai data *testing*, kemudian dilakukan perulangan sebanyak 10 kali seperti yang diilustrasikan pada gambar 3.



Gambar 3. Skema 10 *fold validation*

2.7 Evaluasi Hasil dengan *Confusion Matrix*

Hasil pengujian klasifikasi dilakukan evaluasi untuk mengetahui nilai akurasi dari algoritma yang akan dianalisa apakah model klasifikasi yang dibuat layak digunakan. Metode evaluasi yang digunakan untuk mengukur kinerja dari algoritma klasifikasi adalah *confusion matrix*. *Confusion matrix* menampilkan perbandingan hasil klasifikasi yang dilakukan dengan data sebenarnya dalam bentuk tabel matrix yang dapat dilihat pada tabel 3.

Tabel 3. *Confusion Matrix*

Accuracy = 0%		*Prediction*		*Total Prediction*
		positive	negative	
Actual	positive	TP	FN	P
	negative	FP	TN	N
Total Actual		P^1	N^1	P+N

Dengan TP (*True Positive*) adalah nilai dari label *positive* yang diprediksi dengan benar. TN (*True Negative*) adalah nilai dari label *negative* yang diprediksi dengan benar. FP (*False Positive*) adalah nilai label *negaitive* yang diprediksi sebagai label *positive*. Dan FN (*False Negative*) adalah nilai label *positive* yang diprediksi sebagai label *negative*. P adalah nilai positive yang sebenarnya. N adalah nilai negative yang sebenarnya. P^1 adalah nilai positive hasil prediksi. Dan N^1 adalah nilai negative hasil prediksi. P+N adalah total *instance dataset* yang digunakan pada proses klasifikasi.

Untuk mengukur *performance matrix* ada beberapa paremeter yang dijadikan acuan dalam menilai kinrja algoritma diantaranya adalah akurasi, presisi dan *recall*. Nilai akurasi dapat dihitung dengan persamaan 2.

$$Akurasi = \frac{TP+TN}{P+N} \quad (2)$$

Akurasi adalah nilai rasio prediksi benar dari total data. Nilai presisi adalah perbandingan nilai rasio prediksi benar dengan total nilai yang diprediski dengan benar. Untuk mendapatkan nilai presisi dari masing-masing kelas digunakan persamaan 3 dan 4.

$$Presisi\ (TP) = \frac{TP}{P1} \quad (3)$$

$$Presisi\ (TN) = \frac{TN}{N1} \quad (4)$$

Nilai presisi dari keseluruhan kelas dapat dihitung dengan persamaan 5.

$$Presisi = \frac{P}{P+N} * Presisi\ TP + \frac{N}{P+N} * Presisi\ TN \quad (5)$$

Sedangkan nilai *recall* adalah perbandingan rasio prediksi benar dengan total data yang benar. Nilai *recall* tiap-tiap kelas dapat dihitung dengan persamaan 6 dan 7.

$$Recall\ (TP) = \frac{TP}{P} \quad (6)$$

$$Recall\ (TN) = \frac{TN}{N} \quad (7)$$

Nilai *recall* dari keseluruhan kelas dapat dihitung dengan persamaan 8.

$$Recall = \frac{P}{P+N} * Recall\ TP + \frac{N}{P+N} * Recall\ TN \quad (8)$$

3. Hasil dan Pembahasan

Dataset yang digunakan pada penelitian ini diambil dari *UCI Repository* berupa data tentang *bank marketing*. *Dataset* tersebut kemudian diklasifikasian menggunakan algoritma *Naïve Bayes*. Algoritma optimasi juga digunakan pada penelitian ini untuk menguji nilai akurasi terhadap algoritma *naïve bayes*. Algoritma optimasi yang digunakan adalah *Genetic Algorithm* dan *Bagging*.

3.1 Hasil Penelitian

Pengujian terhadap model klasifikasi menggunakan *cross-validation* untuk membagi data *training* dan data *testing*. *cross validation* akan membagi *dataset* menjadi K partisi dengan bagian data *training* dan data *testing*.

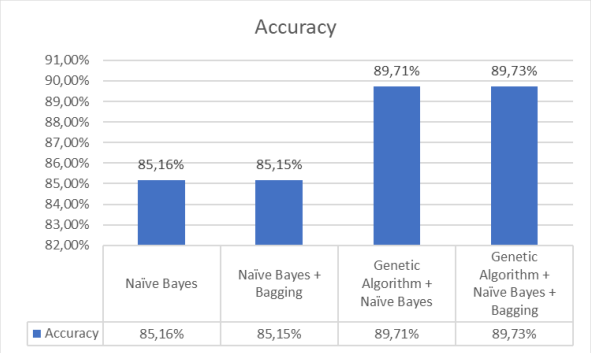
DOI: https://doi.org/10.29207/resti.v5i3.3067

Proses ini dilakukan sebanyak nilai K yang ditetapkan sehingga semua partisi pernah menjadi data *testing*. Hasil pengujian klasifikasi terhadap data *bank marketing* menggunakan algoritma *naïve bayes* dengan algoritma optimasi *genetic algorithm* dan *bagging* dapat dilihat pada tabel 4.

Tabel 4. Hasil Akurasi Pengujian

Algoritma	Akurasi
Naïve Bayes	85,16%
Naïve Bayes + Bagging	85,15%
Genetic Algorithm + Naïve Bayes	89,71%
Genetic Algorithm + Naïve Bayes + Bagging	89,73%

Berdasarkan tabel 4 tersebut dapat dilihat bahwa nilai akurasi untuk algoritma *naïve bayes* sebesar 85,16%. Optimasi algoritma *naïve bayes* dengan *bagging* mendapatkan nilai akurasi sebesar 85,15%, sedangkan optimasi *naïve bayes* dengan *getenic algorithm* mendapatkan nilai akurasi sebesar 89,71%. Selanjutnya nilai akurasi dari kombinasi *genetic algorithm* dan *bagging* terhadap algoritma *naïve bayes* sebesar 89,73%. Hasil pengujian yang diperoleh menunjukan bahwa hasil terbaik adalah optimasi menggunakan *genetic algorithm* yang dikombinasikan dengan *bagging*. Hasil perbandingan nilai akurasi kinerja algoritma ditampilkan dalam bentuk grafik yang dapat dilihat pada gambar 4.



Gambar 4 Grafik perbandingan nilai akurasi kinerja algoritma

Hasil evaluasi kinerja dari masing-masing algoritma digambarkan dalam bentuk *confusion matrix* yang dapat dilihat pada tabel 5.

Tabel 5. *confusion matrix* kinerja Naïve Bayes

Accuracy = 85,16%		Prediction		Recall
		no	yes	
Actual	no	2874	2415	0,89
	yes	4293	35629	0,54
Precission		0,94	0,40	

Tabel 5 menampilkan hasil evaluasi kinerja algoritma *naïve bayes* pada proses pengujian *cross validation* dengan 10 kali iterasi. Dari 45211 data yang diuji, 2874 data kelas *no* diprediksi dengan benar, dan 2415 data dari kelas *no* yang diprediksi sebagai kelas *yes*. Kemudian 35629 kelas *yes* diprediksi dengan benar, dan 4293 kelas *yes* yang diprediksi sebagai kelas *no*. Dari matrix tersebut dapat dilihat nilai presisi dari masing-masing kelas adalah 0,94% untuk kelas no, dan 0,40% untuk kelas yes. Sehingga rata-rata nilai presisinya adalah sebesar 0,67%. Sedangkan nilai *recall* untuk masing-masing kelas adalah 0,89% untuk kelas no dan 0,54% untuk kelas yes, sehingga rata-rata nilai *recall* adalah sebesar 0,72%.

Selanjutnya *confusion matrix* untuk kinerja algoritma naïve bayes dengan optimasi algoritma *bagging* ditampilkan pada tabel 6.

Tabel 6 *confusion matrix* kinerja *Naïve Bayes* dan *Bagging*

Accuracy = 85,15%		Prediction		Recall
		no	yes	
Actual	no	2868	2421	0,89
	yes	4292	35630	0,54
Precission		0,94	0,40	

Tabel 6 menampilkan hasil evaluasi kinerja algoritma *naïve bayes* dengan algoritma otimasi *bagging* pada proses pengujian *cross validation* dengan 10 kali iterasi. Dari 45211 data yang diuji, 2868 data kelas *no* diprediksi dengan benar, dan 2421 data dari kelas *no* yang diprediksi sebagai kelas *yes*. Kemudian 35630 kelas *yes* diprediksi dengan benar, dan 4292 kelas *yes* yang diprediksi sebagai kelas *no*. Dari matrix tersebut dapat dilihat nilai presisi dari masing-masing kelas adalah 0,94% untuk kelas *no*, dan 0,40% untuk kelas *yes*. Sehingga rata-rata nilai presisinya adalah sebesar 0,67%. Sedangkan nilai *recall* untuk masing-masing kelas adalah 0,89% untuk kelas *no* dan 0,54% untuk kelas *yes*, sehingga rata-rata nilai *recall* adalah sebesar 0,72%.

Selanjutnya *confusion matrix* untuk kinerja algoritma *naïve bayes* dengan optimasi *genetic algorthm* ditampilkan pada tabel 7 sebagai berikut:

Tabel 7 *confusion matrix* kinerja *Naïve Bayes* dengan *Genetic Algorithm*

Accuracy = 89,71%		Prediction		Recall
		no	yes	
Actual	no	2062	3227	0,96
	yes	1425	38497	0,39
Precission		0,92	0,59	

Tabel 7 menampilkan hasil evaluasi kinerja algoritma *naïve bayes* dengan otimasi *genetic algorithm* pada proses pengujian *cross validation* dengan 10 kali iterasi dam dengan nilai populasi pada *genetic algorithm* sebesar 50 populasi. Dari 45211 data yang diuji, 2062 data kelas *no* diprediksi dengan benar, dan 1425 data dari kelas *no* yang diprediksi sebagai kelas *yes*. Kemudian 38497 kelas *yes* diprediksi dengan benar, dan 3227 kelas *yes* yang diprediksi sebagai kelas *no*. Dari matrix tersebut dapat dilihat nilai presisi dari masing-masing kelas adalah 0,92% untuk kelas *no*, dan 0,59% untuk kelas *yes*. Sehingga rata-rata nilai presisinya adalah sebesar 0,76%. Sedangkan nilai *recall* untuk

DOI: https://doi.org/10.29207/resti.v5i3.3067

masing-masing kelas adalah 0,96% untuk kelas *no* dan 0,39% untuk kelas *yes*, sehingga rata-rata nilai *recall* adalah sebesar 0,68%.

Selanjutnya *confusion matrix* untuk kinerja algoritma *naïve bayes* dengan optimasi *genetic algirthm* yang dikombinasikan dengan *bagging* ditampilkan pada tabel 8 sebagai berikut:

Tabel 8 *confusion matrix* kinerja *Naïve Bayes* dengan *Genetic Algorithm* dan *bagging*

Accuracy = 89,71%		Prediction		Recall
		no	yes	
Actual	no	2058	3231	0,96
	yes	1410	38512	0,39
Precission		0,92	0,59	

Tabel 8 menampilkan hasil evaluasi kinerja algoritma *naïve bayes* dengan otimasi *genetic algorithm* yang dikombinasikan dengan *bagging* pada proses pengujian *cross validation* dengan 10 kali iterasi dam dengan nilai populasi pada *genetic algorithm* sebesar 50 populasi. Dari 45211 data yang diuji, 2058 data kelas *no* diprediksi dengan benar, dan 1410 data dari kelas *no* yang diprediksi sebagai kelas *yes*. Kemudian 38512 kelas *yes* diprediksi dengan benar, dan 3231 kelas *yes* yang diprediksi sebagai kelas *no*. Dari matrix tersebut dapat dilihat nilai presisi dari masing-masing kelas adalah 0,92% untuk kelas *no*, dan 0,59% untuk kelas *yes*. Sehingga rata-rata nilai presisinya adalah sebesar 0,76%. Sedangkan nilai *recall* untuk masing-masing kelas adalah 0,96% untuk kelas *no* dan 0,39% untuk kelas *yes*, sehingga rata-rata nilai *recall* adalah sebesar 0,68%.

3.2 Analisa Hasil Pengujian

Berdasarkan hasil pengujian dari model klasifikasi yang digunakan pada penelitian ini didapatkan hasil bahwa penggunaan algoritma optimasi *bagging* dengan kombinasi *genetic algorthm* dapat meningkatkan akurasi algoritma klasifikasi naive bayes sebesar 4,57%. Namun demikian penggunaan *bagging* saja justru hasilnya menurun 0,01%. Berdasarkan evaluasi pengujian menggunakan *confusion matrix* didapatkan nilai rata-rata presisi dan *recall* dari algritma *naïve bayes* dan *naïve bayes* dengan *bagging* adalah sama. Begitu juga antara *naïve bayes* dengan *genetic algorithm* dan kombinasi antara *bagging* dan *genetic algorithm* juga memilki nilai yang sama. Berdasarkan hasil tersebut dapat diketahui bahwa penambahan *bagging* pada algoritma *naïve bayes* tidak berpengaruh terhadap nilai presisi dan *recall*.

4. Kesimpulan

Berdasarkan hasil dan pembahasan dari model penelitian yang digunakan pada penelitian dapat disimpulkan bahwa penggunaan Bagging untuk optimasi Naïve Bayes dalam mengklasifikasi data Bank Marketing ternyata memiliki performa akurasi yang lebih kecil dari pada penggunaan algoritma Naïve Bayes saja (tanpa optimasi Bagging). Performa paling baik adalah dengan penggunaan *Genetic Algorithm* dan Bagging secara bersamaan untuk mengoptimasi algoritma Naïve Bayes dalam mengklasifikasi data Bank Marketing dengan akurasi sebesar 89,73%. Akurasi tersebut lebih besar 4,57% dibandingkan penggunaan algoritma Naïve Bayes saja. Apabila dibandingkan penggunaan *Genetic Algorithm* dan Bagging dengan penggunaan *Genetic Algorithm* saja untuk mengoptimasi algoritma Naïve Bayes dalam mengklasifikasi data Bank Marketing ternyata selisih akurasinya hanya 0,02%, yang artinya peran penggunaan Bagging dalam kombinasi *Genetic Algorithm* dan Bagging untuk optimasi Naïve Bayes sangatlah kecil sehingga di butuhkan kombinasi yang lain untuk mengoptimalkan performa klasifikasi.

Penelitian selanjutnya diharapkan dapat membahas tentang optimasi lain yang dapat digunakan untuk dikombinasikan dengan *Genetic Algorithm* untuk mengoptimasi algoritma Naïve Bayes dalam mengklasifikasi data Bank Marketing, misalkan seperti *boosting*. Penggunaan algoritma *boosting* memungkinkan untuk mengurangi bias data dan meningkatkan kesesuaian data dengan model, sehingga diharapkan mampu meningkatkan performa dari penggunaan *Genetic Algorithm* untuk mengoptimasi algoritma Naïve Bayes dalam mengklasifikasi data Bank Marketing.

Daftar Rujukan

[1] N. Hadinata, "Implementasi Metode Multi Attribute Utility Theory (MAUT) Pada Sistem Pendukung Keputusan dalam Menentukan Penerima Kredit," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 7, no. 2, Sep. 2018, doi: 10.32736/sisfokom.v7i2.562.

[2] I. Sugiyarto, "Perbandingan Kinerja Algoritma Data Mining Prediksi Persetujuan Kartu Kredit," *Fakt. Exacta*, vol. 12, no. 3, pp. 180–192, 2019, doi: 10.30998/faktorexacta.v12i3.4310.

[3] S. Somadiyono and T. Tresya, "Tanggung Jawab Pidana Marketing Menurut Undang Undang Perbankan Terhadap Pembiayaan Bermasalah Di Bank Muamalat Indonesia,Tbk," *J. LEX Spec.*, no. 21, pp. 22–38, 2015.

[4] S. Masripah, "Komparasi Algoritma Klasifikasi Data Mining untuk Evaluasi Pemberian Kredit," *BINA Insa. ICT J.*, vol. 3, no. 1, pp. 187–193, 2016 [Online]. Available: http://ejournal-binainsani.ac.id/index.php/BIICT/article/view/815. [Accessed: 23-Dec-2020]

[5] H. Leidiyana, "Penerapan Algoritma K-Nearest Neighbor Untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bemotor," 2013 [Online]. Available: http://jurnal.unismabekasi.ac.id/index.php/piksel/article/view/293. [Accessed: 23-Dec-2020]

[6] P. Singh and N. Singh, "Role of Data Mining Techniques in Bioinformatics," *Int. J. Appl. Res. Bioinforma.*, vol. 11, no. 1, Jan. 2021, doi: 10.4018/IJARB.2021010106.

[7] S. Umadevi and K. S. J. Marseline, "A survey on data mining classification algorithms," in *2017 International Conference on Signal Processing and Communication (ICSPC)*, 2017, doi: 10.1109/CSPC.2017.8305851.

DOI: https://doi.org/10.29207/resti.v5i3.3067

[8] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. Elsevier, 2012.

[9] A. Verma, "Evaluation of Classification Algorithms with Solutions to Class Imbalance Problem on Bank Marketing Dataset using WEKA," *Int. Res. J. Eng. Technol.*, 2019 [Online]. Available: www.irjet.net

[10] Yoga Religia, Agung Nugroho, and Wahyu Hadikristanto, "Klasifikasi Analisis Perbandingan Algoritma Optimasi pada Random Forest untuk Klasifikasi Data Bank Marketing," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 1, Feb. 2021, doi: 10.29207/resti.v5i1.2813.

[11] M. R. Longadge, M. Snehlata, S. Dongre, and D. Latesh Malik, "Class Imbalance Problem in Data Mining: Review," 2013 [Online]. Available: www.ijcsn.org

[12] R. S. Wahono, "Integrasi SMOTE dan Information Gain pada Naive Bayes untuk Prediksi Cacat Software," 2015 [Online]. Available: http://journal.ilmukomputer.org

[13] I. G. A. Socrates, A. L. Akbar, M. S. Akbar, A. Z. Arifin, and D. Herumurti, "Optimasi Naive Bayes Dengan Pemilihan Fitur Dan Pembobotan Gain Ratio," *Lontar Komput. J. Ilm. Teknol. Inf.*, Mar. 2016, doi: 10.24843/LKJITI.2016.v07.i01.p03.

[14] O. Somantri and M. Khambali, "Feature Selection Klasifikasi Kategori Cerita Pendek Menggunakan Naïve Bayes dan Algoritme Genetika," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 6, no. 3, Sep. 2017, doi: 10.22146/jnteti.v6i3.332. [Online]. Available: http://ejnteti.jteti.ugm.ac.id/index.php/JNTETI/article/view/332

[15] C. K. Aridas, S. Karlos, V. G. Kanas, N. Fazakis, and S. B. Kotsiantis, "Uncertainty Based Under-Sampling for Learning Naive Bayes Classifiers under Imbalanced Data Sets," *IEEE Access*, vol. 8, pp. 2122–2133, 2020, doi: 10.1109/ACCESS.2019.2961784.

[16] Youqin Pan and Zaiyong Tang, "Ensemble methods in bank direct marketing," in *2014 11th International Conference on Service Systems and Service Management (ICSSSM)*, 2014, doi: 10.1109/ICSSSM.2014.6874056.

[17] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera, "A Review on Ensembles for the Class Imbalance Problem: Bagging-, Boosting-, and Hybrid-Based Approaches," *IEEE Trans. Syst. Man, Cybern. Part C (Applications Rev.*, vol. 42, no. 4, Jul. 2012, doi: 10.1109/TSMCC.2011.2161285.

[18] R. S. Wahono and N. Suryana, "Combining particle swarm optimization based feature selection and bagging technique for software defect prediction," *Int. J. Softw. Eng. its Appl.*, vol. 7, no. 5, pp. 153–166, 2013, doi: 10.14257/ijseia.2013.7.5.16.

[19] J. Ha and J. S. Lee, "A new under-sampling method using genetic algorithm for imbalanced data classification," in *ACM IMCOM 2016: Proceedings of the 10th International Conference on Ubiquitous Information Management and Communication*, 2016, doi: 10.1145/2857546.2857643.

[20] V. Karia, W. Zhang, A. Naeim, and R. Ramezani, "GenSample: A Genetic Algorithm for Oversampling in Imbalanced Datasets," Oct. 2019 [Online]. Available: http://arxiv.org/abs/1910.10806. [Accessed: 30-Mar-2021]

[21] S. Moro, P. Cortez, and P. Rita, "A data-driven approach to predict the success of bank telemarketing," *Decis. Support Syst.*, vol. 62, pp. 22–31, Jun. 2014, doi: 10.1016/j.dss.2014.03.001. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S016792361400061X

[22] A. Verma, "Study and Evaluation of Classification Algorithms In Data Mining," *Int. Res. J. Eng. Technol.*, 2018.

[23] B. K. Khotimah, M. Miswanto, and H. Suprajitno, "Optimization of feature selection using genetic algorithm in naïve Bayes classification for incomplete data," *Int. J. Intell. Eng. Syst.*, vol. 13, no. 1, pp. 334–343, Feb. 2020, doi: 10.22266/ijies2020.0229.31.

[24] G. Liang, X. Zhu, and C. Zhang, "The effect of varying levels of class distribution on bagging for different algorithms: An empirical study," *Int. J. Mach. Learn. Cybern.*, vol. 5, no. 1, Feb. 2014, doi: 10.1007/s13042-012-0125-5.

[25] L. Breiman, "Bagging Predictors," Kluwer Academic Publishers, 1996.

[26] E. Alfaro, M. Gámez, and N. García, "adabag: An R Package for Classification with Boosting and Bagging," *J. Stat. Softw.*, vol. 54, no. 2, 2013, doi: 10.18637/jss.v054.i02.

[27] O. Kramer, "Genetic Algorithms," 2017.

[28] D. Jorge Martins Sousa, R. Fuentecilla Maia Ferreira Neves Nuno Cavaco Gomes Horta, A. Manuel Raminhos Cordeiro Grilo Supervisor, and R. Fuentecilla Maia Ferreira Neves, "Using Naïve Bayes and Genetic Algorithms to Find Influential Twitter Users to Forecast the S&P 500 Electrical and Computer Engineering Examination Committee," 2017.

[29] A. Wibowo, "10 Fold Cross Validation," 2017. [Online]. Available: https://mti.binus.ac.id/2017/11/24/10-fold-cross-validation. [Accessed: 23-Dec-2020]

DOI: https://doi.org/10.29207/resti.v5i3.3067