

Penerapan Discretization dan Teknik Bagging Untuk Meningkatkan Akurasi Klasifikasi Berbasis Ensemble pada Algoritma C4.5 dalam Mendiagnosa Diabetes

Mirqotussa'adah^{a1}, Much Aziz Muslim^{a2}, Endang Sugiharti^{a3},
Budi Prasetyo^{a4}, Siti Alimah^{b5}

^aJurusan, Fakultas, Universitas

Menyertakan alamat institusi dan negara

^aJurusan Ilmu Komputer, FMIPA, Universitas Negeri Semarang
Kampus Sekaran Gunungpati Semarang, Indonesia

^bJurusan Biologi, FMIPA, Universitas Negeri Semarang
Kampus Sekaran Gunungpati Semarang, Indonesia

¹mirqotussaadah@students.unnes.ac.id

²a212muslim@mail.unnes.ac.id, ³endanghs02@yahoo.com

⁴budipras@mail.unes.ac.id, ⁵siti_alimah@mail.unnes.ac.id

Abstrak

Pada bidang kesehatan, data mining dapat dimanfaatkan untuk memprediksi suatu penyakit dari data rekam medis pasien, diantaranya diabetes. Ada beberapa model data mining salah satunya klasifikasi. Di bidang klasifikasi, ada banyak cabang yang berkembang yaitu pohon keputusan (decision tree). Salah satu decision tree yang populer adalah C4.5. Dalam riset ini, data yang digunakan adalah pima indian diabetes dataset yang diambil dari UCI repository of machine learning. Pada dataset ini seluruh atributnya bertipe numerik yang bersifat continuous dan untuk menangani data continuous digunakan discretization. Akurasi sangat penting dalam pengklasifikasian, ensemble method adalah metode yang digunakan untuk meningkatkan akurasi algoritma klasifikasi dengan membangun beberapa classifier dari data training. Dari hasil penelitian, dengan menerapkan discretization dan teknik bagging untuk klasifikasi berbasis ensemble pada algoritma C4.5 dapat meningkatkan akurasi sebesar 6,26%. Dengan akurasi awal 68,61%, setelah diterapkan discretization dan teknik bagging menjadi 74,87%.

Kata kunci: Data mining, Decision tree, C4.5, Discretization, Ensemble, Bagging, Diabetes.

Abstract

In the field of health, data mining can be used to predict a disease from patient medical record data, diabetes. There are several data mining models which one is classification. In the access field, there are many branches that are developing the decision tree (decision tree). One popular decision tree is C4.5. In this study, the data used were pima indian diabetes dataset taken from UCI machine learning repository. In this dataset all attributes are of continuous numeric type and for combined continuous data discretization is used. Accuracy is very important in the classification, ensemble method is a method used to improve the accuracy of classification algorithm by building some classifier of training data. From the research results, by applying discretization and bagging techniques to ensemble-based classification on C4.5 algorithm can increase the accuracy of 6.26%. With an initial accuracy of 68.61%, after applied discretization and bagging techniques to 74.87%.

Keywords: Data mining, Decision tree, C4.5, Discretization, Ensemble, Bagging, Diabetes.

1. Pendahuluan

Dengan kemajuan teknologi informasi dewasa ini, kebutuhan akan informasi yang akurat sangat dibutuhkan dalam kehidupan sehari-hari. Namun, sering kali informasi tersebut masih harus digali ulang dari data yang jumlahnya sangat besar. Oleh karena itu data tersebut dapat dimanfaatkan atau diolah menjadi pengetahuan dengan istilah yang sering disebut *data mining*. Klasifikasi merupakan bagian penting dari pembelajaran mesin dan aplikasi data mining. Ada banyak metode yang berbeda untuk membandingkan dan menentukan hasil klasifikasi terbaik [1].

Di bidang klasifikasi, ada banyak cabang yang berkembang yaitu pohon keputusan (*decision tree*), klasifikasi *Bayesian*, jaringan saraf dan algoritma genetika [2]. Dalam penelitiannya mengatakan *decision tree* memang populer dan sering digunakan dalam klasifikasi karena memiliki hasil yang cukup baik jika dibanding algoritma lainnya. Salah satu *decision tree* yang populer adalah C4.5 [2].

Pada bidang kesehatan, *data mining* dapat dimanfaatkan untuk memprediksi suatu penyakit dari data rekam medis pasien. Dengan metode klasifikasi pada *data mining*, data seperti umur, jenis kelamin, tekanan darah dan atribut lainnya, dapat digunakan menjadi faktor pendukung dalam memprediksi kemungkinan pasien terkena suatu penyakit [3]. Algoritma klasifikasi *data mining* tersebut dapat dimanfaatkan dan membantu ahli medis dalam mendiagnosa suatu penyakit, salah satunya adalah *diabetes*. Penyakit diabetes merupakan salah satu penyakit yang mematikan, faktor resiko tinggi dalam keluarga yang menyebabkan penyakit *diabetes* antara lain dikarenakan orang gemuk yang tidak melakukan latihan fisik, dan orang-orang yang memiliki gaya hidup yang tidak sehat dan makanan yang berlebih dari apa yang dibutuhkan oleh tubuh [4].

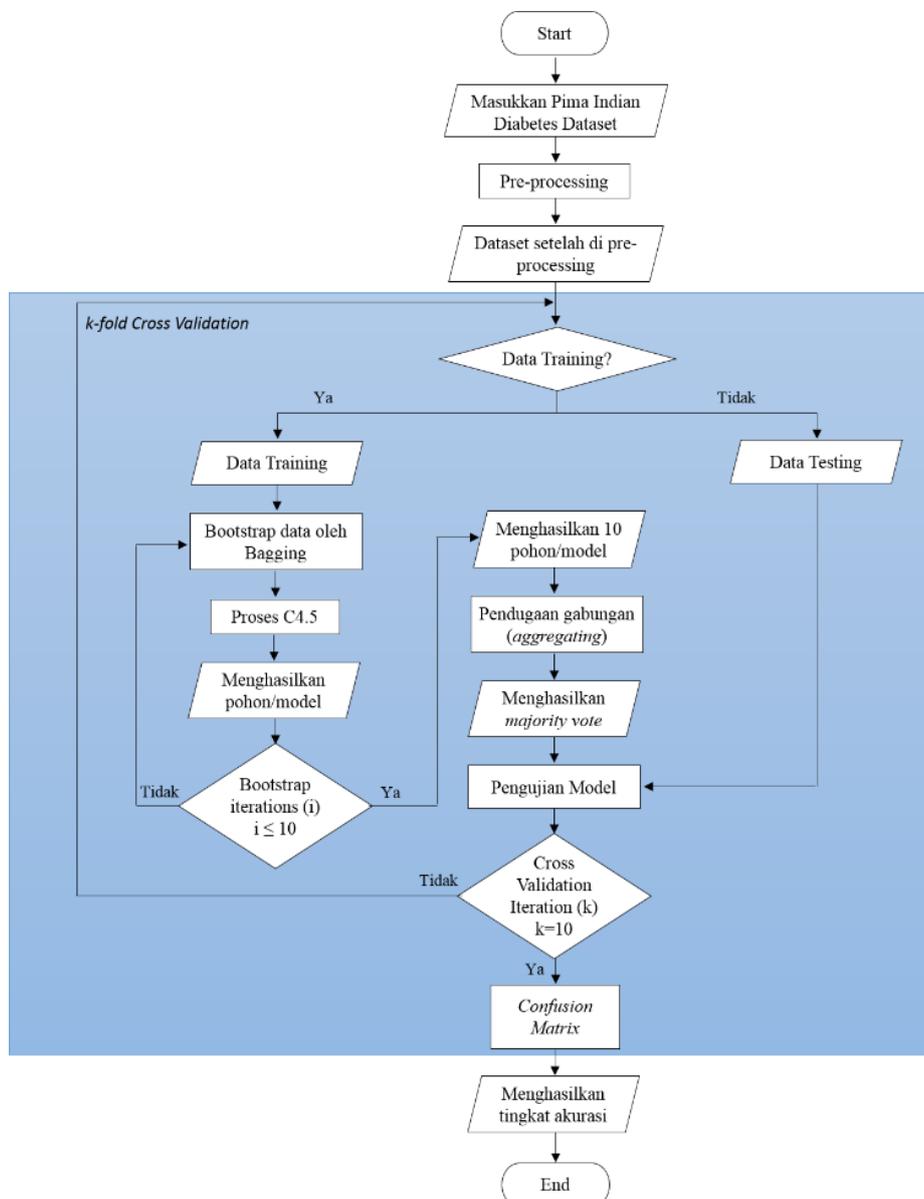
Penelitian diagnosa dilakukan dengan menggunakan *dataset pima indian diabetes*. *Dataset* yang digunakan dalam penelitian ini adalah *dataset pima indian diabetes* yang diperoleh dari UCI *repository of machine learning datasets*. Proses pengolahan dan ekstraksi pengetahuan dari data menjadi tugas penting yang sering dilakukan oleh algoritma pembelajaran. Salah satu tugas yang paling umum dilakukan oleh algoritma pembelajaran adalah pembuatan aturan klasifikasi dari contoh kelas berlabel [5]. Contohnya dijelaskan oleh serangkaian atribut numerik, nominal, atau kontinyu. Banyak algoritma pembelajaran dirancang secara jelas untuk menangani data numerik atau nominal, sementara beberapa algoritma berperforma lebih baik dengan atribut bernilai diskrit terlepas dari kenyataan bahwa mereka juga dapat menangani atribut kontinyu [6]. Sebagian besar aplikasi *data mining* hanya bisa diterapkan pada fitur diskrit. Namun, data di dunia nyata seringkali bersifat kontinyu. Bahkan untuk algoritma yang bisa langsung menangani fitur kontinyu, pembelajaran seringkali kurang efisien dan efektif. Oleh karena itu *discretization* membahas masalah ini dengan menemukan interval angka yang lebih ringkas untuk diwakili dan ditentukan [7]. Salah satu contohnya, algoritma pohon keputusan ID3 bekerja dengan baik pada masalah klasifikasi memiliki dataset dengan nilai-nilai diskrit [8].

Diskritisasi atribut kontinyu merupakan salah satu langkah pengolahan data penting dari ekstraksi pengetahuan. Metode diskritisasi yang efektif tidak hanya dapat mengurangi permintaan memori sistem dan meningkatkan efisiensi algoritma data mining dan *machine learning*, namun juga membuat pengetahuan yang diambil dari *dataset discretized* lebih ringkas, mudah dipahami dan digunakan [7]. *Discretization* adalah proses mengubah nilai atribut kontinyu menjadi sejumlah interval yang terbatas dan berasosiasi dengan masing-masing interval dengan nilai numerik yang diskrit. Proses diskritisasi dilakukan sebelum proses pembelajaran [5].

Akurasi sangat penting dalam pengklasifikasian. *Ensemble method* adalah metode yang digunakan untuk meningkatkan akurasi algoritma klasifikasi dengan membangun beberapa *classifier* dari data *training* kemudian pada saat klasifikasi metode ini menggunakan *voting/aggregating* dari *classifier-classifier* tersebut [9]. Salah satu contoh dari *ensemble method* adalah *bootstrap aggregating* yang biasa disingkat "*bagging*". Metode bagging adalah metode resampling yang telah berhasil diterapkan pada area supervised learning untuk meningkatkan akurasi prediksi [10].

2. Metodologi Penelitian

Metodologi yang digunakan pada penelitian ini terdiri dari beberapa tahap yaitu studi pendahuluan, pengumpulan data, analisis data, perancangan sistem dan penarikan kesimpulan. Studi pendahuluan yaitu merumuskan perumusan permasalahan penelitian. Selanjutnya ke tahap pengumpulan data, pada tahap ini menggunakan studi pustaka dan metode dokumentasi. Studi pustaka adalah pengumpulan bahan, informasi, keterangan dan teori dalam buku serta konsultasi dengan para ahli atau narasumber serta rujukan dari artikel, jurnal dan karya ilmiah lainnya yang relevan dan berhubungan dengan objek penelitian, serta metode yang digunakan dalam penelitian. Metode dokumentasi adalah teknik pengumpulan data yang diambil dari sumber atau sudah tersedia tanpa melakukan pencatatan data. Pada penelitian ini digunakan *dataset pima indian diabetes* yang diambil dari *UCI repository of machine learning datasets*. Selanjutnya tahap analisis data, yaitu menjelaskan teknik penyelesaian metode yang digunakan dalam penelitian. Pada penelitian ini teknik analisis data yang dijelaskan adalah metode *discretization* pada *pre-processing* dalam menangani atribut yang bersifat *continuous* dan teknik *bagging* pada algoritma C4.5 yang diilustrasikan pada Gambar 1.



Gambar 1. Flowchart Penggabungan Algoritma C4.5 dan Teknik Bagging

Pada Gambar 1 pengolahan awal atau yang biasa disebut tahap *pre-processing*, yaitu data yang sudah didapat dibersihkan dan dipilah sehingga mendapatkan data baru untuk data *training* dan data *testing* dari atribut yang sudah ditentukan. Pembagian Training dan testing menggunakan *k-fold cross validation* dengan *default k=10*. Pada data *training*, lakukan *bootstrap* data sebanyak 10 kali. Dan pada setiap *bootstrap* dilakukan proses C4.5 yang akan dijelaskan pada Gambar 2, sehingga diperoleh 10 buah pohon klasifikasi. Selanjutnya, tahap *aggregating* yaitu dilakukan pendugaan gabungan berdasarkan hasil pendugaan 10 buah pohon klasifikasi yang terbentuk menggunakan aturan *majority vote* (suara terbanyak) sehingga mendapatkan model baru yang akan digunakan untuk pengujian data *testing*. Hasil prediksi gabungan yang telah diperoleh selanjutnya digunakan untuk menguji ketepatan klasifikasi pada penerapan metode *bagging*. Uji ketepatan klasifikasi dilakukan menggunakan *confusion matrix*. Pengukuran akurasi dengan *confusion matrix* dapat dilihat pada Tabel 1 [11].

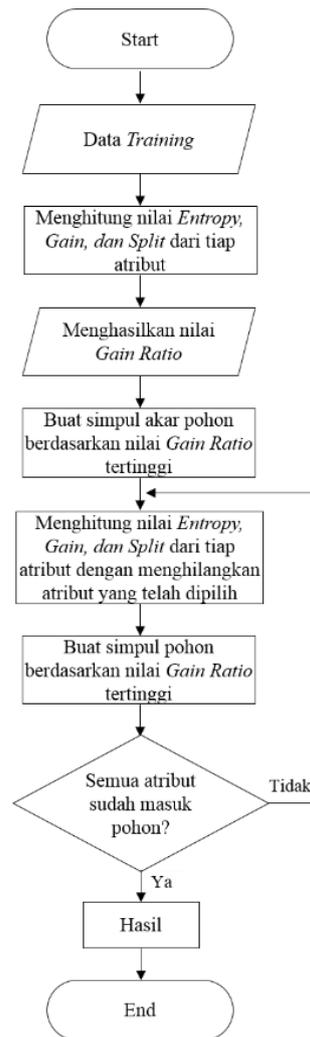
Tabel 1. Confusion Matrix

		Predicted Class	
		Positive	Negative
Actual Class	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

Formulasi perhitungan yang digunakan adalah sebagai berikut :

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \tag{1}$$

Tahapan alur kerja proses algoritma C4.5 diilustrasikan pada Gambar 2.



Gambar 2. Flowchart Algoritma C4.5

Berikut penjelasan dari tiap proses yang terdapat pada Gambar 2, yaitu menghitung nilai *entropy*, *gain* dan *split* dari masing-masing atribut data *training* yang ada sehingga menghasilkan *gain ratio*. Rumus untuk menghitung *entropy*, *gain*, *split* dan *gain ratio* bisa dilihat pada persamaan 2, 3, 4 dan 5 berikut [12]:

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i \quad (2)$$

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (3)$$

$$SplitEntropy_A(S) = - \sum_{i=1}^n \frac{|S_i|}{|S|} * \log_2 \frac{|S_i|}{|S|} \quad (4)$$

$$GainRatio(A) = \frac{Gain(A)}{SplitEntropy(A)} \quad (5)$$

Dimana:

- S = Himpunan Kasus
- A = Atribut
- n = jumlah partisi atribut A
- |S_i| = jumlah kasus pada partisi ke-i
- |S| = jumlah kasus dalam S
- Ph = Proporsi dari S_i terhadap S

Atribut yang memiliki *gain ratio* terbesar dipilih untuk membuat simpul akar. Selanjutnya, menghitung nilai *entropy*, *gain* dan *split* dari masing-masing atribut dengan menghilangkan atribut yang telah dipilih sebelumnya. Atribut yang memiliki *gain ratio* terbesar dipilih untuk membuat simpul internal. Ulangi perhitungan tersebut hingga semua atribut memiliki kelas. Jika semua atribut/pohon sudah memiliki kelas, maka tampilkan pohon keputusan awal dan *generate* aturan keputusan awal.

Selanjutnya, ke tahap perancangan sistem, pada penelitian ini perancangan sistem dilakukan sebagai alat uji penerapan metode *discretization* dalam menangani atribut yang bersifat *continuous* dan teknik *bagging* untuk meningkatkan akurasi klasifikasi berbasis *ensemble* pada Algoritma C4.5 dalam mendiagnosa *diabetes* dengan menggunakan *Rapidminer*. Tahap yang terakhir adalah penarikan kesimpulan, simpulan dalam penelitian ini adalah mendapatkan model dan membandingkan tingkat akurasi algoritma C4.5 yang menerapkan metode *discretization* dan teknik *bagging* dalam mendiagnosa *diabetes* sehingga didapatkan metode yang terbaik.

3. Hasil dan Pembahasan

Pada penelitian ini, data yang digunakan adalah *dataset* yang diambil dari UCI *repository of machine learning*. *Pima indian diabetes dataset* ini terdiri dari 768 data klinis yang semuanya berasal dari jenis kelamin wanita dengan umur sekurang-kurangnya 21 tahun yang terbagi dalam 8 atribut dan 1 atribut kelas yang digunakan untuk mendiagnosa ada tidaknya diabetes pada seseorang. *Pima indian diabetes dataset* terdiri dari 8 data *numeric* dan 1 data *nominal*. Tabel 2 menunjukkan deskripsi atribut dari *dataset pima indian diabetes*.

Tabel 2. Atribut Dataset Pima Indian Diabetes

Atribut	Singkatan	Deskripsi	Satuan	Tipe Data
Pregnant	Pregnant	Angka kehamilan	-	Numerik
Plasma-Glucose	Glucose	Kadar glukosa 2 jam setelah makan	Mg/dL	Numerik
Diastolic Blood-Pressure	DBP	Tekanan darah	Mm Hg	Numerik
Triceps Skin Fold Thickness	TSFT	Ketebalan kulit	Mm	Numerik
Insulin	INS	Insulin	Mu U/ml	Numerik
Body Mass Index	BMI	Berat Tubuh	Kg/m ²	Numerik
Diabetes Pedigree Function	Pedigree	Riwayat Diabetes dalam keluarga	-	Numerik
Age	Age	Umur	Years	Numerik
Class Variable	Class	Positif diabetes (1) dan Negatif diabetes (0)	-	Nominal

Sebelum dilakukan proses klasifikasi menggunakan algoritma C4.5, data harus dipersiapkan terlebih dahulu agar siap untuk diolah (dikenal dengan istilah *pre-processing*) dengan tujuan meminimalkan kesalahan dan mengoptimalkan hasil *mining* dari *classifier* yang digunakan. Tahap persiapan sesuai dengan tahapan data *mining*. Tabel 3 menunjukkan beberapa data dari *dataset pima Indian diabetes*.

Tabel 3. Dataset Pima Indian Diabetes

Preg	Plas	Pres	Skin	Insu	Mass	Pedi	Age	Class
7	136	74	26	135	26	0,647	51	tested_negative
5	155	84	44	545	38,7	0,619	34	tested_negative
1	119	86	39	220	45,6	0,808	29	tested_positive
4	96	56	17	49	20,8	0,340	26	tested_negative
5	108	72	43	75	36,1	0,263	33	tested_negative
0	78	88	29	40	36,9	0,434	21	tested_negative

Pada tahap *cleaning*, karena dalam *record* data terdapat data yang bernilai kosong atau dapat disebut juga dengan data yang salah (*missing value*). Oleh karena itu, diberikan perlakuan pada

data yang salah (*missing value*). *Missing value* dapat diisi secara manual dan juga dapat diisi berdasarkan model *average* yaitu dengan menggantikan nilai yang kosong dengan nilai rata-rata berdasarkan nilai yang tersedia untuk fitur tersebut [13].

Selanjutnya, pada tahap transformasi data, dataset di-diskritisasi menggunakan *Entropy-Based Discretization*. Ini dilakukan untuk mentransformasikan data bertipe *numeric* ke dalam pola yang dapat dikenali guna mendapatkan nilai entropi untuk mencari nilai *Information Gain* (Persamaan 3) dari atribut bertipe *numeric*. Diskritisasi atribut bertujuan untuk menyederhanakan permasalahan dan meningkatkan akurasi dalam proses pembelajaran. Tabel 4 menunjukkan diskritisasi atribut dataset pima Indian diabetes dan Tabel 5 menunjukkan data setelah didiskritisasi.

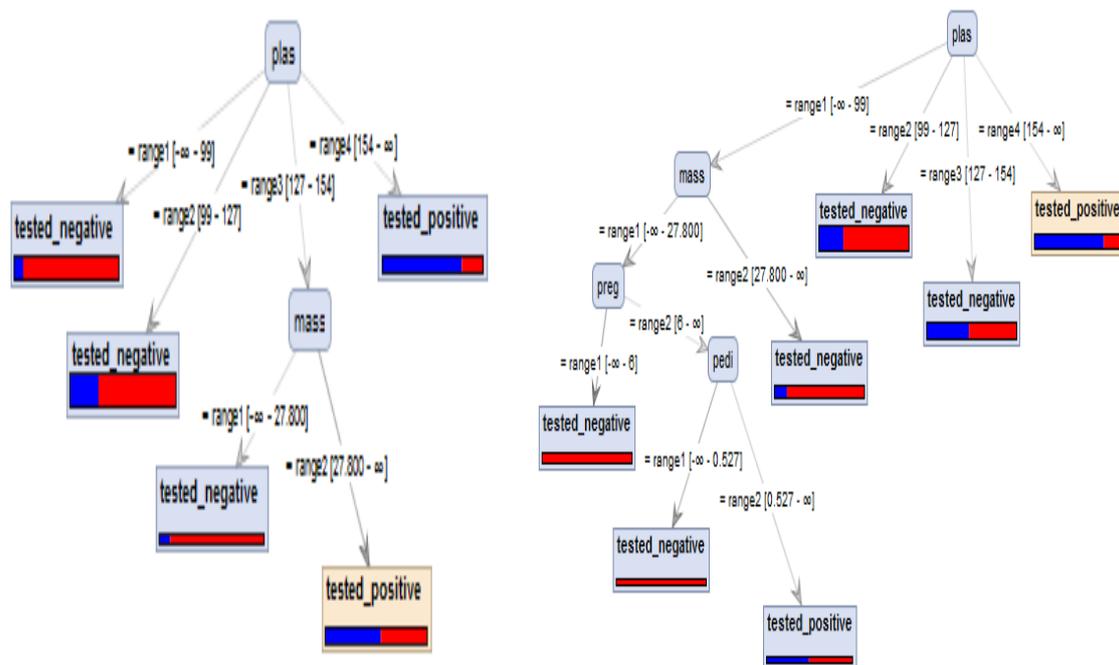
Tabel 4. Diskritisasi atribut dataset pima Indian diabetes

Atribut	Diskritisasi
Pregnant	(≤ 6), (> 6)
Plasma-Glucose	(≤ 127), (> 127)
Diastolic Blood-Pressure	(≤ 68), (> 68)
Triceps Skin Fold Thickness	(≤ 23), (> 23)
Insulin	(≤ 87), (> 87)
Body Mass Index	($\leq 27,3$), ($> 27,3$)
Diabetes Pedigree Function	($\leq 0,527$), ($> 0,527$)
Age	(≤ 28), (> 28)

Tabel 5. Data setelah didiskritisasi

Preg	Plas	Pres	Skin	Insu	Mass	Pedi	Age	Class
> 6	> 127	> 68	> 23	> 87	$\leq 27,3$	$> 0,527$	> 28	tested_negative
≤ 6	> 127	> 68	> 23	> 87	$> 27,3$	$> 0,527$	> 28	tested_negative
≤ 6	≤ 127	> 68	> 23	> 87	$> 27,3$	$> 0,527$	> 28	tested_positive
≤ 6	≤ 127	≤ 68	≤ 23	≤ 87	$\leq 27,3$	$\leq 0,527$	≤ 28	tested_negative
≤ 6	≤ 127	> 68	> 23	≤ 87	$> 27,3$	$\leq 0,527$	> 28	tested_negative
≤ 6	≤ 127	> 68	> 23	≤ 87	$> 27,3$	$\leq 0,527$	≤ 28	tested_negative

Setelah melalui tahap *pre-processing*, selanjutnya dilakukan tahap proses data mining. Penelitian ini diimplementasikan kedalam *tool Rapidminer*. *Rapidminer* adalah koleksi dari algoritma *learning machine* yang digunakan untuk tugas-tugas *data mining*. *Rapidminer* berisi *tool* untuk data *pre-processing*, klasifikasi, regresi, *clustering*, *rule association*, dan memvisualisasikan data tersebut menjadi mudah untuk dapat dipahami. Pada bagian ini, hasil eksperimen dianalisis untuk mengevaluasi kinerja algoritma *data mining* yang diusulkan. Fitur yang dipilih yaitu *discretization* untuk menangani atribut *continuous (numeric)* dan teknik *bagging* untuk klasifikasi berbasis *ensemble* pada algoritma C4.5 guna meningkatkan akurasi dalam mendiagnosa *diabetes*. Gambar 3 menunjukkan pohon klasifikasi yang terbentuk pada proses *bagging*.



Gambar 3. Hasil pohon yang dihasilkan *bagging*

Hasil prediksi gabungan yang telah diperoleh selanjutnya digunakan untuk menguji ketepatan klasifikasi pada penerapan metode *bagging*. Uji ketepatan klasifikasi dilakukan menggunakan matriks konfusi pada Tabel 6. Sedangkan Tabel 7 menunjukkan presentase keakuratan dari algoritma C4.5 dalam mendiagnosa *diabetes*.

Tabel 6. Matriks konfusi hasil klasifikasi

	Prediksi	
	Tested_positive	Tested_negative
Tested_positive	113	38
Tested_negative	155	462

Tabel 7. Tingkat akurasi algoritma C4.5

Data	C4.5	C4.5 + Bagging	Discretization + C4.5	Discretization + C4.5 + Bagging
Diabetes	68,61%	69,79%	74,61%	74,87%

Hasil eksperimen dengan menggunakan C4.5 didapatkan akurasi sebesar 68,61%. Eksperimen dengan menambahkan discretization didapatkan akurasi sebesar 74,61%, mengalami kenaikan akurasi sebesar 6,26%. Sedangkan hasil eksperimen ketiga yaitu dengan menambahkan teknik bagging pada algoritma C4.5 didapatkan akurasi sebesar 69,79%. Hasil dari kedua eksperimen tersebut menjelaskan bahwa dengan menerapkan *discretization* ataupun teknik *bagging* dapat meningkatkan akurasi. *Discretization* dapat meningkatkan akurasi dikarenakan *discretization* memperbaiki kualitas data sebelum dilakukan proses *learning* dilakukan. Sedangkan *bagging* dapat meningkatkan akurasi karena mengurangi *variance* dan *overfitting* pada model. *Bagging* cocok untuk algoritma yang sifatnya *unstable learning algorithms* dimana model akan berubah jika data *training*-nya ikut dirubah, contohnya adalah algoritma CART dan C4.5. Oleh karena itu, hasil akhir dari eksperimen dengan menerapkan *discretization* dan teknik *bagging* pada algoritma klasifikasi dapat meningkatkan performa akurasi secara signifikan pada algoritma C4.5 yaitu didapatkan akurasi sebesar 74,87%.

4. Kesimpulan

Dari empat kali eksperimen, yaitu menggunakan algoritma C4.5 menghasilkan akurasi 68,61%. Eksperimen kedua dengan menambahkan teknik *bagging* pada algoritma C4.5 menghasilkan akurasi 69,79%. Selanjutnya eksperimen ketiga yaitu menambahkan teknik *bagging* pada algoritma C4.5 menghasilkan akurasi 74,61%. Dari ketiga eksperimen, teknik *discretization* dan teknik *bagging* terbukti efektif dapat meningkatkan hasil akurasi algoritma C4.5 pada klasifikasi *dataset diabetes*. Teknik *discretization* selain dapat mengubah atribut *continuous* menjadi diskrit juga dapat meningkatkan hasil akurasi. Selanjutnya, eksperimen keempat yaitu dengan menggabungkan teknik *discretization* dan teknik *bagging* pada algoritma C4.5 menghasilkan akurasi sebesar 74,87%. Dari hasil penelitian dengan menggunakan *discretization* dan teknik *bagging* pada algoritma C4.5 menunjukkan peningkatan 6,26%. Dengan akurasi awal 68,61%, setelah diterapkan *discretization* dan teknik *bagging* menjadi 74,87%. Dapat disimpulkan bahwa penerapan *discretization* dan teknik *bagging* dapat meningkatkan akurasi algoritma C4.5 pada klasifikasi *dataset diabetes*.

Daftar Pustaka

- [1] R. Das, "A Comparison Of Multiple Classification Methods For Diagnosis Of Parkinson Disease," *Expert Systems with Applications*, vol. 37, no 2, pp.1568-1572, 2010.
- [2] C. J. Tsai, C. I. Lee, And W. P. Yang, "A Discretization Algorithm Based On Class-Attribute Contingency Coefficient," *Information Sciences*, vol. 178, no. 3, pp.714-731, 2008.
- [3] R. A. Muzakir, A., & Wulandari, "Model Data Mining Sebagai Prediksi Penyakit Hipertensi Kehamilan Dengan Teknik Decision Tree," *Scientific Journal of Informatics*, vol. 3, no. 1, pp. 19–26, 2016.
- [4] S. M. Nuwangi, C. R. Oruthotaarachchi, J. M. P. P. Tilakaratna, And H. A. Caldera, "Utilization Of Data Mining Techniques In Knowledge Extraction For Diminution Of Diabetes," In *Proceedings - 2nd Vaagdevi International Conference On Information Technology For Real World Problems, VCON*, pp.3-8, 2010.
- [5] A. Al-Ibrahim, "Discretization Of Continuous Attributes In Supervised Learning Algorithms," *Res. Bull. Jordan Acm*, 2011.
- [6] R. Kerber, "Chimerge: Discretization Of Numeric Attributes," . In *Proceedings of the tenth national conference on Artificial intelligence. Aai Press*, 1992.
- [7] Dash, R., Paramguru, R. L., & Dash, R, "Comparative Analysis Of Supervised And Unsupervised Discretization Techniques," *International Journal of Advances in Science and Technology*, Vol. 2, No. 3, Pp. 29–37, 2011.
- [8] A. A. Nurcahyani And R. Saptono, "Identifikasi Kualitas Beras Dengan Citra Digital," *Scientific Journal of Informatics*, vol. 2, no.1, pp.63-72, 2016.
- [9] K. Tan, Pang, N., Michael, S. & Vipin, *Introduction To Datamining*. Boston: Pearson Addison Wesley, 2006.
- [10] O. Somantri, G. W. Sasmito, M. S. Sungkar, And Erwadi, "Optimalisasi Neural Network Dengan Bootstrap Aggregating (Bagging) Untuk Penentuan Prediksi Harga Listrik," *Scientific Journal of Informatics*, vol. 1, no.2, pp.185-192, 2014.
- [11] F. Gorunescu, *Data Mining: Concepts And Techniques*, 1st Ed. Verlag Berlin Heidelberg: Springer, 2011.
- [12] Prasetyo, "Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab," CV. Andi Offset, 2014.
- [13] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, Waltham, MA: Elsevier/Morgan Kaufmann, 2012.