

Perbandingan Kinerja Sistem Klasifikasi Berbasis *K-Fold Cross Validation* pada Algoritma *Decision Tree ID3* dan *C5.0*

Muhammad Yusuf Ashari¹, Abd Charis Fauzan², Huda Maariful Muhamat³

Program Studi Ilmu Komputer, Universitas Nahdlatul Ulama Blitar, Indonesia

Email: ¹yusufashari202@gmail.com, ²abdcharis@unublitar.ac.id, ³muhamatmaarif@unublitar.ac.id

INFORMASI ARTIKEL

Histori artikel:

Naskah masuk, 22 Mei 2022

Direvisi, 08 Juni 2022

Diiterima, 10 Juni 2022

Kata Kunci:

Decision Tree ID3,
Decision Tree C5.0,
K-fold Cross Validation,
Perbandingan Kinerja

ABSTRAK

Abstract- In data mining, many techniques and methods have been carried out in predictive models, classification methods for example, one of which is the Decision Tree or decision tree including CART, ID3, C4.5 C5.0 and others. Of the many types of decision trees, they must have different levels of precision and accuracy. Of course there is a need for research that discusses the comparison of each method. In this study, the author evaluates the performance of the classification and comparison system of the ID3 Decision Tree Algorithm which is the first generation method with the Decision Tree C5.0 Algorithm, where the C5.0 Decision Tree Algorithm is the latest generation method which is an extension of the C4.5 Decision Tree Algorithm and the C5.0 Algorithm. Decision Tree ID3 based on K-fold Cross Validation Algorithm. Therefore, this research aims to compare the ID3 Decision Tree Algorithm with the C5.0 Decision Tree Algorithm to find out which algorithm has good performance and will be used in predicting data. In this research, 215 datasets of factors that influence campus placement are used. This research AIur starts from data collection, pre-processing, calculation of the ID3 and C5.0 Decision Tree Algorithms and then evaluated using K-fold Cross Validation. The results of this study indicate, through a comparison of the performance of the K-fold Cross Validation-based classification system, the ID3 Decision Tree Algorithm is superior to the C5.0 Decision Tree Algorithm. Decision Tree ID3 algorithm with 95% precision, 94% recall/sensitivity and 93% accuracy. While the Decision Tree C5.0 Algorithm with 91% precision, 92% recall and 89% accuracy.

Abstrak- Pada data mining, terdapat teknik dan metode yang telah dilakukan dalam model prediksi, metode klasifikasi contohnya, salah satunya adalah *Decision Tree* atau pohon keputusan diantaranya adalah CART, ID3, C4.5 C5.0 dan lain-lain. Dari banyaknya jenis pohon keputusan, pasti mempunyai tingkat ketepatan dan keakurasian yang berbeda. Tentunya perlu adanya penelitian yang membahas tentang perbandingan dari masing-masing metode. Pada penelitian ini penulis mengevaluasi kinerja sistem klasifikasi dan komparasi Algoritma *Decision Tree ID3* yang merupakan metode generasi pertama dengan Algoritma *Decision Tree C5.0*, dimana Algoritma *Decision Tree C5.0* ini metode generasi terbaru yang merupakan perpanjangan dari Algoritma *Decision Tree C4.5* dan Algoritma *Decision Tree ID3* berbasis Algoritma *K-fold Cross Validation*. Oleh karena itu, dalam penelitian ini bertujuan untuk mengkomparasikan antara Algoritma *Decision Tree ID3* dengan Algoritma *Decision Tree C5.0* untuk mengetahui Algoritma mana yang baik kinerjanya dan akan digunakan dalam memprediksi data. Dalam penelitian ini menggunakan dataset factor yang mempengaruhi penempatan kampus sebanyak 215 data. AIur penietian ini dimulai dari pengumpulan data, pre-prosesing, perhitungan Algoritma *Decision Tree ID3* dan C5.0 kemudian dievaluasi menggunakan *K-fold Cross Validation*. Hasil penelitian ini menunjukkan, melalui perbandingan kinerja sistem klasifikasi berbasis *K-fold Cross Validation* Algoritma *Decision Tree ID3* lebih unggul daripada Algoritma *Decision Tree C5.0*. Algoritma *Decision*

Tree ID3 dengan hasil presisi 95%, recall/sensitivity 94% dan akurasi 93%. Sedangkan Algoritma *Decision Tree* C5.0 dengan hasil presisi 91%, recall 92% dan akurasi 89%.

Copyright © 2019 LPPM - STMIK IKMI Cirebon
This is an open access article under the CC-BY license

Penulis Korespondensi:

Abd. Charis Fauzan

Program Studi Ilmu Komputer,
Universitas Nahdlatul Ulama Blitar
Jl. Masjid No.22, Kota Blitar, Indonesia
Email: abdcharis@unublitar.ac.id

1. Pendahuluan

Ada banyak jenis *Decision Tree* atau pohon keputusan diantaranya adalah ID3, *C4.5* C.50 dan lain-lain. *Algoritma Iterative Dichotomiser Three* (ID3) merupakan Algoritma *Decision Tree* yang paling dasar. ID3 digunakan untuk menentukan nilai bobot dari setiap faktor. Setelah itu menyeleksi alternatif terbaik dari beberapa kombinasi faktor alternatif. Metode *Decision Tree* ini menentukan setiap seleksi yang akan dipilih, hingga menampilkan semua kombinasi kemungkinan dan menampilkan alternatif yang terbaik. (Kristiyani [1]).

Algoritma *Decision Tree C4.5* ialah lanjutan dari Algoritma *Decision Tree* ID3, dimana pengembangan dilakukan dalam berbagai hal seperti, mengatasi data kontinu, pruning dapat mengatasi dan missing data. Sedangkan Algoritma *Decision Tree C5.0* merupakan pengembangan dari Algoritma *Decision Tree C4.5*. Di Algoritma ini semua kesalahan dianggap tidak ada perbedaan. tidak ada pemisahan kesalahan berdasarkan kepentingan atau berdasarkan signifikansinya. Jelas peningkatan Algoritma *Decision Tree C5.0* dibandingkan Algoritma sebelumnya adalah bahwa ia memperlakukan semua kesalahan dengan klasifikasi individu berdasarkan pada besarnya dampaknya terhadap sistem. Itu membangun pengklasifikasi yang bisa membantu mengurangi biaya kesalahan klasifikasi yang tinggi. Karakteristik ini dari Algoritma *Decision Tree C5.0* dikenal sebagai biaya kesalahan klasifikasi variabel. Masalah ini juga dapat diatasi dengan sangat baik di Algoritma *Decision Tree C5.0* dengan cara menerapkan karakteristik yang disebut case dengan atribut berat. Dengan menggunakan fitur ini, Algoritma *Decision Tree C5.0* dapat mengurangi biaya salah perhitungan prediktif yang bias (Arif [2]).

(Defiyanti [3]) telah menyelesaikan penelitian dengan membandingkan Algoritma *Decision Tree* ID3 dan melakukan penelitian dengan membandingkan kinerja Algoritma *Decision Tree*

C4.5 dalam data klasifikasi pesan beraturan. Kinerja sistem diukur sesuai jumlah atribut dan jumlah data. Hasil dari penelitian ini dapat disimpulkan bahwa Algoritma *Decision Tree* ID3 memiliki kinerja (*precision*, *recall/sensitivity*, dan *accuracy*) yang lebih baik daripada Algoritma *Decision Tree C4.5*. (Galathya [4]) (2012) telah menyelesaikan penelitian dengan mengembangkan Algoritma *Decision Tree C4.5* dan mengkomparasikannya dengan Algoritma *Naive Bayes*, *Sequential Minimal Optimization (SMO)*, dan *multilayer perceptron*, dengan menggunakan beberapa dataset. Pengujian akurasi berdasarkan *Relative Absolute Error* dan *Correctly Classified Instances*. Hasil dari penelitian ini, Algoritma *Decision Tree C4.5* memiliki tingkat akurasi sekitar 1-3% lebih baik pada 7 dataset dari total 9 dataset. Pada penelitian (Wibowo [5]) yang mengkomparasikan Algoritma *Decision Tree* dengan *naive bayes*. Hasilnya, *Decision Tree* lebih baik dari pada Algoritma *naive bayes* berdasarkan tingkat akurasi.

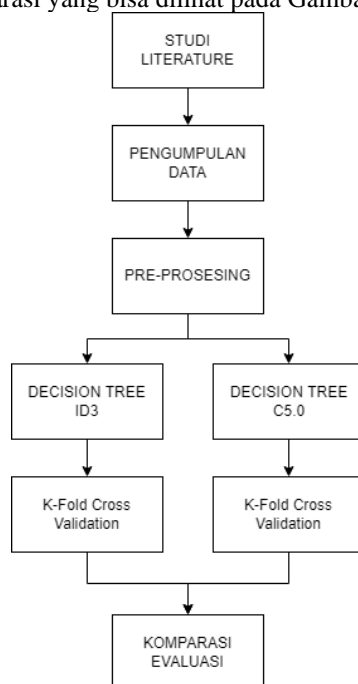
Dalam penelitian kali ini penulis mengevaluasi sistem klasifikasi dan komparasi Algoritma *Decision Tree* ID3 yang merupakan lanjutan dari Algoritma *Decision Tree C4.5* dengan Algoritma *Decision Tree C5.0* menggunakan Algoritma *K-fold Cross Validation*. dalam klasifikasi data yang bisa mengurangi estimasi waktu komputasi serta menjaga keakuratan estimasi model. Seperti yang dikatakan oleh (Hsi-jen Chiang [6]) Dalam kasus klasifikasi, ada yang perlu diperhatikan dalam proses validasi silang, semua data dibagi menjadi k lipatan yang sama (atau hampir sama. Dengan kata lain, distribusi kelas setiap partisi harus sama antar kelas, yang berarti juga sama dengan distribusi kelas di set data originalnya.

(Carleo [7]) Dengan metode *K-fold Cross Validation* ini bisa mengurangi waktu komputasi dengan tetap menjaga keakuratan estimasi model. *Crossvalidation* atau yang biasa disebut dengan rotasi estimasi merupakan teknik validasi untuk menilai hasil analisis yang akan menggeneralisasi

kumpulan data bebas. Teknik *Crossvalidation* biasanya dipakai memprediksi model dan memperkirakan keakuratan dari suatu model prediktif saat diaplikasikan dalam praktiknya. Salah satunya ialah *k-fold crossvalidation*, yaitu membagi data menjadi K partisi set data dengan jumlah data yang sama(Crammer [8]). Adapun sumber *data sets* dari www.kaggle.com dimana setiap dataset memiliki jumlah data (*attributes*) dan jumlah atribut (*instances*) yang tidak sama. Dimana *K-fold Cross Validation* akan menghasilkan penilaian *precision*, *recall*, dan *accuracy*.

2. Metode Penelitian

Pada metode penelitian ini, dilakukan berdasarkan rancangan diagram alur penelitian yang dimulai dari studi literatur dan pengumpulan data, setelah memperoleh data kemudian memasuki tahap preproesing data yang nantinya akan dimuat dalam *Decision Tree* baik Algoritma *Decision Tree* ID3 maupun Algoritma *Decision Tree* C5.0. Setelah selesai membuat *Decision Tree* dilanjutkan pada metode *K-fold Cross Validation* dan akan dikomparasi yang bisa dilihat pada Gambar 1.



Gambar 1. Alur penelitian

2.1 Studi Literatur

Pada tahap studi literature ini digunakan penulis sebagai bahan referensi untuk melakukan penelitian. Referensi didapat melalui penelitian-penelitian terdahulu atau jurnal nasional/internasional yang berkaitan dengan pembahasan dalam penelitian ini..

2.2 Pengumpulan Data

Pada tahap pengumpulan data ini dilakukan dengan mencari dataset jenis klasifikasi. Dataset

yang digunakan dalam penelitian kali ini diperoleh dari

<https://www.kaggle.com/datasets/benroshan/factors-affecting-campus-placement> . Data yang diambil selanjutnya dilakukan preproesing dahulu agar dapat dilakukan perhitungan pada Algoritma *Decision Tree* ID3 maupun Algoritma *Decision Tree* C5.0 yang selanjutnya dilakukan evaluasi sistem.

2.3 Perhitungan Algoritma *Decision Tree* ID3

ID3 adalah Algoritma berbasis perolehan informasi yang dikembangkan oleh Ross Quinlan. Konstruksi Algoritma ID3 pohon keputusan berdasarkan perolehan informasi yang didapat dari data pelatihan(Ibomoiyee [9]). Algoritma pohon keputusan ID3 dimulai dari simpul akar. *Node* akar adalah salah satu dari atribut terbaik. Kemudian nilai properti dihasilkan sesuai dengan masing-masing cabang. Setiap cabang memiliki dihasilkan simpul baru. Untuk kriteria seleksi atribut terbaik, ID3 menggunakan definisi *information gain* berbasis entropi untuk memilih atribut uji dalam *node*. Entropi mencirikan kemurnian set sampel apapun(Wang [10]). Ringkasnya, kinerja Algoritma *Decision Tree* ID3 ditunjukkan sebagai berikut :

- Pilih seluruh atribut yang tidak terpakai kemudian hitung *entropy*-nya yang berhubungan dengan test sample.
- Ambil atribut nilai entropinya yang paling sedikit.
- Buat simpul yang ada atribut tersebut.

Untuk menghitung Algoritma *Decision Tree* ID3 harus menghitung nilai *entropy* dan *gain* *information*-nya ditunjukkan pada persamaan 1 dan persamaan 2.

$$Entropy(S) = - \sum_{j=1}^k p_j \log_2 p_j \quad (1)$$

dimana :

S : Kumpulan kasus

k : Semua atribut pada variabel *A*

p_j : Ukuran dari *S_j* dan *S*

$$Gain(S, A) = Entropy(S) - \sum_{S_v} \frac{S_v}{S} Entropy(S_v) \quad (2)$$

dimana :

S : Kumpulan kasus

A : Variabel

S_v : Semua kasus pada kategori ke-*v*

S : Semua kasus dalam *S*

2.4 Perhitungan Algoritma *Decision Tree* C5.0

Algoritma *Decision Tree* C5.0 ini merupakan pengembangan Algoritma *Decision Tree* ID3 dan C4.5 yang dibuat oleh Ross Quinlan pada tahun 1987. Dalam Algoritma *Decision Tree* C5.0 atributnya dipilih berdasarkan *gain Ratio*. Algoritma *Decision Tree* C5.0 menghasilkan pohon

3. Analisis Dan Pembahasan

3.1 Hasil Pegumpulan Data

Dataset yang digunakan dalam penelitian kali ini diperoleh dari

<https://www.kaggle.com/datasets/benroshan/factors>

-affecting-campus-placement. Banyak data pada dataset yang diambil yaitu 215. Banyak dataset dibagi menjadi 70% sebagai data latih (training data) dan 30% sebagai data uji (testing data).

Tabel 2. Dataset penelitian

sl_no	gender	ssc_p	ssc_b	hsc_p	hsc_b	hsc_s	degree_p	degree_t	workex	etest_p	specialisation	mba_p	status	salary
1	M	67.00	Others	91.00	Others	Commerce	58.00	Sci&Tech	No	55	Mkt&HR	58.8	Placed	27000
2	M	79.33	Central	78.33	Others	Science	77.48	Sci&Tech	Yes	86.5	Mkt&Fin	66.28	Placed	20000
3	M	65.00	Central	68.00	Central	Arts	64.00	Comm&Mgmt	No	75	Mkt&Fin	57.8	Placed	25000
4	M	56.00	Central	52.00	Central	Science	52.00	Sci&Tech	No	66	Mkt&HR	59.43	Not Placed	
5	M	85.80	Central	73.60	Central	Commerce	73.30	Comm&Mgmt	No	96.8	Mkt&Fin	55.5	Placed	42500
...
215	M	62.00	Central	58.00	Others	Science	53.00	Comm&Mgmt	No	89	Mkt&HR	60.22	Not Placed	

Berdasarkan Tabel 2. terdapat 14 atribut dengan data dengan 6 atribut numerik, dan 8 atribut kategorik.

Informasi atribut dari dataset Tabel 1. yaitu,

1. gender: Jenis Kelamin [M: Pria, F: Wanita]
2. ssc_p: Persentase Pendidikan Menengah
3. ssc_b : Dewan Pendidikan
4. hsc_p: Persentase Pendidikan Menengah Tinggi
5. hsc_b: Dewan Pendidikan
6. hsc_s: Spesialisasi dalam Pendidikan Menengah
7. degree_p: Persentase Gelar
8. degree_t: Jenis gelar
9. workex : bekerja
10. etest_p: Persentase tes kelayakan kerja
11. specialisation: keahlian
12. mba_p : persentase MBA

13. status : status

14. salary : gaji

Dalam penelitian ini perhitungan dibagi menjadi dua pengujian, yaitu pengujian dengan metode *Algotima Decision Tree ID3* dan metode *Algotima Decision Tree C5.0*. Kemudian setelah selesai pengujian dari kedua metode tersebut, hasil dari perhitungan data dari metode tersebut dilanjutkan pada proses validasi data dengan menggunakan metode *K-fold Cross Validation*.

3.2 Hasil Preprocessing data

Tabel 3. Hasil Pre-processing data

sl_no	gender	ssc_p	ssc_b	hsc_p	hsc_b	hsc_s	degree_p	degree_t	workex	etest_p	specialisation	mba_p	status
1	M	67.00	Others	91.00	Others	Commerce	58.00	Sci&Tech	No	55	Mkt&HR	58.8	Placed
2	M	79.33	Central	78.33	Others	Science	77.48	Sci&Tech	Yes	86.5	Mkt&Fin	66.28	Placed
3	M	65.00	Central	68.00	Central	Arts	64.00	Comm&Mgmt	No	75	Mkt&Fin	57.8	Placed
4	M	56.00	Central	52.00	Central	Science	52.00	Sci&Tech	No	66	Mkt&HR	59.43	Not Placed
5	M	85.80	Central	73.60	Central	Commerce	73.30	Comm&Mgmt	No	96.8	Mkt&Fin	55.5	Placed
...
215	M	62.00	Central	58.00	Others	Science	53.00	Comm&Mgmt	No	89	Mkt&HR	60.22	Not Placed

3.3 Hasil Perhitungan Decision Tree ID3

berdasarkan pernyataan Masulloh dan (Fitriyani [13]). Untuk mendapatkan akar dan cabang pada pohon keputusan menggunakan *gain* tertinggi atau informasi *gain*. Menurut (Budiman [14]),

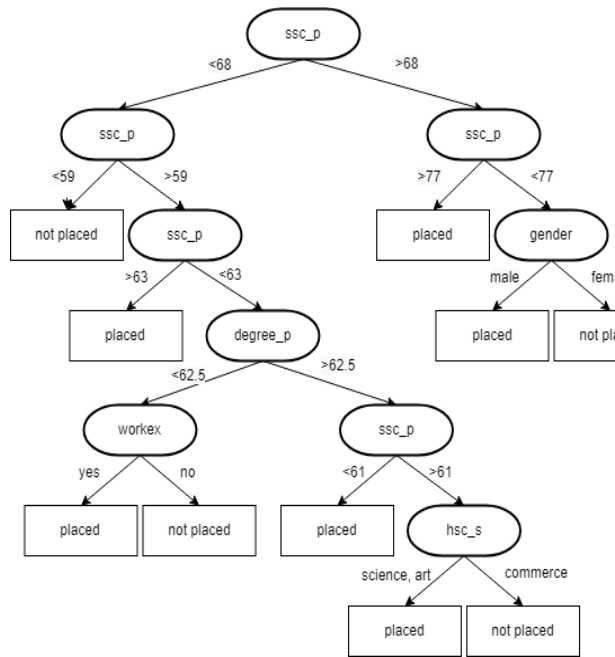
informasi *gain* dapat diperoleh berdasarkan persamaan *entropy*. *Entropy* dihasilkan dari *entropy* seluruh data dan *entropy* pada masing masing atribut yang terdapat pada dataset.

Tabel 4. Perhitungan nilai *gain*

Node 1		Total	Placed	Not Placed	Entropy	Gain
gender	male	110	171	120	51	0,8791358
	female	61	41	20	0,9127342	0,0016645
ssc_p	<68	87	42	45	0,9991421	
	>68	83	77	6	0,3744088	0,18907
ssc_b	Others	76	54	22	1,3735509	
	Central	95	66	29	0,8680404	0,0002124
hsc_p	<66,8	88	47	41	0,8876297	
	>66,8	82	72	10	1,7556701	0,10972
hsc_b	Others	101	72	29	0,996644	
	Central	70	48	22	0,8649833	0,0006128
hsc_s	Commerce	88	61	27	0,8980588	
	Science	74	54	20	1,7630421	
	Arts	9	5	4	0,8894664	0,0049265
degree_p	<66,7	92	49	43	0,8418522	
	>66,7	79	71	8	0,9910761	0,1242576
degree_t	Sci&Tech	48	34	14	1,4699251	
	Comm&Mgmt	116	83	33	0,8708645	
	Others	7	3	4	0,8614872	0,0099509
workex	No	110	68	42	0,9852281	
	Yes	61	52	9	1,7323517	0,0466939
etest_p	<72,1	96	62	34	0,959316	
	>72,1	75	58	17	0,6036522	0,014024
specialisation	Mkt&HR	74	41	33	1,5629683	
	Mkt&Fin	97	79	18	0,9377343	
mba_p	<62,4	88	60	28	0,7721551	
	>62,4	83	60	23	1,7098894	0,057445
					1,6836582	
					0,9915529	
					0,9023933	
					0,8514809	0,001454
					1,7538742	

Dalam membuat sebuah pohon keputusan, membutuhkan *root node*, *internal node* dan *leaf node* agar menyerupai pohon sesungguhnya (Fahma[15]). Menurut Patel dan (Prajapati [16]), *root node* atau akar merupakan induk dari semua *node* dan merupakan *node* tertinggi pohon pada Algoritma *Decision Tree*. Dalam menentukan *root* atau akar dari pohon keputusan ditentukan berdasarkan *gain* tertinggi atau *gain* informasi.

Tabel 4. merupakan hasil perhitungan *gain* pada setiap atribut yang ada pada dataset. Hasil *gain* informasi yang memiliki nilai tertinggi pada *node* pertama digunakan sebagai *root* pada pohon keputusan. Pohon keputusan pada prediksi penerimaan tenaga kerja menghasilkan rule pada Gambar 2. Rule yang didapat berdasarkan pohon keputusan kemudian diimplementasikan pada evaluasi Sistem menggunakan metode *K-fold Cross Validation*.



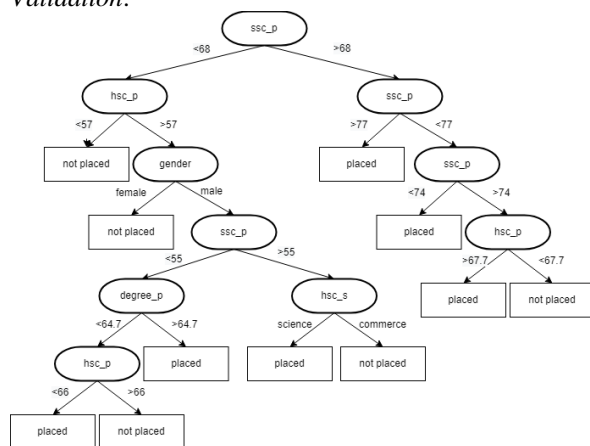
Gambar 2. Pohon Keputusan *Decision Tree* ID3.
 3.4 Hasil Perhitungan *Decision Tree* C5.0

Tabel 5. Perhitungan Nilai *Ratio Gain*

Node 1		Total	Placed	Not Placed	Entropy	Gain	Ratio
gender	male	172	121	51	0,87698	0,001853	0,00105
	female	111	80	31	0,85447		
		61	41	20	0,91273		
ssc_p	<68	87	42	45	0,99914	0,19031	0,1389
	>68	84	78	6	0,37123		
ssc_b	Others	77	55	22	1,37037	0,000327	0,00019
	Central	95	66	29	0,88763		
hsc_p	<66,8	88	47	41	0,99664	0,11096	0,0726
	>66,8	83	73	10	0,53074		
hsc_b	Others	102	73	29	1,52739	0,000747	0,00042
	Central	70	48	22	0,89806		
hsc_s	Commerce	89	62	27	0,88537	0,004805	0,00177
	Science	74	54	20	0,84185		
	Arts	9	5	4	0,99108		
degree_p	<66,7	92	49	43	2,7183	0,125606	0,0857
	>66,7	80	72	8	0,99693		
degree_t	Sci&Tech	48	34	14	1,46593	0,01006	0,00371
	Comm&Mgmt	117	84	33	0,87086		
	Others	7	3	4	0,85823		
workex	No	110	68	42	0,98523	0,048056	0,03087
	Yes	62	53	9	0,95932		
etest_p	<72,1	97	63	34	1,55691	0,01326	0,0078
	>72,1	75	58	17	0,93453		
specialisation	Mkt&HR	74	41	33	1,70668	0,058359	0,03475
	Mkt&Fin	98	80	18	0,99155		

					1,6796		
mba_p	<62,4	88	60	28	0,90239		
	>62,4	84	61	23	0,84688	0,0017	0,001
					1,74927		

Tabel 5. merupakan hasil perhitungan *Ratio gain* pada setiap atribut yang ada pada dataset. Hasil *Ratio gain* informasi yang memiliki nilai tertinggi pada *node* pertama digunakan sebagai *root* pada pohon keputusan. Pohon keputusan pada prediksi penerimaan tenaga kerja menghasilkan rule pada Gambar 3. Rule yang didapat berdasarkan pohon keputusan kemudian diimplementasikan pada evaluasi Sistem menggunakan metode *K-fold Cross Validation*.



Gambar 3. Pohon Keputusan *Decision Tree C5.0*

3.5 Hasil *Decision Tree ID3* Dengan *K-fold Cross Validation*.

	Presisi	Recall	Akurasi
	ID3	ID3	ID3
K-1	100%	100%	100%
K-2	100%	93,75%	95,23%
K-3	93,33%	100%	95,23%
K-4	93,33%	93,33%	90,47%
K-5	93,75%	100%	95,23%
K-6	100%	100%	100%
K-7	100%	100%	100%
K-8	92,85%	86%	85,71%
K-9	80%	80%	80,95%
K-10	94,44%	85%	84,61%

3.5 Hasil *Decision Tree ID3* Dengan *K-fold Cross Validation*.

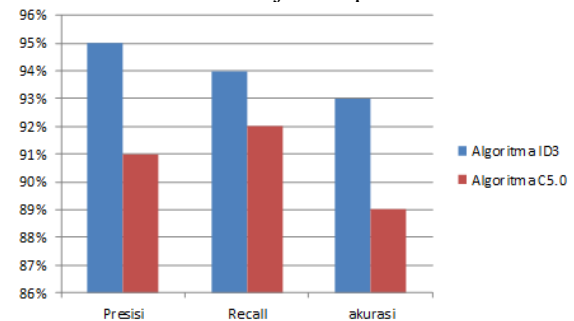
Tabel 7. Hasil C5.0 Dengan 10-fold

	Presisi	Recall	Akurasi
	C5.0	C5.0	C5.0
K-1	92%	100%	95%
K-2	93%	100%	95%
K-3	80%	92%	81%
K-4	100%	93%	95%
K-5	93%	100%	95%
K-6	88%	88%	81%

K-7	100%	100%	100%
K-8	86%	92%	86%
K-9	90%	75%	81%
K-10	83%	83%	77%

3.5 Komparasi dengan *K-fold Cross Validation*

Pada tahap komparasi ini akan dibandingkan hasil evaluasi sistem klasifikasi menggunakan metode *K-Fold Cross Validation* antara Algoritma *Decision Tree ID3* dan Algoritma *Decision Tree C5.0*, dimana keluaran dari metode evaluasi *K-fold Cross Validation* ini akan dihitung dari sisi hasil presisi, recall dan akurasi. Ditunjukkan pada Gambar 4.



Gambar 4. Rata-rata final setiap Algoritma

4. Kesimpulan

Hasil dari penelitian ini tentang perbandingan kinerja sistem klasifikasi berbasis *K-fold Cross Validation* pada Algoritma *Decision Tree id3* dan *Decision Tree C5.0* dapat disimpulkan bahwa berdasarkan dataset kelayakan tenaga kerja dari www.kaggle.com, Algoritma *Decision Tree ID3* lebih unggul dengan nilai presisi sebesar 95%, recall/sensitivity sebesar 94%, dan akurasi sebesar 93%. Sedangkan Algoritma *Decision Tree C5.0* memperoleh nilai presisi sebesar 91%, recall/sensitivity sebesar 92%, serta nilai akurasi sebesar 89%.

Daftar Pustaka

- [1] A. R. Kristiyani, N., Fibriani, C., & Tanaamah, "Sistem Pendukung Keputusan dengan Menggunakan Algoritma Iterative Dichotomizer Three.," *J. Teknol. Informasi-IT*, pp. 1-100., 2011.
- [2] M. Arif, "Decision Tree Algorithms C4.5 and C5.0 in Data Mining: A Review," *Int. J. Database Theory Appl.*, vol. Vol. 11, N, pp. 4-5, 2018.
- [3] S. D. and D. L. C. Pardede, "PERBANDINGAN KINERJA ALGORITMA ID3 DAN C4.5 DALAM KLASIFIKASI SPAM-MAIL," *gunadarma Univ. Repos.*, 2010.
- [4] E. Galathiya, "Pengembangan Algoritma pohon keputusan C4.5 dan melakukan perbandingan

- terhadap algoritma Naive Bayes, Sequential Minimal Optimization (SMO),” 2012.
- [5] P. I. C. Wibowo, A. C. Fauzan, M. Dwi and and F. A. Q. Yustiana, “Komparasi Algoritma Naive Bayes dan *Decision Tree* Untuk Memprediksi Lama Studi Mahasiswa,” vol. 1, pp. 65–74, 2019.
- [6] C.-C. T. Hsi-Jen Chiang, Chun-Chieh Tseng, “A retrospective analysis of prognostic indicators in dental implant therapy using the C5.0 *Decision Tree* algorithm,” *J. Dent. Sci.*, vol. 8, pp. 248–255, 2013.
- [7] M. J. H. and G. Carleo, “Neural-Network Approach to Dissipative Quantum Many-Body Dynamics,” *Phys. Rev. Lett.*, vol. 122, 2019.
- [8] K. Crammer, “On the algorithmic implementation of multiclass kernel-based vector machines,” *J. Mach. Learn. Res. - JMLR*, vol. 2, pp. 265–292, 2002.
- [9] Z. W. Ibomoiey Domor Mienyea, Yanxia Suna, “Prediction performance of improved *Decision Tree*-based algorithms: a review,” *2nd Int. Conf. Sustain. Mater. Process. Manuf.*, pp. 698–703, 2019.
- [10] L. N. Wang Xiaohu, Wang Lele, “An Application of *Decision Tree* Based on ID3,” *2012 Int. Conf. Solid State Devices Mater. Sci.*, pp. 1017 – 1021, 2012.
- [11] N. Putri, Y. R., Mukhlash, I. dan Hidayat, “Prediksi Pola Kecelakaan Kerja pada Perusahaan Non Ekstraktif Menggunakan Algoritma *Decision Tree*: C4.5 dan C5.0,” *J. Sains dan Seni Pomits*, vol. 2, 2013.
- [12] & M. C. Nurhayati., Iwan K, Hadihardaja., Indratmo Soekarno., “A Study of Hold-Out and *K-fold Cross Validation* for Accuracy of Groundwater Modeling in Tidal Lowland Reclamation Using Extreme Learning Machine,” *2nd Int. Conf. Technol. Informatics, Manag. Eng. Environ.*, pp. 228 – 233, 2014.
- [13] I. M. and Fitriyani, “Implementasi Algoritma C4.5 Untuk Klasifikasi Anak Berkebutuhan Khusus Di Ibnu Sina Stimulasi Center,” *eProsiding Sist. Inf*, vol. 1, pp. 136–144, 2020.
- [14] X. A. Budiman, A. S., & Parandani, “Uji Akurasi Klasifikasi Dan Validasi Data Pada Penggunaan Metode Membership Function Dan Algoritma C4.5 Dalam Penilaian Penerima Beasiswa,” *J. Tek. Mesin, Elektro Dan Ilmu Komput.*, pp. 565–578, 2018.
- [15] H. D. Fahma and A. C. Fauzan, “Prediksi Keberlangsungan Studi Mahasiswa Fakultas Ilmu Pendidikan dan Sosial Universitas Nahdlatul Ulama Blitar,” vol. 1, pp. 110–119, 2021.
- [16] H. H. P. and P. Prajapati, “Study and Analysis of *Decision Tree* Based Classification Algorithms,” *J. Comput. Sci. Eng*, vol. 6, pp. 76–78, 2018.