



Komparasi Penerapan *Forward Selection* Pada Algoritma C4.5 Dan Algoritma *Naïve Bayes* Untuk Memprediksi Kompetensi Karyawan (Studi Kasus: PT. XYZ)

Fuza Putra Arwanda¹, Hannie², Budi Arif Dermawan³

^{1,2,3}Program Studi Teknik Informatika, Fakultas Ilmu Komputer Universitas Singaperbangsa Karawang. Jl. H.S Ronggowaluyo Teluk Jambe Timur Kabupaten Karawang. 41361.

Email : fuza.putra17102@student.unsika.ac.id

Info Artikel

Sejarah Artikel:

Diterima: 30 Mei 2021

Direvisi: 8 Juni 2021

Dipublikasikan: Juni 2021

e-ISSN: 2089-5364

p-ISSN: 2622-8327

DOI: 10.5281/zenodo.4970680

Abstract:

Monitoring and evaluation that is not well targeted towards employee assessment will cause problems in the development of human resources within a company. Currently the company PT. XYZ, is a new company which just started its production in 2020, thus monitoring and evaluation of the performance of each employee is very necessary. This study aims to compare C4.5 and the Naïve Bayes algorithm and add forward selection with the KDD (Knowledge Discovery in Database) methodology to predict competent and incompetent employees. Evaluation of the comparative use of the two algorithms, namely C4.5 and Naïve Bayes, was carried out by testing using Cross-Validation using 10-fold cross-validation.

Keywords: *C4.5, Naïve Bayes, Employee Competence, KDD, Cross-Validation*

PENDAHULUAN

Memasuki era revolusi industri 4.0 yang mana teknologi informasi menjadi basis utama dalam kehidupan manusia. Segala sesuatu hal dilakukan tanpa batas dengan penggunaan daya komputasi dan data yang tidak terbatas, dikarenakan pengaruh oleh internet juga teknologi digital yang masif menjadi tulang punggung pergerakan dan konektivitas manusia dan mesin. Pada era revolusi industri 4.0 akan terjadi masalah dalam berbagai hal aktivitas manusia, masalah sumber daya manusia menjadi salah satu hal penting dalam sebuah organisasi. Sumber daya manusia berperan menentukan arah

dalam kemajuan sebuah organisasi (Rohida, 2018).

Monitoring dan evaluasi yang tidak tepat sasaran terhadap penilaian karyawan akan menimbulkan masalah dalam pengembangan sumber daya manusia di dalam sebuah perusahaan. Penentuan karyawan pada sebuah perusahaan mempunyai penilaian yang berbeda-beda seperti menyeleksi berkas, maupun wawancara. Cara tersebut umum dilakukan dalam penentu karyawan disuatu perusahaan (Wahyono, 2018). *Data mining* adalah proses penambangan informasi terpendam dalam sebuah *database* yang

sebelumnya tidak diketahui (Werdiningsih, 2020). Metode klasifikasi merupakan metode yang mengekstraksi model dengan mengelompokan data latih pada sebuah kelas (Prasetyowati, 2017).

Metode ini telah dipakai pada beberapa penelitian sebelumnya, Penelitian (Sunge, 2018) memprediksi kompetensi karyawan menggunakan algoritma C4.5 pada PT. Hankook Tire Indonesia menghasilkan hasil klasifikasi dalam prediksi kompetensi menggunakan *data training* dengan *Confusion Matrix* nilai akurasi sebesar 78.64% dan data *testing* nilai akurasi sebesar 56.00% untuk nilai kurva ROC *data training* akurasi sebesar 0.907 sehingga masuk ke dalam *Excellent Classification* dan data *testing* dengan akurasi sebesar 0.783 sehingga masuk ke dalam *Fair Classification*. Algoritma C4.5 itu sendiri memiliki kelebihan dalam pengambilan keputusan yang kompleks dapat diubah menjadi sederhana dan juga dapat menghilangkan perhitungan yang tidak penting. Permasalahannya untuk kualitas keputusan yang didapatkan sangat tergantung dengan bagaimana pohon keputusan yang dihasilkan (Haryanto, 2017). Maka dari itu atas saran dari penelitian sebelumnya yang dilakukan menggunakan algoritma C4.5 dalam prediksi kompetensi karyawan PT. Hankook Tire Indonesia ini ditambahkan beberapa atribut agar memaksimalkan dalam prediksi kompetensi semakin akurat. Kemudian dapat dilakukan juga pengujian dengan penambahan metode optimasi agar dapat meningkatkan nilai akurasi pada algoritma juga bisa dilakukan komparasi menggunakan algoritma lain salah satunya *Naïve Bayes* untuk memperoleh perbandingan akurasi yang paling tinggi dalam membuat kualifikasi prediksi kompetensi karyawan (Sunge, 2018). Penelitian (Retnasari & Rahmawati, 2017) yang berjudul Diagnosa penyakit jantung dengan model algoritma *Naïve Bayes* dan algoritma C4.5 dihasilkan jika akurasi dari algoritma *Naïve Bayes* sebesar 86.67%

dengan nilai AUC 0.909 termasuk kedalam *Excellent Classification* sedangkan akurasi algoritma C4.5 sebesar 83.70% dengan nilai AUC 0.834 termasuk ke dalam *Good Classification*. Dengan perbandingan kedua Algoritma tersebut terlihat jika algoritma *Naïve Bayes* memiliki tingkat akurasi yang bagus dibandingkan algoritma C4.5. Adapun kekurangan pada algoritma *Naïve Bayes* ini pada saat melakukan proses *training* dan *testing* klasifikasi tidak optimal jika tidak disimpan pada dataset untuk itu perlu disimpan hasil klasifikasi pada dataset (Muslehatin, 2017). Untuk itu agar mengoptimalkan dari masing-masing algoritma ditambahkan teknik Optimasi *Forward Selection* pada algoritma klasifikasi *data mining* untuk mengurangi tingkat kompleksitas, meningkatkan akurasi pada algoritma klasifikasi, juga melihat atribut yang berpengaruh terhadap tingkat akurasi (Supriyanti, 2018).

Berdasarkan permasalahan yang ada, maka penelitian ini bertujuan untuk membandingkan tingkat akurasi algoritma C4.5 berbasis *Forward Selection* dengan algoritma *Naïve Bayes* berbasis *Forward Selection* dalam memprediksi kompetensi karyawan pada PT. XYZ.

METODOLOGI PENELITIAN

Penelitian dilakukan dengan menggunakan data Kompetensi karyawan pada PT. XYZ yang dimiliki dari setiap karyawan per tanggal 9 November 2020 - 8 Maret 2021 yang berjumlah 54 dengan 34 atribut yaitu Nik, Usia, Jenjang Karir, Tingkat Kompetensi terdiri dari 30 kode (A1, A2, A3, A4, A5, A6, A7, A8, A9, A10, B1, B2, B3, B4, B5, B6, B7, B8, C1, C2, C3, C4, C5, C6, C7, C8, C9, C10, C11, C12), Posisi, dan Hasil. Metodologi yang dipakai pada penelitian ini adalah *Knowledge Discovery in Database (KDD)* yang mana Merupakan sebuah ekstraksi pada informasi yang potensial implisit serta tak diketahui pada kumpulan data. Tahapan KDD juga mengaitkan dengan hasil proses *data mining* (proses ekstrak kecenderungan

suatu pola data), setelah itu merubah hasil dengan akurat sebagai suatu informasi yang nantinya dapat dimengerti [9]. KDD terdiri dari lima tahap yaitu :

1. Seleksi Data

Tidak semua atribut pada data yang telah dikumpulkan digunakan untuk proses pengolahan *data mining*. Operator *Select Attributes* pada *RapidMiner* digunakan untuk menyeleksi atribut mana saja yang akan digunakan. Atribut yang akan dipakai yaitu Usia, Jenjang Karir, Tingkat Kompetensi dan Hasil. Atribut “Nik” dan “Posisi” tidak dipakai dikarenakan hanya digunakan sebagai identitas pada data saja, sedangkan atribut lainnya berpengaruh yang kemudian akan diolah menggunakan 2 metode klasifikasi *data mining* yaitu dengan Algoritma *c4.5* dan *Naïve Bayes*

2. *Preprocessing*

Preprocessing merupakan salah satu tahapan penting pada proses mining dikarenakan data yang digunakan dalam proses mining tidak selamanya dalam kondisi yang ideal untuk diproses terkadang pada data tersebut terdapat berbagai permasalahan yang dapat mengganggu hasil dari proses mining itu sendiri diantaranya seperti *missing value*, data *redundant*, *outliers*. Setelah melihat dan menganalisis data yang ada pada tahapan *data preprocessing* dengan data yang ada yaitu data kompetensi karyawan yang telah diseleksi dan ternyata tidak ada *missing value* ataupun data yang tidak sempurna sehingga pada tahap ini tidak dilakukan penanganan data karena data sudah bagus.

3. *Transformation*

Pada data kompetensi karyawan data yang akan ditransformasi yaitu pada atribut “Usia” dikarenakan

tipe data pada atribut ini numerik jadi agar dapat mempermudah dalam proses pengklasifikasian maka dilakukan transformasi pada atribut “Usia” dengan mengubahnya menjadi beberapa interval kategori dengan menggunakan operator *Discretization by Binning* pada *software RapidMiner*.

Discretization by Binning mengubah atribut numerik menjadi ke dalam beberapa kelompok atau *bin* yang telah ditentukan, dimana nilai terbagi beberapa rentang yang sama pada setiap *bin* (Hofman & Klinkenberg, 2016). Proses *Discretization by Binning* akan dilakukan pada atribut “Usia” dengan tipe numerik yang memiliki nilai 18 sampai 62 diubah menjadi 5 interval yang terdiri dengan rentang nilai setiap interval

Tabel 1. Hasil Transformasi dengan *Discretization by Binning*

Interval	Keterangan	Inisialisasi
Range 1	$-\infty - 18,3$	18 Tahun kebawah
Range 2	18,0 – 29,0	Usia 19 - 29 Tahun
Range 3	29,0 – 40,0	Usia 30 - 40 Tahun
Range 4	40,0 – 51,10	Usia 41 - 51 Tahun
Range 5	50,0 – 61,0	Usia 52 - 62 Tahun

Pada data kompetensi karyawan juga akan mentransformasikan atribut “Jenjang Karir” dikarenakan pada atribut tersebut memiliki tipe data kategorikal sehingga untuk melakukan proses prediksi dilakukan transformasi untuk mengubah tipe data juga nilai *record* menjadi numerik untuk mempermudah proses klasifikasi. Proses transformasi ini dilakukan dengan menggunakan *tools* *RapidMiner* dengan bantuan operator *Map*. Operator *Map*

digunakan untuk memetakan nilai tertentu dari atribut yang dipilih dengan nilai baru.

Tabel 2. Proses Transformasi

No	Jenjang Karir	Perubahan
1	SD	1
2	SMP	2
3	SMA/SMK/MA	3
4	D2/D3	4
5	S1	5

4. Data Mining

Pada tahap ini dilakukan penerapan algoritma C4.5 dan juga *Naïve Bayes* untuk memprediksi kompetensi karyawan pada PT. Xyz. Algoritma C4.5 bekerja dengan menghitung nilai *entropy* dan gain pada setiap atribut data kemudian membentuk suatu pohon keputusan (Mardi, 2017). Sedangkan *Naïve Bayes* bekerja dengan melakukan pengklasifikasian berdasarkan probabilitas nilai pada setiap atributnya (Pratiwi, 2020).

C4.5 memiliki persamaan rumus sebagai berikut :

$$Entropy(S) = \sum_i^n = 0 - pi * \log_2 pi$$

$$Gain(S, A) = Entropy(S) - \sum_i^n = 1 \frac{[Si]}{[S]} * Entropy(S)$$

Naïve Bayes memiliki persamaan rumus sebagai berikut :

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}$$

Pengujian algoritma C4.5 dan *Naïve Bayes* pada penelitian ini dilakukan dengan menggunakan metode *k-fold cross-validation* dengan menggunakan *10-fold*. *K-fold cross-validation* merupakan metode validasi yang membagi secara random seluruh data menjadi *data*

training dan *data testing*, dengan keuntungan memanfaatkan data sebanyak mungkin untuk pelatihan dengan beberapa kali iterasi pengujian sesuai dengan nilai k yang ditentukan (Anindika, 2020)

5. Evaluasi

Dari pengujian yang telah dilakukan selanjutnya dievaluasi untuk menilai performa dari penggunaan kedua algoritma sebelum dan sesudah menggunakan *forward selection* pada dataset kompetensi karyawan PT. XYZ berdasarkan nilai *accuracy* dari *confussion matrix* dan nilai AUC. AUC dihitung untuk mengukur perbandingan performa (Widaningsih 2019).

0.90 – 1.00 : *excellent classification*

0.80 – 0.90 : *good classification*

0.70 – 0.80 : *Mfair classification*

0.60 – 0.70 : *poor classification*

0.50 – 0.60 : *failure*

HASIL PENELITIAN DAN PEMBAHASAN

- Penelitian ini dilakukan untuk
- (1) memprediksi karyawan yang kompeten dan tidak kompeten dengan menggunakan metodologi *Knowledge Discovery in Database* (KDD). Data yang digunakan adalah data kompetensi karyawan pada PT. XYZ. diambil dari data karyawan juga kompetensi yang dimiliki dari setiap karyawan per tanggal 9 November 2020 - 8 Maret 2021 yang berjumlah 54 dengan 34 atribut yaitu Nik, Usia, Jenjang Karir,
 - (2) Tingkat Kompetensi terdiri dari 30 kode (A1, A2, A3, A4, A5, A6, A7, A8, A9, A10, B1, B2, B3, B4, B5, B6, B7, B8, C1, C2, C3, C4, C5, C6, C7, C8, C9, C10, C11, C12), Posisi, dan Hasil. Untuk atribut kode tingkat kompetensi sendiri penulis membagi ke dalam 3 kelompok kompetensi terdiri dari Kompetensi Umum, Kompetensi Khusus dan Kompetensi Teknikal.

Tabel 3. Kategori Kompetensi Karyawan

Kode	Kategori Kompetensi Umum	Kode	Kategori Kompetensi Khusus	Kode	Kategori Kompetensi Teknikal
A1	Pemahaman ISO 9001	B1	Kedisiplinan absensi	C1	Membuat report logbook dan kartu stock barang
A2	Pemahaman ISO 14400 1	B2	Bekerja secara efektif dan efisien	C2	Membuat laporan
A3	Leadership	B3	Mampu berkomunikasi dengan baik	C3	Melakukan reparasi dan perawatan mesin
A4	Memahami profil perusahaan	B4	Melakukan kontrol	C4	Melakukan instalasi pada mesin baru
A5	Memahami peraturan perusahaan	B5	Memahami prosedur tim	C5	Melakukan kontrol dalam proses pengemasan barang
A6	Teliti terhadap tugas kerja	B6	Mengantahui jenis mesin Ties On dan Earloop	C6	Pengecekan hasil kemas target Ties On dan Earloop
A7	Proaktif	B7	Mengetahui jenis Material Masker	C7	Melakukan packing barang setiap hasil produksi
A8	Pemahaman proses	B8	Mampu mengatasi permasalahan	C8	Mengoperasikan mesin produksi Ties On dan Earloop

A9	Memahami 5S	C9	Membuat rencana jadwal produksi
A10	Pemahaman PL (Product Liability)	C10	Mengelola anggaran produksi
		C11	Mengelola arsip data
		C12	Membuat agenda kantor

Tabel 4. Data Training untuk prediksi

No	USIA	JENJANG KARIR	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	B1	B2	B3	B4	B5	B6	B7	B8	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	HASIL		
1	50	SL	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	4	5	3	3	3	3	3	4	4	4	5	KOMPETEN			
2	51	SL	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	4	4	3	3	4	3	3	5	5	5	4	KOMPETEN			
3	41	SL	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	4	4	3	3	4	4	4	4	4	4	4	KOMPETEN			
4	41	SL	5	5	4	5	5	5	5	5	5	5	5	5	4	5	5	5	5	5	3	3	5	5	3	3	4	5	3	3	3	KOMPETEN			
5	33	D3	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	4	4	4	4	4	4	4	4	5	5	4	KOMPETEN			
6	33	D3	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	4	4	4	4	4	4	4	4	5	5	4	KOMPETEN			
7	26	D3	4	4	3	5	5	4	5	5	4	5	5	5	4	5	4	5	5	5	4	1	0	1	3	0	0	3	3	5	5	KOMPETEN			
8	21	D2	4	4	3	5	5	4	5	5	4	4	5	5	4	5	4	5	5	5	3	4	1	0	1	3	0	0	3	3	5	5	KOMPETEN		
9	21	D2	4	4	3	5	5	4	5	5	4	4	5	5	4	5	4	5	5	5	3	4	1	0	1	3	0	0	3	3	5	5	KOMPETEN		
10	14	SMK	4	4	3	5	5	4	5	5	4	5	5	5	4	5	4	5	5	5	3	4	1	0	1	3	0	0	3	3	5	5	KOMPETEN		
11	19	SMK	4	4	3	5	5	4	5	5	4	4	5	5	4	5	4	5	5	5	3	4	1	0	1	3	0	0	3	3	5	5	KOMPETEN		
12	22	SMA	4	4	3	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	2	4	4	4	4	5	5	0	0	3	3	3	KOMPETEN		
13	27	SMA	4	4	3	5	5	4	5	5	5	5	5	5	5	5	5	5	5	5	2	4	4	4	4	5	5	0	0	3	3	3	KOMPETEN		
14	21	SMA	4	4	3	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	2	4	4	4	4	5	5	3	3	3	3	3	KOMPETEN		
15	13	SMA	4	4	3	5	5	5	5	4	5	5	5	5	5	5	5	5	5	5	2	5	4	3	5	5	4	3	3	3	3	3	KOMPETEN		
16	10	SMK	4	4	3	5	5	5	5	4	5	5	5	5	5	5	5	5	5	5	4	4	4	4	4	4	4	2	3	3	3	3	KOMPETEN		
17	19	SMA	3	3	3	5	5	4	4	5	4	4	5	4	3	5	4	5	5	5	1	1	3	2	0	0	2	2	4	2	2	2	KOMPETEN		
18	20	SMA	3	3	3	5	5	4	4	5	4	4	5	3	3	4	5	5	5	5	1	1	3	2	0	0	2	2	4	2	2	2	KOMPETEN		
19	20	SMK	3	3	3	5	5	5	4	4	4	4	4	4	3	5	5	5	5	0	0	2	1	2	2	0	5	0	0	1	0	0	KOMPETEN		
20	20	SMK	3	3	3	5	5	4	4	4	4	4	4	4	3	5	5	5	5	4	0	0	2	1	2	2	0	5	0	0	1	0	0	KOMPETEN	
21	19	SMK	3	3	3	5	5	4	4	4	4	4	4	4	3	5	5	5	5	5	0	0	2	1	2	2	0	5	0	0	1	0	0	KOMPETEN	
22	21	SMA	3	3	3	5	5	4	4	4	4	4	4	4	3	5	5	5	5	0	0	2	1	2	2	0	5	0	0	1	0	0	KOMPETEN		
23	20	SMK	3	3	3	5	5	4	4	4	4	4	4	4	3	5	5	5	5	5	0	0	2	1	2	2	0	5	0	0	1	0	0	KOMPETEN	
24	20	SMA	3	3	3	5	5	4	4	4	4	4	4	4	3	5	5	5	5	5	0	0	2	1	2	2	0	5	0	0	1	0	0	KOMPETEN	
25	19	SMK	3	3	3	5	5	4	4	4	4	4	4	4	3	5	5	5	5	5	0	0	2	1	2	2	0	5	0	0	1	0	0	KOMPETEN	
26	21	SMK	3	3	3	5	5	4	4	4	4	4	4	4	3	5	5	5	5	5	4	0	0	2	1	2	2	0	5	0	0	1	0	0	KOMPETEN
27	22	SMA	3	3	3	5	5	4	4	4	4	4	4	4	3	5	5	5	5	5	0	0	2	1	2	2	0	5	0	0	1	0	0	KOMPETEN	
28	19	SMA	3	3	3	5	5	4	4	4	4	4	4	4	3	5	5	5	5	5	0	0	2	1	2	2	0	5	0	0	1	0	0	KOMPETEN	

Setelah melewati tahap *data selection*, *data preprocessing* dan *data transformation* diperoleh atribut yang akan diproses yaitu “Usia”, “Jenjang Karir”, “Tingkat Kompetensi”, dan “Hasil”.

- Perhitungan Algoritma C4.5 Mencari nilai *entropy* dan Gain yang paling besar dari setiap atribut.

Tabel 5. Perhitungan Algoritma C4.5

Atribut	Nilai	Jumlah	Kompeten	Tidak	Entropy	Gain Ratio
		54	40	14	0,82	
					5627	
A8						0,43
						6948
	5	18	18	0	0	
	4	28	22	6	0,74	
					9595	
	3	0	0	0	0	
	2	0	0	0	0	
	1	8	0	8	0	

Atribut Kode A8 merupakan menjadi atribut yang memiliki nilai gain besar yang mana menjadi akar pada pohon keputusan juga menjadi salah satu atribut yang berpengaruh.

- Perhitungan Algoritma Naïve Bayes Menghitung probabilitas terdapat 2 class dengan class Kompeten berjumlah 40 dan Tidak Kompeten berjumlah 14.

Tabel 6. Probabilitas Setiap Kelas

Kelas	Jumlah	Probabilitas Kelas P(H)
Kompeten	40	0,740
Tidak Kompeten	14	0,259

- Confusion Matrix

Tabel 7. Model C4.5 Confusion Matrix 10-fold Cross-Validation

Accuracy: 84.67% +/- 18.27% (micro average: 85.18%)

	Real KOMPETEN	Real TIDAK	Class predicted
pred KOMPETEN	35	5	82.11%
pred TIDAK	5	11	68.75%
Class total	87.50%	78.57%	

Dari pengujian Model C4.5 dalam memprediksi kompetensi karyawan diperoleh nilai AUC sebesar 0,500 yang termasuk kategori *Failure*.

Tabel 8. Model Naïve Bayes Confusion Matrix 10-fold Cross-Validation

Accuracy: 77.57% +/- 11.52% (micro average: 77.78%)

	Real KOMPETEN	Real TIDAK	Class predicted
pred KOMPETEN	25	4	86.96%
pred TIDAK	4	10	68.89%
Class total	80.00%	71.43%	

Dari pengujian Model *Naïve Bayes* dalam memprediksi kompetensi karyawan diperoleh nilai AUC sebesar 0,940 yang termasuk kategori *Excellent Classification*.

Tabel 9. Model C4.5 berbasis forward selection Confusion Matrix 10-fold Cross-Validation

Accuracy: 93.87% +/- 11.81% (micro average: 93.87%)

	Real KOMPETEN	Real TIDAK	Class predicted
pred KOMPETEN	40	0	100.00%
pred TIDAK	0	6	100.00%
Class total	100.00%	67.14%	

Dari pengujian Model C4.5 berbasis fitur *forward Selection* dalam memprediksi kompetensi karyawan diperoleh nilai AUC sebesar 0,500 yang termasuk kategori *Failure*.

Tabel 10. Model Naïve Bayes berbasis forward selection Confusion Matrix 10-fold Cross-Validation

Accuracy: 95.88% +/- 11.25% (micro average: 94.48%)

	Real KOMPETEN	Real TIDAK	Class predicted
pred KOMPETEN	40	0	83.02%
pred TIDAK	0	11	100.00%
Class total	100.00%	78.57%	

Dari pengujian Model *Naïve Bayes* berbasis fitur *forward Selection* dalam memprediksi kompetensi karyawan diperoleh nilai AUC sebesar 0,944 yang termasuk kategori *Excellent Classification*

KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, maka dapat disimpulkan beberapa hal yaitu sebagai berikut :

- Penerapan pada algoritma C4.5 dan *Naïve Bayes* dengan metodologi KDD digunakan untuk memprediksi Kompetensi Karyawan pada PT. XYZ. Algoritma C4.5 melakukan pembelajaran menghitung nilai *entropy* dan nilai gain pada setiap atribut data juga menghasilkan pohon keputusan yang memiliki 1 akar pohon

yaitu A8 yang mana itu menjadi atribut yang memiliki nilai Gain besar yaitu 0,436948 sekaligus menjadi atribut yang berpengaruh. Sedangkan Algoritma *Naïve Bayes* melakukan pembelajaran menggunakan nilai probabilitas setiap atributnya.

2. Evaluasi penggunaan komparasi kedua algoritma yaitu C4.5 dan *Naïve Bayes* dilakukan dengan pengujian menggunakan *Cross-Validation* penggunaannya *10-fold cross-validation* menghasilkan pada Algoritma *Naïve Bayes* didapat nilai *accuracy* sebesar 77,78% dan nilai *AUC* sebesar 0,940 sehingga masuk kedalam kategori *Excellent Classification*. Menggunakan algoritma *Naïve Bayes* berbasis *forward selection* didapat nilai *accuracy* sebesar 94,44% dan nilai *AUC* sebesar 0,944 sehingga masuk kedalam kategori *Excellent Classification*. Setelah digunakan optimasi *forward selection* nilai *accuracy* meningkat 16,66% nilai *AUC* juga meningkat 4. Maka, algoritma *Naïve Bayes* dinilai mampu memprediksi kompetensi karyawan pada PT. XYZ dengan baik dibandingkan algoritma C4.5.

DAFTAR PUSTAKA

- Rohida, Leni. (2018). Pengaruh Revolusi Industri 4.0 Terhadap Kompetensi Sumber Daya Manusia. *Jurnal Manajemen dan Bisnis Indonesia*, Vol. 6(1), pp. 114-36.
- Wahyono., & Agung, Nugroho. (2018). Penerapan Algoritma C4.5 Untuk Prediksi Tingkat Kompetensi Karyawan Pt. Multistrada Arah Sarana. *Joutica*, Vol. 3(1), 145.
- Haryanto., Fandy, Ferdian., & Seng, Hansun. (2017). Penerapan Algoritma C4.5 Untuk Memprediksi Penerimaan Calon Pegawai Baru Di PT. Wise. *Teknik Informatika Dan Sistem Informasi*, Vol. 3(2), pp. 95-103.
- Sunge, Aswan, Supriadi. (2018). Prediksi Kompetensi Karyawan Menggunakan Algoritma C4.5 (Studi Kasus PT. Hankook Tire Indonesia). *Seminar Nasional Teknologi Informasi dan Komunikasi 2018*, pp. 23-24.
- Muslehatin, Wiwik., Muhammad, Ibnu., & Mustakim. (2017). Penerapan *Naïve Bayes Classification* Untuk Klasifikasi Tingkat Kemungkinan Obesitas Mahasiswa Sistem Informasi UIN Suska Riau. *Seminar Nasional Teknologi Informasi Komunikasi dan Industri (SNTIKI)*, Vol 7.
- Supriyanti., Wiwit., & Norma Puspitasari. (2018). Implementasi Teknik Seleksi Fitur Forward Selection Pada Algoritma Klasifikasi Data Mining Untuk Prediksi Masa Studi Mahasiswa Politeknik Indonusa Surakarta. *Jurnal INFORMA Politeknik Indonesia Surakarta*, Vol. 4(2), pp. 49-54.
- I. Werdiningsih., B. Nuqoba., & Muhammadun. (2020). Data Mining Menggunakan Android Weka, dan SPSS: Airlangga University Press, Surabaya.
 - E. Prasetyowati. (2017). Data <i>iing Pengelompokan Data Untuk Informasi dan Evaluasi. *Duta Media Publishing*, Pamekaran.
 - M. Afdal., & Muhammad, Rosadi. (2019). Penerapan Association Rule Data Mining Untuk Analisis Penempatan Tata Letak Buku Di Perpustakaan Menggunakan Algoritma Apriori. *Jurnal Ilmiah Rekayasa dan Manajemen Sistem*, Vol. 5(1), pp. 99.
 - Mardi, Yuli. (2017). Data Mining: Klasifikasi Menggunakan Algoritma C4.5. *Jurnal Edik Informatika*, Vol. 2(2), pp. 213-219.

- D. A. Pratiwi., R. M. Awangga., & M. Y. H. Setyawan. (2020). Seleksi Calon Kelulusan Tepat Waktu Mahasiswa Teknik Informatika Menggunakan Metode Naïve Bayes. *Kreatif*, Bandung.
- E. Anindika, Sari., M. Thereza. Br. Saragih., I. Ali Shariati., S. Sofyan., R. Al Baihaqi., & R. Nooraeni. (2020). Klasifikasi Kabupaten Tertinggal di Kawasan Timur Indonesia Dengan Support Vector Machine. *JIKO(Jurnal Inform dan Komputer)*, Vol. 3(3), pp. 188-195. Doi:10.33387/v3i3.2364.
- Widianingsih, Sri. (2019). Perbandingan Metode Data Mining Untuk Prediksi Nilai dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4.5, Naïve Bayes, Knn dan Svm. *Jurnal Tekno Insentif*, Vol. 13(1), pp. 16-25