

## **Perbandingan Teknik Klasifikasi Untuk Memprediksi Kualitas Kinerja Karyawan**

**Hijrah\*<sup>1</sup>, Mukhlizar\*<sup>2</sup>, T.M. Azis Pandria\*<sup>3</sup>**

<sup>1</sup>Jurusan Dakwah dan Komunikasi Penyiaran Islam, STAIN Teungku Dirundeng, Meulaboh

<sup>2</sup>Jurusan Tarbiyah, STAIN Teungku Dirundeng, Meulaboh

<sup>3</sup>Jurusan Teknik Industri, Fakultas Teknik, Universitas Teuku Umar

e-mail: \*<sup>1</sup>hijrah@staindirundeng.ac.id, \*<sup>2</sup>mukhlizar@staindirundeng.ac.id,

\*<sup>3</sup>azispandria@utu.ac.id

### **Abstrak**

Penelitian ini dilakukan untuk membandingkan beberapa teknik klasifikasi yang digunakan untuk menentukan kualitas kinerja karyawan. kualitas kinerja karyawan sangat berpengaruh terhadap perkembangan sebuah perusahaan, ada banyak atribut yang digunakan sebagai pengukur kinerja karyawan semakin banyak atribut yang digunakan maka akan semakin baik hasil dari kualitas penilaian yang dilakukan. untuk mendapatkan hasil yang lebih maksimal maka dilakukan perbandingan teknik klasifikasi, perbandingan dilakukan dengan melihat hasil kinerja dari metode yang digunakan. metode yang digunakan algoritma C4.5, Naive Bayes, dan k-nearest neighbors, metode yang akan digunakan tersebut akan dilakukan pengujian menggunakan alat rapid miner

**Kata kunci** – Kinerja Karyawan, Atribut, Metode Klasifikasi, Rapid Miner

### **Abstract**

*This study was conducted to compare several classification techniques used to determine the quality of employee performance. the quality of employee performance is very influential on the development of a company, there are many attributes that are used as a measure of employee performance the more attributes used then the better the results of the quality of assessment performed. to get maximum result hence done by comparison of technique of classification, comparison done by seeing result of performance of method used. method used by algorithm C4.5, Naive Bayes, and k-nearest neighbors, method to be used will be tested using tool rapid miner*

**Keyword** – Employee Performance, Attributes, Classification Method, Rapid Miner

## **1. PENDAHULUAN**

Kinerja perusahaan sangat bergantung pada kinerja setiap individunya, setiap individu karyawan maupun kelompok akan memberikan kekuatan pada kinerja perusahaan sebab motivasi yang diperoleh akan berpengaruh terhadap kinerja perusahaan [1]. Pada penelitian ini dilakukan pemanfaatan serta penerapan metode data mining, Metode *Data mining* dibutuhkan karena adanya sejumlah data yang besar, data yang besar tersebut digunakan agar diperoleh informasi dan pengetahuan yang berguna sehingga menghasilkan penilaian kinerja karyawan yang bersifat obyektif [2]. Sebelum dilakukan penerapan dengan metode *data mining* pada penelitian ini akan dilakukan pengujian terhadap beberapa metode tersebut, pengujian dilakukan sebagai pembuktian bahwa metode yang diterapkan adalah metode yang lebih baik atau lebih tepat pada penelitian ini metode *data mining* yang akan dilakukan pengujian adalah algoritma C4.5, Naive Bayes, dan k-nearest neighbors.

Tinjauan studi yang dilakukan pada penelitian ini menyimpulkan bahwa algoritma C4.5 yang diterapkan dalam proses penilaian kinerja karyawan selama ini nilai akurasi yang diperoleh

mencapai 90 % dan termasuk sebagai *excellent classification* akan tetapi dari penelitian tersebut atribut yang digunakan untuk dilakukan pengujian hanya terdiri dari beberapa atribut saja, Dengan menggunakan atribut yang lebih banyak maka hasil yang diperoleh bisa lebih baik oleh sebab itu penelitian ini dilakukan dengan atribut yang lebih banyak sehingga memungkinkan nilai akurasi tersebut berubah jika digunakan pada kriteria atau atribut yang lebih banyak. proses pembuktian akurasi tersebut dilakukan dengan mencari nilai gain information dari setiap metode data mining.

## 2. METODE PENELITIAN

Penilaian kinerja karyawan dilakukan untuk memberikan informasi yang berkaitan dengan pekerjaan karyawan [3]. penilaian kinerja karyawan menggunakan bobot skala sebagai berikut:

**Tabel 1.** Skala Penilaian Kinerja Karyawan [4].

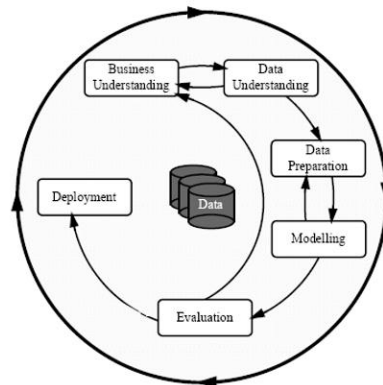
Skala Penilaian	Keterangan
1	Tidak memuaskan/Sangat Jelek
2	Perlu perbaikan/Jelek
3	Memenuhi harapan/Cukup
4	Melebihi harapan/Baik
5	Luar biasa/Amat baik

Penilaian bobot akan dikalikan dengan setiap kriteria, setiap kriteria akan dibagi dengan total nilai yang diperoleh sehingga keseluruhan penilaian dapat dinilai dari skala penilaian berikut:

**Tabel 2.** Skala Nilai Total Dalam Interval [4]

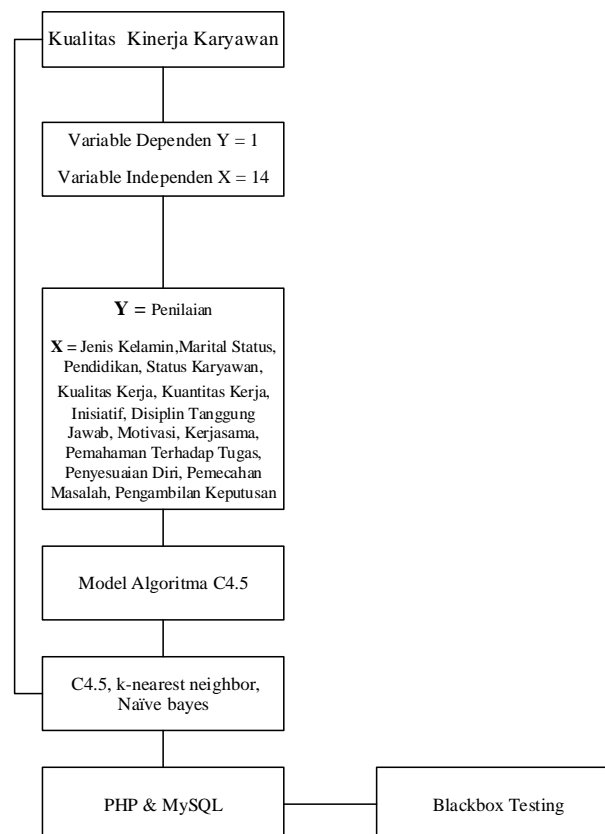
No	Skala Nilai	Kategori	Interval Nilai
1	A	Kinerja Sangat Tinggi	$4,20 < n \leq 5,00$
2	B	Kinerja Tinggi	$3,40 < n \leq 4,20$
3	C	Kinerja Sesuai Standar	$2,60 < n \leq 3,40$
4	D	Kinerja Rendah	$1,80 < n \leq 2,60$
5	E	Kinerja Tidak Efektif	$1 < n \leq 1,80$

Semakin berkembangnya penerapan *data mining* maka kelompok analis *data mining* yakni DaimlerChrysler, SPSS dan NCR bergabung untuk membangun *Cross-Industry Standard Process For Data mining* (CRISP-DM), kelompok tersebut membuat sebuah model proses *data mining* yang netral terhadap jenis industri, *tools*, dan aplikasi [5].



**Gambar 1.** Model Proses CRISP-DM [5].

Metodologi CRISP-DM merupakan standarisasi *data mining* yang digagas oleh Daimler Chrysler (Daimler-Benz), SPSS (ISL) yang selanjutnya dikembangkan pada berbagai workshop (antara 1997-1999), NCR [6]. pada penelitian ini juga diterapkan metodologi tersebut dengan tahap - tahap yang dilakukan sebagai berikut :



**Gambar 2.** Metodologi Penerapan CRISP-DM

Pada penelitian ini digunakan metode penelitian kuantitatif, penelitian tersebut merupakan penelitian yang menggunakan desain eksplanasi, dimana objek penelitiannya digunakan untuk menguji antar variable yang dihipotesiskan, adapun pada penelitian tersebut

hipotesis akan diuji kebenarannya [7]. Dalam proses penerapan mining C4.5, maka proses yang dilakukan adalah sebagai berikut :

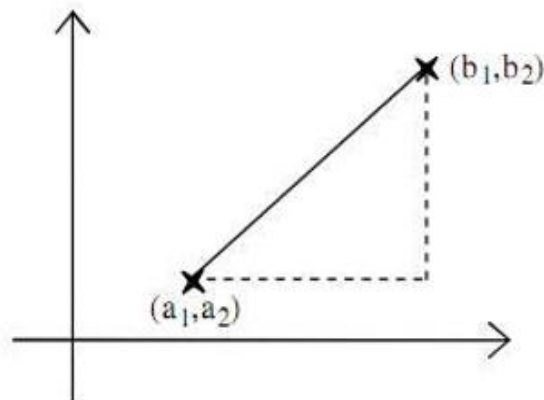
1. Hitung frekuensi kemunculan masing-masing nilai atribut pada data survey
2. Hitung nilai Entropy dari masing-masing nilai atribut
3. Hitung nilai Information Gain dengan menggunakan nilai Entropy yang telah dihitung sebelumnya.
4. Hitung nilai Split Info dari tiap atribut
5. Hitung nilai Gain Ratio menggunakan nilai Information Gain dan Split Info.
6. Ambil nilai Gain Ratio terbesar dan jadikan simpul akar.
7. Hilangkan atribut yang dipilih sebelumnya dan ulangi perhitungan nilai Entropy, Information Gain, Split Info dan Gain Ratio dengan memilih Gain Ratio terbesar dan dijadikan simpul internal pohon.
8. Ulangi perhitungan tersebut hingga semua atribut pohon memiliki kelas.
9. Jika semua pohon sudah memiliki kelas, maka tampilkan pohon keputusan awal dan *generate* aturan keputusan awal.

Naive bayes adalah klasifikasi probabilitas yang sesuai dengan teorema bayes dimana metode pada naive bayes menganggap bahwasanya efek dari nilai atribut pada kelas tertentu independen dari nilai - nilai atribut lainnya [8]. Naive bayes memiliki kemampuan klasifikasi yang sama dengan *decison tree* dan *neural network* dalam tingkat akurasi dan kecepatan yang tinggi pada saat diaplikasikan ke dalam database. Bayesian Classification merupakan statistik yang digunakan untuk memprediksi probabilitas keanggotaan suatu class dimana bentuk umum yang digunakan adalah sebagai berikut [2]:

$$P(H | X) = \frac{P(X | H)P(H)}{P(X)} \dots\dots\dots(1)$$

- X = Data dengan class yang belum diketahui
- H = Hipotesis data X merupakan suatu class spesifik
- P(H|X) = Probabilitas hipotesis H berdasarkan kondisi X
- P(H) = Probabilitas hipotesis H
- P(X|H) = Probabilitas X berdasarkan kondisi tersebut
- P(X) = Probabilitas dari X

Dalam pengklasifikasiannya pengujian pada KNN dilakukan dengan memberikan nilai K dimana nilai yang terbaik diperoleh dengan menggunakan pada validasi silang [9]. k-nearest neighbors dapat dimodifikasi untuk melakukan atribut yang bersifat kategorikal. *Atribut continue* pada jarak terdekat akan digunakan rumus jarak euclidean berikut adalah contoh jarak Euclidean [2]:



**Gambar 3.** Contoh jarak Euclidean [2].

Pada jarak euclidean antara titik  $(a_1, a_2, \dots, a_n)$  dan  $(b_1, b_2, \dots, b_n)$  pada ruang dimensi merupakan generalisasi dari dua hasil jarak Euclidean, sehingga dapat diberikan rumus sebagai berikut :  $\sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \dots + (a_n - b_n)^2}$

Confusion matrix merupakan sebuah metode evaluasi yang menggunakan tabel sebagai berikut :

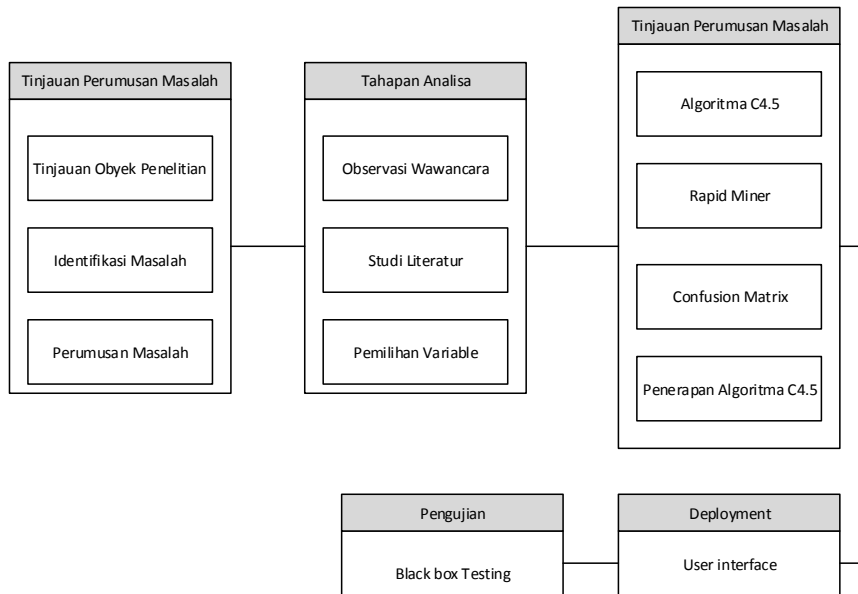
**Tabel 3.** Model Confusion Matrix [10].

<i>Correct Classification</i>	<i>Classified as</i>	
	+	-
+	<i>True positives</i>	<i>False negatives</i>
-	<i>False positives</i>	<i>True negatives</i>

Pada tabel tersebut dijelaskan bahwa dataset memiliki dua kelas yakni kelas positif dan kelas negatif, dengan penggunaan tabel tersebut nantinya akan diperoleh nilai *accuracy*, *precision*, dan *recall*. Nilai *Accuracy* adalah nilai yang mempresentasikan jumlah record data yang akan dilakukan pengklasifikasian secara benar oleh algoritma, nilai *precision* adalah nilai yang diperoleh dari jumlah kasus yang diprediksi dengan positif, dan nilai *recall* adalah proporsi dari sejumlah kasus positif yang diprediksi secara benar.

Pengujian pada penelitian ini menggunakan dua *tools data mining* adapun salah satu *tools* tersebut adalah rapid miner. Rapid miner merupakan perangkat lunak yang digunakan pada ilmu pengetahuan dan dikembangkan oleh perusahaan yang bergerak dalam pembelajaran mesin, penambangan teks, serta proses analisis dan prediktif [11]. Rapid miner merupakan platform analisis prediktif yang open source, pada rapid miner disediakan kemudahan dalam mengimport data dan mengevaluasi model dalam peramalan [12]. Rapid miner dapat membaca inputan sampai tiga angka di belakang koma sehingga menjelaskan bahwa rapid miner lebih teliti dimana hasil perhitungan pada rapid miner menghasilkan dua aturan asosiasi [13].

Berikut adalah bagan atau alur proses yang dilakukan untuk mencapai tujuan yang diharapkan dalam penelitian ini :



**Gambar 4.** Langkah-langkah Penelitian.

### 3. HASIL DAN PEMBAHASAN

Sebelum digunakan setiap data akan memiliki kelas binominal atau polynominal sesuai aturan yang telah dibuat serta berdasarkan nilai datanya.

**Tabel 4.** Pembagian Variable dan Kelas Data.

Variable	Nama Field	Jenis Kelas Data	Kelas Data Yang Digunakan
Y	Penilaian	Binominal	Efektif, Tidak Efektif
X1	Jenis Kelamin	Binominal	Male, Female
X2	Marital Status	Binominal	Menikah, Belum Menikah
X3	Pendidikan	Polynomial	SMA, D3, S1, S2
X4	Status Karyawan	Polynomial	Tetap, Kontrak, Outsourcing
X5	Kualitas Kerja	Polynomial	Sangat Jelek, Jelek, Cukup, Baik, Sangat Baik
X6	Kuantitas Kerja	Polynomial	Sangat Jelek, Jelek, Cukup, Baik, Sangat Baik
X7	Inisiatif	Polynomial	Sangat Jelek, Jelek, Cukup, Baik, Sangat Baik
X8	Disiplin	Polynomial	Sangat Jelek, Jelek, Cukup, Baik, Sangat Baik
X9	Tanggung Jawab	Polynomial	Sangat Jelek, Jelek, Cukup, Baik, Sangat Baik
X10	Motivasi	Polynomial	Sangat Jelek, Jelek, Cukup, Baik, Sangat Baik
X11	Kerjasama	Polynomial	Sangat Jelek, Jelek, Cukup, Baik, Sangat Baik
X12	Pemahaman Terhadap Tugas	Polynomial	Sangat Jelek, Jelek, Cukup, Baik, Sangat Baik
X13	Penyesuaian Diri	Polynomial	Sangat Jelek, Jelek, Cukup, Baik, Sangat Baik
X14	Pemecahan Masalah	Polynomial	Sangat Jelek, Jelek, Cukup, Baik, Sangat Baik
X15	Pengambilan Keputusan	Polynomial	Sangat Jelek, Jelek, Cukup, Baik, Sangat Baik

Pada tahap penelitian ini dilakukan perhitungan manual menggunakan C4.5 dengan 170 sample data, adapun langkah – langkah klasifikasi data tersebut dengan menggunakan algoritma C4.5 sebagai berikut :

1. Data training dibutuhkan, adapun data yang digunakan adalah 170 sample data training
2. Setelah data disiapkan, maka dilakukan proses penghitungan jumlah data karyawan yang efektif dan tidak efektif berdasarkan nilai atribut
3. Setelah melakukan proses penghitungan tersebut, maka dilakukan penghitungan nilai entropy total dimana diketahui pada proses penghitungan sebelumnya jumlah karyawan yang efektif berjumlah 154 dan karyawan yang tidak efektif berjumlah 16

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

$$= (-154/170) * \log_2(154/170) + (-16/170) * \log_2(16/170) = 0.2495$$

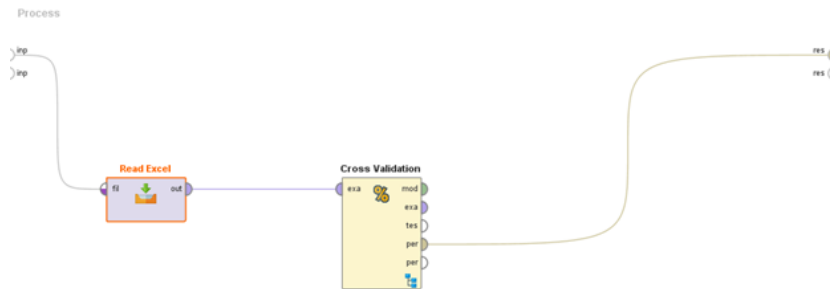
4. Setelah dilakukan penghitungan nilai entropy, maka dilakukan juga penghitungan nilai gain untuk masing – masing atribut sehingga diperoleh nilai gain tertinggi nantinya. Perhitungan nilai gain dilakukan untuk menentukan atribut dengan nilai gain tertinggi guna menjadikan atribut tersebut akar. Berikut adalah contoh salah satu atribut yang dilakukan penghitungan nilai gain nya

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

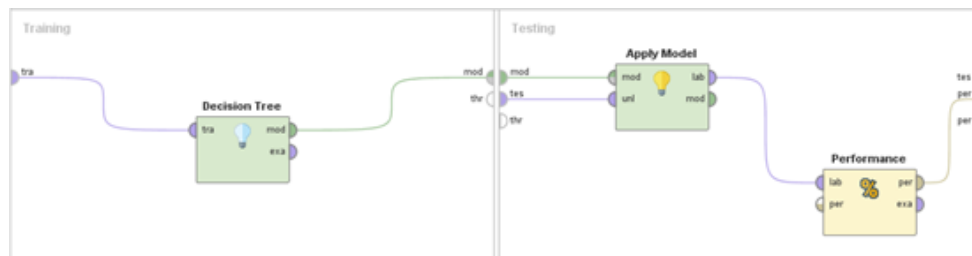
$$= 0,2495 - ((140/170 * 0,2432) + (30/170 * 0,2814)) = 0.003499$$

Untuk memperoleh nilai gain tertinggi dimana nilai gain yang diperoleh akan dijadikan sebagai akar maka harus ditentukan oleh hasil perhitungan nilai entropy dan gain dari semua atribut

Dengan data *training* yang sudah disiapkan ke dalam *confusion matrix* maka pengujian pertama sekali dilakukan dengan mencari nilai confusion matrix yang diperoleh dari tools rapid miner. tahapan dalam memperoleh nilai tersebut dapat dilihat pada gambar dibawah ini :



**Gambar 5.** Proses design dan import data dengan tools Rapid Miner.



**Gambar 6.** Design cross validation dan metode yang digunakan.

Berikut adalah hasil perhitungan confusion matrix untuk mencari nilai *Accuracy*, *precision* dan *recall* dari metode C4.5, Naïve Bayesian dan k-nearest neighbors :

**Tabel 5.** Nilai *Accuracy*, *Precision*, dan *Recall* untuk C4.5

	<i>True</i> Tidak Efektif	<i>True</i> Efektif
<i>pred</i> Tidak Efektif	9	7
<i>pred</i> Efektif	7	147

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{147+9}{147+9+7+7} = \frac{156}{170} = 0,9176 = 91,76\%$$

$$Precision = \frac{TP}{TP+FP} = \frac{147}{147+7} = \frac{147}{154} = 0,9545 = 95,45\%$$

$$Recall = \frac{TP}{TP+FN} = \frac{147}{147+7} = \frac{147}{154} = 0,9545 = 95,45\%$$

**Tabel 6.** Nilai *Accuracy*, *Precision*, dan *Recall* untuk Naive Bayes.

	<i>True</i> Tidak Efektif	<i>True</i> Efektif
<i>pred</i> Tidak Efektif	9	11
<i>pred</i> Efektif	7	143

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{143+9}{143+9+11+7} = \frac{152}{170} = 0,8941 = 89,41\%$$

$$Precision = \frac{TP}{TP+FP} = \frac{143}{143+11} = \frac{143}{154} = 0,9285 = 92,85\%$$

$$Recall = \frac{TP}{TP+FN} = \frac{143}{143+7} = \frac{143}{150} = 0,9533 = 95,33\%$$



**Tabel 7.** Nilai *Accuracy*, *Precision*, dan *Recall* untuk k-nearest neighbors.

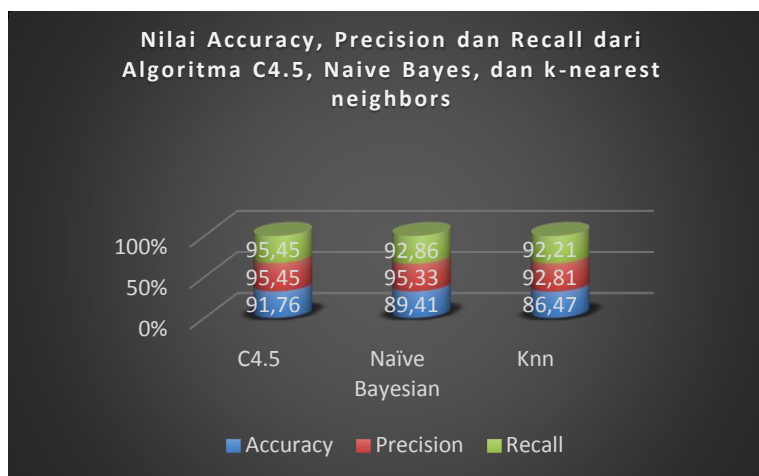
	<i>True Tidak Efektif</i>	<i>True Efektif</i>
<i>pred Tidak Efektif</i>	5	12
<i>pred Efektif</i>	11	142

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{142+5}{142+5+12+11} = \frac{147}{170} = 0,8647 = 86,47\%$$

$$Precision = \frac{TP}{(TP+FP)} = \frac{142}{142+12} = \frac{142}{154} = 0,9220 = 92,21\%$$

$$Recall = \frac{TP}{(TP+FN)} = \frac{142}{142+11} = \frac{142}{153} = 0,9281 = 92,81\%$$

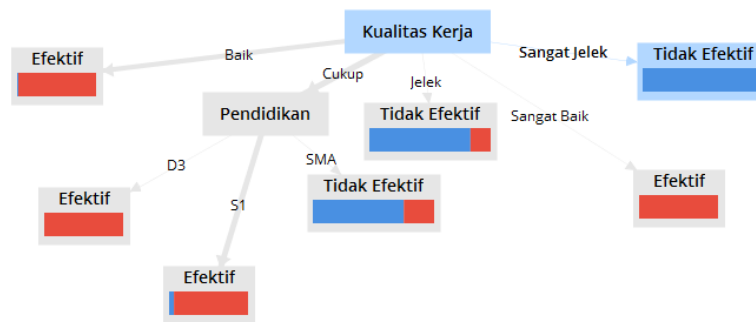
Berdasarkan perhitungan nilai yang diperoleh diatas maka nilai dari masing - masing metode dapat dilihat pada grafik di bawah ini sebagai gambaran perbandingan nilai dari ketiga metode tersebut.



**Gambar 7.** Grafik Nilai *Accuracy*, *Precision* dan *Recall*

Sistem rekomendasi akan dianggap baik jika nilai *precision* dan *recall*nya tinggi, pada diagram di atas diketahui bahwa yang memiliki nilai *recall* dan nilai *Precision* yang paling tinggi adalah algoritma C4.5 tidak hanya nilai *Precision* dan *recall*nya bahkan untuk nilai akurasi C4.5 lebih tinggi dibandingkan kedua algoritma tersebut [14]. Akurasi dalam klasifikasi adalah persentase ketepatan record data yang diklasifikasikan secara benar setelah dilakukan pengujian pada hasil klasifikasi.

Pohon keputusan yang terbentuk dari algoritma C4.5 yang telah diketahui information gain nya dapat dilihat pada gambar 8, dari pohon keputusan tersebut akan membentuk aturan yang dapat diterapkan



**Gambar 8.** Gambar pohon keputusan.

Dari pohon keputusan yang telah dibuat maka aturan yang akan diterapkan adalah sebagai berikut:

- Kualitas Kerja = Baik: Efektif {Tidak Efektif=1, Efektif=68}
- Kualitas Kerja = Cukup
  - | Pendidikan = D3: Efektif {Tidak Efektif=0, Efektif=2}
  - | Pendidikan = S1: Efektif {Tidak Efektif=5, Efektif=76}
  - | Pendidikan = SMA: Tidak Efektif {Tidak Efektif=3, Efektif=1}
- Kualitas Kerja = Jelek: Tidak Efektif {Tidak Efektif=5, Efektif=1}
- Kualitas Kerja = Sangat Baik: Efektif {Tidak Efektif=0, Efektif=6}
- Kualitas Kerja = Sangat Jelek: Tidak Efektif {Tidak Efektif=2, Efektif=0}

Pada penelitian ini dilakukan perhitungan menggunakan data testing, perhitungan dilakukan untuk memperoleh nilai akurasi. Pada aplikasi, data testing dilakukan dengan mempartisi data yang ada di dalam system. proses partisi data dilakukan dengan random. Berikut proses partisi data yang dimaksud :

**Tabel 8.** Partisi Data Testing

Status Data	Data Training	Data Testing
Ya (154 Data)	92	62
Tidak (16 Data)	10	6
170 Data	102 Data	68 Data

Dari proses partisi data sebesar 60 % maka diperoleh 68 data testing, dimana setelah dilakukan proses penyesuaian data maka di peroleh 8 data yang tidak sesuai dan ada 60 data yang sesuai. Langkah selanjutnya dilakukan proses pencarian nilai *Accuracy* dan error rate sebagai berikut:

$$Accuracy = \frac{Jumlah\ Pengujian\ Benar}{Jumlah\ Pengujian} = \frac{60}{68} = 0,8823 = 88,23\%$$

$$Error\ rate = \frac{Jumlah\ Pengujian\ Salah}{Jumlah\ Pengujian} = \frac{8}{68} = 0,1176 = 11,76\%$$

#### 4. KESIMPULAN

Dari hasil proses penelitian yang dilakukan dapat disimpulkan bahwa algoritma C4.5 memiliki kinerja (*precision*, *recall*, dan *Accuracy*) yang lebih baik dibandingkan dengan Naive Bayes, dari k-nearest neighbors sehingga penilaian kinerja karyawan dengan studi kasus pada penelitian ini dapat dianggap baik. Nilai *precision* dan *recall* yang diperoleh 95,45%, sedangkan untuk *Accuracy* 91,76%. Berdasarkan nilai kinerja tersebut maka penelitian ini bisa dikatakan lebih baik dibandingkan dengan penelitian sebelumnya karena atribut yang digunakan lebih banyak dan nilai kinerja yang diperoleh di atas 90 %

#### 5. SARAN

Penelitian ini dapat dikembangkan lagi dengan menggunakan metode klasifikasi yang berbeda atau menambahkan jumlah metodenya, serta dapat juga memanfaatkan aplikasi pengujian selain aplikasi rapid miner.

#### UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada keluarga dan sahabat yang telah memberi dukungan moril terhadap penelitian ini.

#### DAFTAR PUSTAKA

- [1] Brahmasari, I. A., & Suprayetno, A., 2008, Pengaruh motivasi kerja, kepemimpinan dan budaya organisasi terhadap kepuasan kerja karyawan serta dampaknya pada kinerja perusahaan (Studi kasus pada PT. Pei Hai International Wiratama Indonesia), *Jurnal Manajemen dan kewirausahaan*, vol 10, no 2, hal 124-135.
- [2] Jananto, A., 2013, Algoritma Naive Bayes untuk Mencari Perkiraan Waktu Studi Mahasiswa, *Dinamik*, vol 18, no 1, hal 9-16.
- [3] Astuti, D. A. L. W., 2006, Penciptaan sistem penilaian kinerja yang efektif dengan assessment centre, *Jurnal Manajemen Maranatha*, vol 6, no 1, hal 23-34.
- [4] Sylvia, T., Hidayat, A., & Putri, S. A., 2013, Penilaian Kinerja Karyawan Bagian Personalial Berdasarkan Kompetensi dengan Menggunakan Metode Analytic Network Process (ANP) dan Rating Scale, *Industria: Jurnal Teknologi dan Manajemen Agroindustri*, vol 2, no 2.
- [5] Rahmayuni, I., 2014, Perbandingan Performansi Algoritma C4.5 Dan Cart dalam Klasifikasi Data Nilai Mahasiswa Prodi Teknik Komputer Politeknik Negeri Padang, *Jurnal TeknoIf*, vol 2, no 1.
- [6] Fadillah, A. P., 2015, Penerapan Metode CRISP-DM untuk Prediksi Kelulusan Studi Mahasiswa Menempuh Mata Kuliah (Studi Kasus Universitas XYZ), *Jurnal Teknik Informatika Dan Sistem Informasi*, vol 1, no, 3, hal 260-269.
- [7] Mulyadi, M., 2011, Penelitian Kuantitatif dan Kualitatif serta Pemikiran Dasar Menggabungkannya, *Jurnal Studi Komunikasi dan Media*, vol 15, no 1, hal 127-138.
- [8] Fatmawati., 2016, Perbandingan Algoritma Klasifikasi *Data mining* Model C4.5 dan Naive Bayes untuk Prediksi Penyakit Diabetes, *Jurnal Techno Nusa Mandiri*, vol 13, no 1, hal 50-59.
- [9] Prahudaya, T. Y., & Harjoko, A., 2017, Metode Klasifikasi Mutu Jambu Biji Menggunakan Knn Berdasarkan Fitur Warna Dan Tekstur, *Jurnal Teknosains*, vol 6, no 2, hal 113-123.

- [10] Andriani, A., 2013, Sistem Pendukung Keputusan Berbasis Decision Tree Dalam Pemberian Beasiswa Studi Kasus: Amik “Bsi Yogyakarta”, *Seminar Nasional Teknologi Informasi Dan Komunikasi*, vol 2013, hal 163-168.
- [11] Kori, A., 2017, Comparative Study of Data Classifiers Using Rapidminer, *IJEDR*, vol 5, no 2, hal 1041-1043.
- [12] Celik, U., & Basarir, C., 2017, The Prediction of Precious Metal Prices via Artificial Neural Network by Using RapidMiner, *Alphanumeric Journal*, vol 5, no 1, hal 45-45.
- [13] Retnosari, P., & Jananto, A., 2013, Implementasi *data mining* untuk menemukan hubungan antara kota kelahiran mahasiswa dengan tingkat kelulusan mahasiswa pada fakultas teknologi informasi unisbank, vol 5, no 2.
- [14] Mayadewi, P., & Rosely, E., 2015, Prediksi Nilai Proyek Akhir Mahasiswa Menggunakan Algoritma Klasifikasi Data Mining, *SESINDO 2015*.