

# PENERAPAN METODE NAÏVE BAYES DAN C4.5 PADA PENERIMAAN PEGAWAI DI UNIVERSITAS POTENSI UTAMA

*APPLICATION OF NAÏVE BAYES AND C4.5 METHODS IN RECEIVING EMPLOYEES IN  
UNIVERSITY OF POTENSI UTAMA*

**Cindy Paramitha Lubis<sup>1</sup>, Dr. Zakarias Situmorang, MT<sup>2</sup>**

<sup>1</sup>Mahasiswa Program Studi Pascasarjana Ilmu Komputer, Fakultas Teknik dan Ilmu Komputer

<sup>2</sup>Dosen Program Studi Pascasarjana Ilmu Komputer, Fakultas Teknik dan Ilmu Komputer  
Universitas Potensi Utama, Jl.K.L. Yos Sudarso KM.6,5 No. 3A, Tj. Mulia, Medan

E-mail : cindyparamitha96@gmail.com<sup>1</sup>, zakarias65@yahoo.com<sup>2</sup>

## **ABSTRAK**

*Penerapan metode Naïve Bayes dan C4.5 dibuat untuk digunakan terhadap seleksi dan klasifikasi calon pegawai yang berpotensi untuk masuk ke dalam kampus dengan cara membuat perhitungan dari persamaan pada setiap kriteria. Permasalahan yang sering ditemukan adalah tidak efektifnya penggunaan metode yang digunakan untuk menghasilkan penerimaan pegawai yang di perlukan sehingga belum sesuai dengan bidang keahlian bagi pelamar. Metode Naïve Bayes dan C4.5 tersebut merupakan metode klasifikasi yang diterapkan pada data mining. Tujuan dibuatnya penelitian ini untuk menentukan tingkat akurasi antara kedua metode tersebut berdasarkan ketepatan perhitungan *Correctly Classified Instance* dan *Incorrectly Classified Instance*. Pengujian metode pada penelitian ini dilakukan dengan menggunakan tools Weka 3.8. Hasil yang didapat Pada metode Naïve Bayes tingkat akurasi yang didapat 77,7778% dan C4.5 memiliki tingkat akurasi 94,444% dari 36 data latih berhasil diuji. Sehingga hasil yang didapat C4.5 merupakan metode yang lebih tepat di gunakan dari pada Naïve Bayes.*

**Kata Kunci** : Naïve Bayes, C4.5 dan Akurasi

## **ABSTRACT**

*The application of the Naïve Bayes and C4.5 methods is made to be used for the selection and classification of prospective employees who have the potential to enter the campus by making a calculation of the equation for each criterion. The problem that is often found is the ineffective use of the method used to produce the necessary recruitment of employees so it is not in accordance with the field of expertise for applicants. Naïve Bayes and C4.5 methods are classification methods that are applied to data mining. The purpose of this research is to determine the level of accuracy between the two methods based on the accuracy of *Correctly Classified Instance* and *Incorrectly Classified Instance* calculations. Testing methods in this study were carried out using Weka 3.8 tools. The results obtained in the Naïve Bayes method obtained an accuracy rate of 77.7778% and C4.5 has an accuracy rate of 94.444% from 36 training data successfully tested. So the results obtained by C4.5 are a more appropriate method to be used than Naïve Bayes.*

**Keywords**: Naïve Bayes, C4.5 and Accuracy

## 1. PENDAHULUAN

Klasifikasi Bayes adalah penjumlahan statistic yang dapat memprediksi dati sutau probabilitas dari setiap klasifikasi. Klasifikasi Bayes yang terlihat sederhana lebih dikenal sebagai naïve Bayes Classifier dapat di artikan bahwa efek dari sebuah kelas atribut yang dihasilkan adalah nilai bebas dari setiap atribut. Asumsi ini disebut class conditional independence yang dibuat untuk

memudahkan pada perhitungan yang dianggap “naive”, dalam bahasa lebih sederhana naïve lagi yaitu itu mendefinisikan bahwa kemunculan suatu term dalam sebuah kalimat tidak mempengaruhi kemungkinan kata-kata yang lain padahal dalam kenyataanya bahwa kemungkinan kata dalam kalimat sangat dipengaruhi kemungkinan keberadaan kata-kata yang dalam kalimat[1].

Data mining merupakan tahap pengklasifikasian data dengan jumlah data yang besar dengan menghubungkan tiap pola yang menghubungkan pada setiap set data yang berukuran besar. Data mining dapat diartikan juga dengan menggali data dari banyaknya informasi yang akan dicari sehingga data yang perlu diketahui akan lebih mudah dicari dengan adanya sistem pola yang dibuat berdasarkan titik terdekat dengan informasi yang sering diperlukan[2].

Naïve Bayes merupakan metode yang digunakan untuk mengembangkan klasifikasi yang akan dijabarkan untuk meningkatkan akurasi dari tiap unit permasalahan yang diselesaikan. Metode yang terkait halnya dilakukan dengan menyiapkan perkalian dan penjumlahan tiap matriks untuk mendapatkan probabilitas tertentu dari tiap atribut yang ada[3].

Metode C4.5 merupakan suatu metode yang digunakan dalam data mining untuk klasifikasi dalam memprediksi dan sudah terbukti *powerfull*[4]. Metode ini merupakan metode yang berfungsi untuk merubah suatu informasi menjadi sebuah pohon keputusan yang dapat mudah untuk dimengerti dengan bahasa yang alami. Proses dari *Decision Tree* ini dimulai dari *node* akar hingga *node* daun yang dibuat dengan rekursif dimana setiap percabangan memiliki kondisi dan setiap ujungnya menghasilkan keputusan[4].

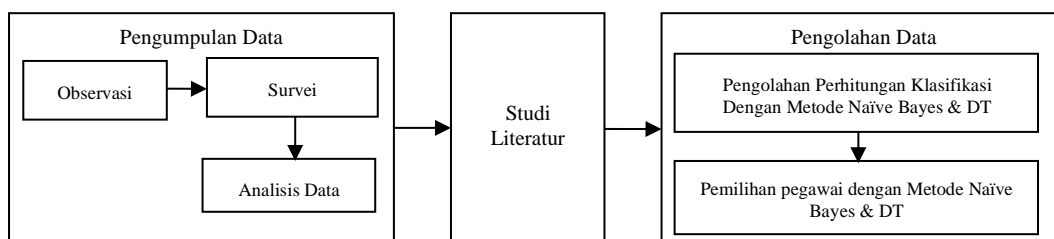
Penerapan kedua metode tersebut dilakukan untuk mencari tingkat akurasi yang lebih tepat dan akurat dalam menentukan penerimaan pegawai tersebut agar hasil dari klasifikasi yang digunakan. Hasil yang didapatkan menggunakan aplikasi dari Weka 3.8 yang dapat dilihat nilai yang *correct and Incorrect* pada aplikasi tersebut. Nilai akurasi yang didapat berdasarkan nilai yang berhasil diidentifikasi dan tidak berhasil diidentifikasi[5].

Pegawai merupakan sumber daya manusia yang dapat memenuhi kompetensi yang ada dan dapat bersaing dengan dunia luar. Pegawai dalam tingkat perguruan tinggi yang tentu saja harus memiliki berbagai ilmu yang berkompetensi terhadap bidang yang susai dengan bagian yang harus dipenuhi[4]. Keilmuan dan keterampilan dari seorang pegawai di kampus dapat mendukung sarana dan prasarana[4]. Menurut *Human Resource Development (HRD)* di Universitas Potensi Utama menyatakan ada beberapa faktor yang menentukan apakah seseorang dapat diterima. Faktor-faktor tersebut adalah: usia yang produktif, pendidikan terakhir, pengalaman bekerja, jenis kelamin, nilai tes tertulis dan komputer, pengecekan kesehatan, dan perilaku atau tes saat wawancara.

## 2. METODE PENELITIAN

### 2.1. Teknik Pengumpulan Data

Adapun metodologi penelitian dari penerapan metode naïve bayes pada penerimaan pegawai di Universitas Potensi Utama pada gambar 1



Gambar 1. Alur Analisis Penelitian

Tahap awal adalah pengumpulan data melalui obseravasi langsung kelapangan dan mensurvei data dengan wawancara kebagian HRD dan menganalisis data yang digunakan, tahap berikutnya studi literatur metode dengan pengolahan data melalui setiap kriteria penerimaan pegawai dan melakukan perhitungan dengan metode Naïve Bayes dan membuat Pohon Keputusan, melakukan pembuatan.

## 2.2. Metode Naïve Bayes

Naive Bayes merupakan sebuah klasifikasi didasarkan dengan perhitungan probabilitas sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari satu dataset yang digunakan. Algoritma menggunakan teorema Bayes dan mengasumsikan beberapa atribut yang independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel pada tiap kelas[6].

Naïve Bayes lebih tepat digunakan pada data yang besar dan dapat menangani data yang dianggap tidak lengkap serta kuat terhadap atribur yang tidak relevan pada data[7]. Naïve Bayes juga memiliki kelemahan dari nilai akurasi yang dapat terpengaruh dengan jumlah atribut[8]. Keuntungan penggunaan Naive Bayes adalah bahwa metode ini hanya butuh beberapa jumlah data pelatihan (Data Training) yang kecil untuk menentukan estimasi paremeter yang diperlukan dalam proses pengklasifikasian[9].

## 2.2. Persamaan Metode Naive Bayes

Persamaan dari teorema Bayes adalah [6]:

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

Di mana :

X : Kelas data yang belum diketahui.

H : Hipotesis merupakan suatu kelas yang lebih spesifik

P(H|X) : Probabilitas hipotesis nilai H berdasarkan kondisi nilai X (posteriori probabilitas)

P(H) : Probabilitas hipotesis H (prior probabilitas)

P(X|H) : Probabilitas nailai X berdasarkan kondisi pada hipotesis nilai H

P(X) : Probabilitas X

Untuk menjelaskan metode Naive Bayes, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis tersebut. Karena itu, metode Naive Bayes di atas disesuaikan sebagai berikut:

$$P(C|F1 \dots Fn) = \frac{P(C)P(F1\dots Fn|C)}{P(F1\dots Fn)} \quad (2)$$

Di mana Variabel C menampilkan kelas, sedangkan variabel F1 ... Fn menampilkan karakteristik petunjuk yang dibutuhkan untuk melaksanakan klasifikasi. Maka rumus tersebut menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam sebuah kelas C (Posterior) adalah peluang yang muncul dari kelas C (sebelum sampel tersebut masuk, dapat disebut prior), dikali dengan peluang munculnya karakteristik-karakteristik sampel pada kelas C (disebut juga *likelihood*), dibagi dengan peluang kemunculan karakteristik-karakteristik sampel secara global (disebut juga *evidence*). Karena itu, rumus di atas dapat pula ditulis secara sederhana sebagai berikut:

$$Posterior = \frac{\text{prior} \times \text{likelihood}}{\text{evidence}} \quad (3)$$

Nilai *Evidence* akan selalu tetap untuk setiap kelas pada setiap satu sampel. Nilai dari posterior tersebut nantinya akan dibandingkan dengan nilai-nilai posterior kelas lainnya untuk menentukan ke kelas apa suatu sampel akan diklasifikasikan. Penjabaran lebih lanjut rumus Bayes tersebut dilakukan dengan memaparkan  $(C|F_1, \dots, F_n)$  Rumus berikut ini akan menjelaskannya sebagai berikut:

$$\begin{aligned}
 & P(C|F_1, \dots, F_n) = \frac{P(C)P(F_1, \dots, F_n|C)}{P(F_1, \dots, F_n)} \\
 & = \frac{P(C)P(F_1|C)P(F_2, \dots, F_n|C, F_1)}{P(F_1, \dots, F_n)} \\
 & = \frac{P(C)P(F_1|C)P(F_2|C, F_1)P(F_3, \dots, F_n|C, F_1, F_2)}{P(F_1, \dots, F_n)} \\
 & = \frac{P(C)P(F_1|C)P(F_2|C, F_1)P(F_3|C, F_1, F_2)P(F_4, \dots, F_n|C, F_1, F_2, F_3)}{P(F_1, \dots, F_n)} \\
 & = \frac{P(C)P(F_1|C)P(F_2|C, F_1)P(F_3|C, F_1, F_2) \dots P(F_n|C, F_1, F_2, F_3, \dots, F_{n-1})}{P(F_1, \dots, F_n)}
 \end{aligned} \tag{4}$$

Penjelasan dari rumus tersebut dapat menghasilkan semakin banyak dan semakin rumitnya faktor yang dapat berpengaruh terhadap nilai probabilitas, yang mungkin saja mustahil untuk dianalisis pada setiap nilainya. Oleh karena itu, perhitungannya akan lebih sedikit rumit dilakukan. Rumus ini menggunakan asumsi yang independensi naif, karena setiap petunjuk ( $F_1, F_2, \dots, F_n$ ) tidak terikat atau bebas (independen) terhadap satu dengan yang lainnya. Dari penjelasan berikut, maka diterapkan suatu kesamaan sebagai berikut:

$$P(F_i|F_j) = \frac{P(F_i \cap F_j)}{P(F_j)} = \frac{P(F_i)P(F_j)}{P(F_j)} = P(F_i) \tag{5}$$

Untuk  $i \neq j$ , sehingga

$$P(F_i|C, F_j) = P(F_i|C) \tag{6}$$

Persamaan di atas merupakan model dari Naive Bayes yang selanjutnya akan digunakan dalam proses klasifikasi. Untuk klasifikasi dengan data kontinyu digunakan rumus Densitas Gauss :

$$P(X_i = x_i | Y = y_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} e^{-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2}} \tag{7}$$

Di mana :

- P : Peluang
- $X_i$  : Atribut nilai ke i
- $x_i$  : Nilai atribut ke i
- Y : Kelas yang dihitung
- $y_i$  : Sub kelas Y yang dicari
- $\mu$  : mean, merupakan rata – rata dari seluruh atribut
- $\sigma$  : Deviasi standar, menyatakan varian dari seluruh atribut.

### 2.3. Algoritma C4.5

Algoritma C4.5 merupakan algoritma yang diterapkan untuk proses pengklasifikasian data dengan menghasilkan sebuah pohon keputusan[10]. Pembuatan pohon keputusan ini berdasarkan dari atribut yang sudah di buat dan memiliki tingkat bobot yang tinggi pula. Algoritma ini juga memiliki nilai *gain* tertinggi berdasarkan nilai *entropy* atribut tersebut sebagai poros dari atribut klasifikasinya[4].

Menurut Jinaidi [4] data yang ada harus dibuat dalam sebuah tabel berdasarkan permasalahan dan jumlah responden sebelum melakukan perhitungan agar dapat mencari nilai *entropy* dan gain

$$Entropy(S) = \sum_{i=0}^n -p_i * \log_2 p_i \tag{1}$$

Rumus (1) merupakan perhitungan dari *entropy* yang digunakan untuk menentukan seberapa informatif atribut tersebut[11]. Berikut keterangannya:

$S$  : Himpunan dari Kasus

$n$  : Jumlah dari banyaknya partisi  $S$

$p_i$  : Jumlah kasus dari partisi ke- $i$

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (2)$$

Rumus (2) merupakan rumus yang digunakan dalam perhitungan *gain* setelah melakukan perhitungan *gain* setelah melakukan perhitungan *entropy*[12]. Berikut keterangannya :

$S$  : Himpunan dari Kasus

$n$  : Jumlah dari banyaknya partisi  $A$

$|S_i|$  : Jumlah kasus dari partisi ke- $i$

$|S|$  : Jumlah dari kasus  $S$

### 3. HASIL DAN PEMBAHASAN

#### 3.1. Penerapan Metode Naive Bayes

Naive Bayes memiliki penjelasan yang cukup sederhana bahwa nilai dari masing masing atribut secara kondisi independen jika nilai dari nilai hasil diberikan. Oleh karena itu, diberikan nilai keluaran, probabilitas individu dilihat dan dicek secara bersama-sama[6]. Keuntungan penggunaan Naive Bayes adalah dengan pengklasifikasian menggunakan data latih yang relative kecil sudah dapat menentukan estimasi dari parameter yang dibutuhkan. Pada metode Naive Bayes yang bersifat konstanta data string sering membedakan data numerik yang bersifat kontinyu, sehingga akan terlihat dari nilai probabilitas dari setiap kriteria. Kriteria tersebut dapat berupa nilai string maupun numerik[1]. Penerapan dari metode Naive Bayes dapat dilihat sebagai berikut.

##### 3.1.1. Data Latih

Dalam analisis metode Naïve Bayes terlebih dahulu harus melihat data apa yang akan digunakan dan menentukan data latih dan dilakukan perhitungan terhadap data tersebut yang kemudian bisa diuji dengan data yang lainnya. Data latih yang digunakan pada penelitian ini dapat dilihat pada tabel 2 berikut:

Tabel 2. Data Latih

No	Jenis Kelamin	USIA	Pendidikan Terakhir	Pengalaman Kerja	Nilai Test Tertulis	Nilai Test Komputer	Wawancara	Persetujuan HRD
1	L	Muda	S1	Belum Ada	Sangat Bagus	Sangat Bagus	Sangat Bagus	Diterima
2	L	Muda	SMA	Ada	Kurang Bagus	Kurang Bagus	Kurang Bagus	Ditolak
3	L	Muda	SMA	Belum Ada	Sangat Bagus	Kurang Bagus	Kurang Bagus	Ditolak
4	L	Muda	SMA	Ada	Bagus	Bagus	Sangat Bagus	Diterima
5	L	Muda	S1	Ada	Bagus	Sangat Bagus	Sangat Bagus	Diterima
6	L	Midle	SMA	Belum Ada	Kurang Bagus	Kurang Bagus	Bagus	Diterima
7	L	Midle	SMA	Belum Ada	Kurang Bagus	Kurang Bagus	Kurang Bagus	Ditolak
8	L	Tua	S1	Belum Ada	Kurang Bagus	Kurang Bagus	Kurang Bagus	Ditolak
9	P	Midle	SMA	Ada	Kurang Bagus	Kurang Bagus	Bagus	Diterima
10	L	Midle	S1	Belum Ada	Kurang Bagus	Kurang Bagus	Bagus	Ditolak
11	L	Midle	S1	Ada	Sangat Bagus	Bagus	Sangat Bagus	Diterima
12	P	Midle	SMA	Ada	Sangat Bagus	Sangat Bagus	Bagus	Diterima
13	P	Muda	S1	Ada	Kurang Bagus	Kurang Bagus	Kurang Bagus	Ditolak
14	P	Muda	S1	Ada	Sangat Bagus	Sangat Bagus	Sangat Bagus	Diterima
15	L	Muda	S1	Belum Ada	Kurang Bagus	Kurang Bagus	Kurang Bagus	Ditolak
16	P	Tua	S1	Belum Ada	Kurang Bagus	Bagus	Kurang Bagus	Ditolak

17	P	Muda	SMA	Ada	Bagus	Sangat Bagus	Bagus	Diterima
18	P	Midle	S1	Ada	Bagus	Bagus	Kurang Bagus	Ditolak
19	P	Midle	SMA	Ada	Bagus	Bagus	Sangat Bagus	Diterima
20	P	Muda	S1	Belum Ada	Kurang Bagus	Kurang Bagus	Sangat Bagus	Diterima
21	P	Midle	S1	Belum Ada	Bagus	Bagus	Sangat Bagus	Diterima
22	L	Muda	SMA	Belum Ada	Bagus	Sangat Bagus	Bagus	Diterima
23	L	Midle	SMA	Belum Ada	Bagus	Sangat Bagus	Sangat Bagus	Diterima
24	L	Midle	S1	Belum Ada	Bagus	Sangat Bagus	Sangat Bagus	Diterima
25	L	Midle	S1	Ada	Kurang Bagus	Kurang Bagus	Kurang Bagus	Ditolak
26	L	Muda	SMA	Ada	Sangat Bagus	Bagus	Kurang Bagus	Ditolak
27	P	Tua	SMA	Belum Ada	Kurang Bagus	Kurang Bagus	Kurang Bagus	Ditolak
28	P	Muda	SMA	Belum Ada	Kurang Bagus	Kurang Bagus	Bagus	Ditolak
29	L	Muda	SMA	Belum Ada	Kurang Bagus	Kurang Bagus	Bagus	Diterima
30	P	Muda	SMA	Belum Ada	Sangat Bagus	Bagus	Sangat Bagus	Diterima
31	P	Muda	SMA	Belum Ada	Kurang Bagus	Kurang Bagus	Bagus	Diterima
32	P	Midle	SMA	Ada	Bagus	Sangat Bagus	Kurang Bagus	Ditolak
33	P	Muda	SMA	Ada	Kurang Bagus	Kurang Bagus	Kurang Bagus	Ditolak
34	P	Midle	S1	Ada	Kurang Bagus	Kurang Bagus	Bagus	Diterima
35	L	Tua	SMA	Belum Ada	Kurang Bagus	Bagus	Kurang Bagus	Ditolak
36	L	Muda	S1	Belum Ada	Kurang Bagus	Kurang Bagus	Bagus	Diterima

### 3.1.2. Kriteria dan Probabilitas

Kriteria dan Probabilitas yang diterapkan harus dihitung dari masing – masing atribut dan kriteria. Data latih pada tabel 2 akan dilihat dari penjelasan pada setiap masing – masing atribut sebagai berikut :

#### 3.1.2.1. Probabilitas Kriteria Jenis Kelamin

Dari data latih pada tabel 2 diketahui jumlah data latih adalah sebanyak 36 data, berdasarkan data dari 36 data tersebut terdapat 2 jenis kelamin data dengan jumlah jenis kelamin perempuan yang diterima oleh HRD sejumlah 10 orang, 7 orang berjenis kelamin perempuan persetujuan HRD ditolak, 10 orang jenis kelamin laki-laki yang persetujuan HRD diterima sejumlah, 9 orang berjenis kelamin laki-laki yang persetujuan HRD ditolak, Probabilitas kriteria jenis kelamin dapat dilihat pada tabel 3.

Tabel 3. Probabilitas Kriteria Jenis Kelamin

Jenis Kelamin	Persetujuan HRD		Probabilitas	
	Diterima	Ditolak	Diterima	Ditolak
Laki-Laki	10	9	0,5	0,5625
Perempuan	10	7	0,5	0,5625
Jumlah	20	16	1	1,125

#### 3.1.2.2. Probabilitas Kriteria Usia

Dari data kriteria usia dapat dilihat dari 36 data terdapat sub atribut muda, midle, dan tua. Ada 11 orang berusia muda yang di setujui oleh HRD, 7 orang berusia muda yang ditolak oleh HRD, 9 orang yang berusia midle yang disetujui oleh HRD, 5 orang yang berusia midle yang tidak disetujui oleh HRD, Tidak ada yang diusia tua yang disetujui oleh HRD, 4 orang yang berusia tua yang tidak disetujui oleh HRD, Probabilitas kriteria jenis Usia dapat dilihat pada tabel 4.

Tabel 4. Probabilitas Kriteria Usia

Usia	Persetujuan HRD		Probabilitas	
	Diterima	Ditolak	Diterima	Ditolak
Muda	11	7	0,55	0,4375
Midle	9	5	0,45	0,3125
Tua	0	4	0	0,25
Jumlah	20	16	1	1

### 3.1.2.3. Probabilitas Kriteria Pendidikan Terakhir

Kriteria Pendidikan terakhir dapat diketahui dari 36 data terdapat sub kriteria S1 dan SMA. Ada 9 orang berpendidikan terakhirnya S1 yang di setujui oleh HRD, 7 orang berpendidikan S1 yang tidak disetujui oleh HRD, 11 orang berpendidikan terakhir SMA yang disetujui oleh HRD, 9 orang berpendidikan S1 yang tidak disetujui oleh HRD, Probabilitas kriteria jenis Pendidikan dapat dilihat pada tabel 5.

Tabel 5. Probabilitas Kriteria Pendidikan Terakhir

Pendidikan Terakhir	Persetujuan HRD		Probabilitas	
	Diterima	Ditolak	Diterima	Ditolak
S1	9	7	0,45	0,4375
SMA	11	9	0,55	0,5625
Jumlah	20	16	1	1

### 3.1.2.4. Probabilitas Kriteria Pengalaman Kerja

Kriteria pengalaman kerja dapat diketahui dari 36 data. Ada 9 orang memiliki pengalaman kerja yang di terima oleh HRD, 7 orang yang memiliki pengalaman kerja ditolak oleh HRD , 11 orang yang belum memiliki pengalaman kerja yang diterima oleh HRD, 9 orang yang tidak memiliki pengalaman kerja ditolak oleh HRD, Probabilitas kriteria pengalaman kerja dapat dilihat pada tabel 6.

Tabel 6. Probabilitas Kriteria Pengalaman Kerja

Pengalaman Kerja	Persetujuan HRD		Probabilitas	
	Diterima	Ditolak	Diterima	Ditolak
Ada Pengalaman	9	7	0,45	0,4375
Belum Ada Pengalaman	11	9	0,55	0,5625
Jumlah	20	16	1	1

### 3.1.2.5. Probabilitas Kriteria Nilai Tes Tertulis

Pada setiap calon pegawai dibidang pendidikan harus di tes secara tertulis dan hasil nilai yang didapatkan oleh calon pegawai kriteria dari tes tersebut dapat diketahui dari 36 data. Ada 5 orang memiliki nilai sangat bagus yang di terima oleh HRD, 2 orang yang memiliki nilai tes sangat bagus ditolak oleh HRD , 8 orang memiliki nilai tes bagus yang diterima oleh HRD, 2 orang memiliki nilai tes bagus ditolak oleh HRD, 7 orang memiliki nilai tes kurang bagus yang diterima oleh HRD, 12 orang memiliki nilai tes kurang bagus yang ditolak oleh HRD, Probabilitas kriteria nilai tes tertulis dapat dilihat pada tabel 7.

Tabel 7. Probabilitas Kriteria Nilai Tes Tertulis

Nilai Tes Tertulis	Persetujuan HRD		Probabilitas	
	Diterima	Ditolak	Diterima	Ditolak
Sangat Bagus	5	2	0,25	0,125
Bagus	8	2	0,4	0,125
Kurang Bagus	7	12	0,35	0,75
Jumlah	20	16	1	1

### 3.1.2.6. Probabilitas Kriteria Nilai Tes Komputer

Pada setiap calon pegawai harus di tes secara komputer dan hasil nilai yang didapatkan oleh calon pegawai kriteria dari tes tersebut dapat diketahui dari 36 data. Ada 8 orang memiliki nilai sangat bagus yang di terima oleh HRD, 1 orang yang memiliki nilai tes sangat bagus ditolak oleh HRD , 5

orang memiliki nilai tes bagus yang diterima oleh HRD, 4 orang memiliki nilai tes bagus ditolak oleh HRD, 7 orang memiliki nilai tes kurang bagus yang diterima oleh HRD, 11 orang memiliki nilai tes kurang bagus yang ditolak oleh HRD, Probabilitas kriteria nilai tes komputer dapat dilihat pada tabel 8.

Tabel 8. Probabilitias Kriteria Nilai Tes Komputer

Nilai Tes Komputer	Persetujuan HRD		Probabilitas	
	Diterima	Ditolak	Diterima	Ditolak
Sangat Bagus	8	1	0,4	0,0625
Bagus	5	4	0,25	0,25
Kurang Bagus	7	11	0,35	0,6875
Jumlah	20	16	1	1

### 3.1.2.7. Probabilitas Kriteria Kesehatan

Pada setiap HRD harus melihat riwayat kesehatan/penyakit dari si calon pegawai apakah baik atau tidak jika keterangan kesehatannya tidak baik alangkah lebih baiknya tidak diterima dan dapat diketahui dari 36 data. Ada 20 orang yang dilihat riwayat kesehatannya baik dan di terima oleh HRD, 16 orang yang dilihat riwayat kesehatannya baik yang ditolak oleh HRD , dan tidak ada yang mendaftar menjadi pegawai dengan riwayat kesehatannya tidak baik, Probabilitas kriteria kesehatan dapat dilihat pada tabel 9.

Tabel 9. Probabilitias Kriteria Kesehatan

Kesehatan	Persetujuan HRD		Probabilitas	
	Diterima	Ditolak	Diterima	Ditolak
Baik	20	16	1	1
Tidak Baik	0	0	0	0
Jumlah	20	16	1	1

### 3.1.2.8. Probabilitas Kriteria Wawancara

Pada atribut wawancara dapat diketahui dari 36 data. Ada 11 orang memiliki nilai wawancara sangat bagus yang di terima oleh HRD, tidak ada memiliki nilai sangat bagus yang ditolak oleh HRD , 9 orang yang memiliki nilai bagus yang diterima oleh HRD, 2 orang memiliki nilai bagus yang ditolak oleh HRD, tidak ada nilai kurang bagus yang diterima oleh HRD, 14 orang bernilai kurang bagus yang ditolak oleh HRD, Probabilitas pada kriteria wawancara dapat dilihat pada tabel 10.

Tabel 10. Probabilitias Kriteria Wawancara

Wawancara	Persetujuan HRD		Probabilitas	
	Diterima	Ditolak	Diterima	Ditolak
Sangat Bagus	11	0	0,55	0
Bagus	9	2	0,45	0,125
Kurang Bagus	0	14	0	0,875
Jumlah	20	16	1	1

## 3.2. Perhitungan Manual Metode Naïve Bayes dan C.45

Dari data diatas telah menjelas kan dari atribut yang digunakan dan nilai probabilitas dari setiap atribut yang ada. Pada perhitungan manual ini peneliti membuat dengan menggunakan *Microsoft Excel* sehingga dapat terlihat lebih jelas perhitungan dari probabilitas pada setiap masing-masing kriteria yang ada. Perhitungan ini menggunakan 36 Data latihan yang akan diuji.



## 3.2.1. Perhitungan Manual Metode Naïve Bayes

Perhitungan Manual Metode Naïve Bayes dapat dijelaskan pada tabel 11.

Tabel 11. Perhitungan Manual Metode Naïve Bayes

	Klasifikasi		Persetujuan = "Diterima"	Persetujuan = "Ditolak"
	Xc1=	Wawancara = Sangat Bagus	0,550	0,000
	Usia = Muda	0,550	0,438	
	Nilai Tes Komputer = Sangat Bagus	0,400	0,063	
	Nilai Tes Tertulis = Sangat Bagus	0,250	0,125	
	Jenis Kelamin = Laki-Laki	0,500	0,563	
	Pendidikan Terakhir = S1	0,450	0,438	
	Pengalaman Kerja = Belum Ada	0,450	0,438	
	#P(X Ci) : P(X persetujuan)	0,003	0,000	
	#P(X Ci)*P(Ci) = P(X persetujuan) *	0,002	0,000	
Xc2=	Wawancara = Kurang Bagus	0,350	0,688	
	Usia = Muda	0,550	0,438	
	Nilai Tes Komputer = Kurang Bagus	0,350	0,688	
	Nilai Tes Tertulis = Kurang Bagus	0,350	0,750	
	Jenis Kelamin = Laki-Laki	0,500	0,563	
	Pendidikan Terakhir = SMA	0,550	0,563	
	Pengalaman Kerja = Ada	0,450	0,438	
	#P(X Ci) : P(X persetujuan)	0,003	0,021	
	#P(X Ci)*P(Ci) = P(X persetujuan) *	0,002	0,010	
Xc3=	Wawancara = Kurang	0,350	0,688	
	Usia = Muda	0,550	0,438	
	Nilai Tes = Kurang	0,350	0,688	
	Nilai Tes = Sangat	0,250	0,125	
	Jenis Kelamin = Laki-Laki	0,500	0,563	
	Pendidikan = SMA	0,550	0,563	
	Pengalaman = Belum Ada	0,550	0,563	
	#P(X Ci) : P(X persetujuan)	0,003	0,005	
	#P(X Ci)*P(Ci) = P(X persetujuan) *	0,001	0,002	
Xc4=	Wawancara = Sangat	0,400	0,063	
	Usia = Muda	0,550	0,438	
	Nilai Tes = Kurang	0,350	0,688	
	Nilai Tes = Sangat	0,250	0,125	
	Jenis Kelamin = Laki-Laki	0,500	0,563	
	Pendidikan = SMA	0,550	0,563	
	Pengalaman = Ada	0,450	0,438	
	#P(X Ci) : P(X persetujuan)	0,002	0,000	
	#P(X Ci)*P(Ci) = P(X persetujuan) *	0,001	0,000	

## 3.2.2. Perhitungan Manual Metode C4.5

Pada perhitungan dengan Metode C4.5 terlebih dahulu harus mencari nilai Entropy dan nilai Gain sehingga dapat kita paparkan lebih jelas lagi. Berikut hasil perhitungan manual :

1. Menghitung Jumlah yang "Diterima" dan "Ditolak" dari data latih dan mengetahui nilai entropy. Jumlah kasus yang diterima sebanyak 20 dan ditolak sebanyak 16. Total keseluruhan menjadi 36 orang. Sehingga dapat dihitung nilai entropy keseluruhan :

$$Entropy = - \sum_{j=1}^m f(i,j) \cdot \log_2 f(i,j)$$

$$= (-20/36 * \log_2(20/36)) + (-16/36 * \log_2(16/36))$$

$$= 0,9911$$

2. Menghitung Entropy dan gain dari masing-masing atribut yaitu Jenis Kelamin, Usia, Pendidikan terakhir, pengalaman berkerja, Nilai tes tertulis, nilai tes computer, wawancara, dan surat kesehatan. Berdasarkan 36 data tersebut terdapat 2 jenis kelamin data dengan jumlah jenis kelamin perempuan yang diterima oleh HRD sejumlah 10 orang, 7 orang berjenis kelamin perempuan persetujuan HRD ditolak, 10 orang jenis kelamin laki-laki yang persetujuan HRD diterima sejumlah, 9 orang berjenis kelamin laki-laki yang persetujuan HRD ditolak, Sehingga nilai tersebut dapat terlihat dari tabel 12.

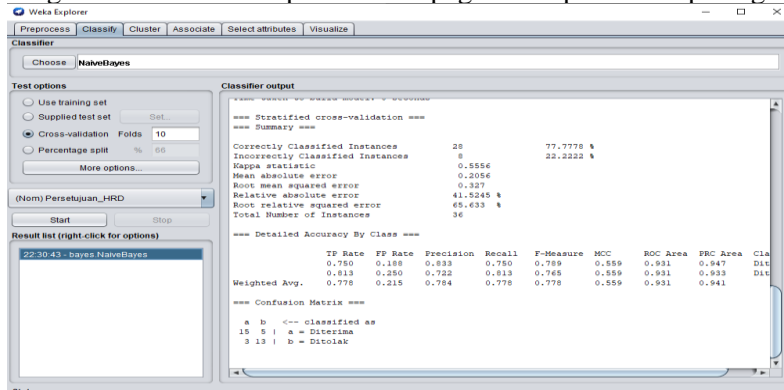
Tabel 12. Perhitungan Manual C4.5

1. Jenis Kelamin			log(x)	log(y)	Nilai	Execute Nilai
Laki-Laki	Diterima	10	0,487	0,511	0,998	0,527
	Ditolak	9				
	JumlahData	19				
Perempuan	Diterima	10	0,450	0,527	0,977	0,462
	Ditolak	7				
	JumlahData	17				
JUMLAH NILAI KESELURUHAN						0,988
Nilai Entropy						0,991

Total	0,003
-------	-------

### 3.3. Pengujian Metode Naïve Bayes

Dari nilai probabilitas di atas akan diuji data sebanyak 36 data dan diuji dengan menggunakan tools weka sehingga menghasilkan klasifikasi penerimaan pegawai dapat dilihat pada gambar 2 berikut:



Gambar 2. Hasil Klasifikasi Metode Naive Bayes

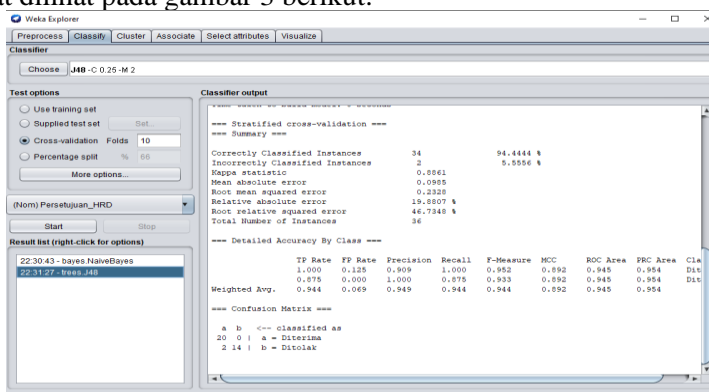
Sumber : Aplikasi Weka 3.8.

Pada gambar 2 di atas mendapatkan persentase untuk Correctly Classified Instance adalah sebesar 77,7778% sementara persentase untuk Incorrectly Classified Instance adalah sebesar 22,222%. Dari data 36 penerimaan pegawai, ada sebanyak 28 data yang disetujui oleh HRD berhasil diklasifikasikan dengan benar dan sebanyak 8 data yang disetujui oleh HRD tidak berhasil diklasifikasikan dengan benar.

Berdasarkan hasil penjumlahan yang telah dibahas sebelumnya ada 8 data yang tidak sesuai perhitungannya dapat dilihat pada kriteria 2,10,13,15,18,28,30,35. Sehingga mempengaruhi nilai akurasi yang didapatkan menjadi 77,778% dengan 22,222% data dapat teridentifikasi dengan baik.

### 3.4. Pengujian Dengan Metode C4.5

Hasil dari probabilitas di atas maka akan diuji data sebanyak 36 data dengan menggunakan tools weka sehingga menghasilkan klasifikasi penerimaan pegawai yang memiliki tingkat akurasi 94.444% sementara persentase untuk Incorrectly Classified Instance adalah sebesar 5,5666%. Di mana dari 36 penerimaan pegawai, ada sebanyak 34 data yang disetujui oleh HRD berhasil diklasifikasikan dengan benar dan sebanyak 2 data yang disetujui oleh HRD tidak berhasil diklasifikasikan dengan benar. Hasil tersebut dapat dilihat pada gambar 3 berikut:

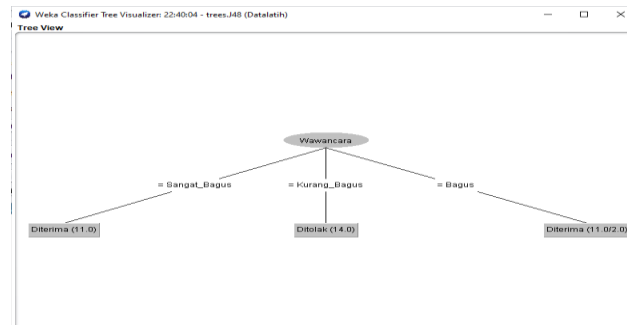


Gambar 3. Hasil Algoritma C4.5

Sumber : Aplikasi Weka 3.8.

Berdasarkan hasil penjumlahan yang telah dibahas sebelumnya ada 2 data yang tidak sesuai perhitungannya dapat dilihat pada kriteria 13 dan 15. Sehingga mempengaruhi nilai akurasi yang didapatkan menjadi 94,4444% dengan 5,5556% data dapat teridentifikasi dengan baik.

Dari hasil yang didapat dari aplikasi weka terbut maka dapat diperoleh sebuah pohon keputusan berdasarkan hasil yang didapat dari yang sudah di implementasikan sebelum. Hasil pohon keputusan tersebut dapat dilihat dari gambar 4 berikut ini :



Gambar 4. Hasil Pohon Keputusan C4.5

Sumber : Aplikasi Weka 3.8.

### 3.5. Hasil Perbandingan kedua metode

Berdasarkan hasil yang didapat pada implementasi yang diterapkan dengan menggunakan tools weka 3.8 maka perbandingan tersebut dapat dilihat dari akurasi yang didapat sebagai berikut :

Tabel 13. Hasil Akurasi

Naïve Bayes	C4.5
77,7778%	94.444%

Dari hasil akurasi diatas dapat ditarik kesimpulan bahwa metode C4.5 lebih akurat dari pada metode Naïve Bayes sehingga dapat di gunakan dan lebih efisien dengan adanya pohon keputusan tersebut. Perlu diketahui tingkat akurasi akan berbeda jika kita lebih memiliki banyak data lagi untuk dicoba dan data tersebut merupakan data yang valid sehingga dapat lebih efektif melakukan uji joba data latih dan data testing.

## 4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan pada penerapan metode naïve bayes untuk penentuan penerimaan pegawai di Universitas Potensi Utama, dari penjelasan diatas maka menghasilkan kesimpulan sebagai berikut:

1. Metode Naïve Bayes dan C4.5 diterapkan untuk menentukan penerimaan pegawai yang diterima oleh Universitas Potensi Utama berdasarkan klasifikasi yang diuji untuk menentukan setiap kriteria seperti jenis kelamin, usia, pendidikan terakhir, pengalaman kerja, nilai tes tertulis, nilai tes computer, surat kesehatan, dan interview.
2. Berdasarkan data penerimaan pegawai yang dijadikan data latih, metode Naive Bayes memiliki tingkat akurasi 77,7778% sedangkan C4.5 memiliki tingkat akurasi 94,444%
3. Sehingga dapat ditarik kesimpulan bahwa metode C4.5 merupakan metode yang lebih akurat dan efisien dari pada Metode Naïve Bayes

## 5. SARAN

Sebagai penelitian ini mengklasifikasikan kriteria penentuan penerimaan pegawai pengujian tersebut sebaiknya dilakukan dengan menggunakan metode lain atau membandingkan dengan yang sudah pernah ada untuk dapat melihat metode perbandingan yang lebih akurat dalam memprediksi penerimaan pegawai.

DAFTAR PUSTAKA

- [1] A. Saleh, "Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga," *Creat. Inf. Technol. J.*, vol. 2, no. 3, pp. 207–217, 2015.
- [2] A. Amrin, "Perbandingan Metode Neural Network Model Radial Basis Function Dan Multilayer Perceptron Untuk Analisa Risiko Kredit Mobil," *Paradigma*, vol. XX, no. 1, pp. 31–38, 2018.
- [3] A. Saleh and F. Nasari, "Penggunaan Teknik Unsupervised Discretization pada Metode Naive Bayes dalam Menentukan Jurusan Siswa Madrasah Aliyah," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 3, p. 353, 2018, doi: 10.25126/jtiik.201853705.
- [4] F. F. Harryanto and S. Hansun, "Penerapan Algoritma C4.5 untuk Memprediksi Penerimaan Calon Pegawai Baru di PT WISE," *J. Tek. Inform. Dan Sist. Inf.*, vol. 3, no. 2, pp. 95–103, 2017.
- [5] A. Tulis, K. Di, and C. V Mega, "Penerapan apriori pada penjualan produk alat tulis kantor di cv. mega stationery," pp. 561–571.
- [6] H. Muhamad, C. A. Prasojo, N. A. Sugianto, L. Surtiningsih, and I. Cholissodin, "Optimasi Naïve Bayes Classifier Dengan Menggunakan Particle Swarm Optimization Pada Data Iris," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 4, no. 3, p. 180, 2017, doi: 10.25126/jtiik.201743251.
- [7] R. N. Devita, H. W. Herwanto, and A. P. Wibawa, "Perbandingan Kinerja Metode Naive Bayes dan K-Nearest Neighbor untuk Klasifikasi Artikel Berbahasa indonesia," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 4, p. 427, 2018, doi: 10.25126/jtiik.201854773.
- [8] M. Ridwan, H. Suyono, and M. Sarosa, "Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier," *Eeccis*, vol. 7, no. 1, pp. 59–64, 2013, doi: 10.1038/hdy.2009.180.
- [9] S. Syarli and A. Muin, "Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi)," *J. Ilm. Ilmu Komput.*, vol. 2, no. 1, pp. 22–26, 2016.
- [10] B. Sugara, D. Widyatmoko, B. S. Prakoso, and D. M. Saputro, "Penerapan Algoritma C4.5 untuk Deteksi Dini Autisme Pada Anak," *Semin. Nas. Teknol. Inf. dan Komun.*, vol. 2018, no. Sentika, pp. 87–96, 2018.
- [11] N. Azwanti, "Algoritma C4.5 Untuk Memprediksi Mahasiswa Yang Mengulang Mata Kuliah (Studi Kasus Di Amik Labuhan Batu)," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 9, no. 1, pp. 11–22, 2018, doi: 10.24176/simet.v9i1.1627.
- [12] A. Andriani, "Penerapan Algoritma C4.5 Pada Program Klasifikasi Mahasiswa Dropout," *Semin. Nas. Mat. 2012*, pp. 139–147, 2012.