



Perbandingan Tingkat Akurasi Klasifikasi Penerimaan Dosen Tetap Menggunakan Metode *Naive Bayes Classifier* dan C4.5

Muhammad Sadikin^{1*}, Rika Rosnelly¹, Roslina², Teddy Surya Gunawan¹, Wanayumini¹

¹ Fakultas Teknik dan Ilmu Komputer, Prodi Pascasarjana Ilmu Komputer, Universitas Potensi Utama, Medan, Indonesia

² Prodi Manajemen Informatika, Politeknik Negeri Medan, Medan, Indonesia

E-mail : ^{1*}dicky.aries.3@gmail.com, ²rikarosnelly@gmail.com, ³roslina@polmed.ac.id, ⁴tsgunawan@gmail.com,

⁵wanayumini@gmail.com

Email Penulis Korespondensi: dicky.aries.3@gmail.com

Abstrak—Rekrutmen merupakan salah satu langkah penting dalam menciptakan SDM (Sumber Daya Manusia) yang profesional. Penerapan metode klasifikasi seperti metode *Naive Bayes* dan C4.5 dapat digunakan dalam pengklasifikasian calon dosen yang berpotensi dan dapat diterima oleh kampus dengan melakukan perhitungan dari persamaan pada masing-masing kriteria. Kesulitan yang dialami adalah tidak efektifnya penggunaan metode untuk menghasilkan penerimaan dosen yang dibutuhkan sehingga belum sesuai dengan keahlian pelamar. Salah satu metode klasifikasi yang diterapkan pada data mining adalah metode *naive bayes* dan C4.5. Tujuan penelitian ini untuk menentukan tingkat akurasi terhadap kedua metode yang digunakan dengan menggunakan tools Weka 3.8 berdasarkan hasil perhitungan *Correctly Classified Instance* dan *Incorrectly Classified Instance*. Hasil akurasi yang didapat dengan metode *naive bayes* sebesar 83,7838 % dan metode c4.5 sebesar 91,8919 % dari 37 data latih. Sehingga metode C4.5 merupakan metode yang lebih tepat untuk digunakan dari pada *naive bayes*.

Kata Kunci: Klasifikasi, Data Mining, Rekrutmen, *Naive Bayes*, Weka

Abstract—Recruitment is an important step in creating professional HR (Human Resources). The application of classification methods such as the *Naive Bayes* method and C4.5 can be used in the classification of potential lecturers and can be accepted by the campus by calculating the equations for each criterion. The difficulty experienced is the ineffective use of the method to generate the required lecturer acceptance so that it is not in accordance with the applicant's expertise. One of the classification methods applied to data mining is the *naive Bayes* method and C4.5. The purpose of this study is to determine the level of accuracy of the two methods used by using the Weka 3.8 tool based on the calculation of *Correctly Classified Instance* and *Incorrectly Classified Instance*. The accuracy results obtained with the *naive Bayes* method are 83.7838% and the C4.5 method is 91.8919% from 37 training data. So the C4.5 method is a more appropriate method to use than *naive Bayes*.

Keywords: Classification, Data Mining, Recruitment, *Naive Bayes*, Weka

1. PENDAHULUAN

Rekrutmen adalah salah satu langkah yang paling penting di dalam pemilihan SDM (Sumber Daya Manusia)[1]. Tujuan yang ingin ditempuh berdasarkan proses seleksi dan rekrutmen Sumber Daya Manusia adalah SDM yang profesional dan memiliki karakteristik tingkah laku mulia, terampil dan memiliki kontribusi[2]. Proses rekrutmen dan seleksi ini merupakan langkah utama dalam menentukan seorang SDM yang diterima apakah berkualitas dan profesional. SDM yang profesional tentunya memberikan *impact* dan *output* yang sangat berpengaruh yaitu berupa hasil ataupun jasa yang juga berkualitas. Produk dan jasa yang dihasilkan berkualitas tentunya berdampak pula kepada kesejahteraan terhadap organisasi, individu bahkan masyarakat luas.

Perguruan Tinggi Negeri (PTN) maupun Perguruan Tinggi Swasta (PTS) tentu saja berkeinginan memiliki dosen yang memiliki kualitas yang baik serta dengan kuantitas yang sesuai dengan harapan dan kriteria yang telah ditetapkan oleh perguruan tinggi tersebut. Kualitas dari calon pendidik dapat dinilai dengan memahami pola dari variable dosen yang sudah terlaksana di tahun-tahun sebelumnya salah satunya dengan mempertimbangkan nilai Index Prestasi Kumulatif (IPK) beserta nilai-nilai test lainnya sebagai acuan penerimaan calon dosen baru.

Data mining merupakan tahap dalam pengklasifikasian terhadap data dengan menghubungkan masing-masing pola pada setiap data set yang berukuran besar dengan jumlah data yang besar pula. Data mining dapat juga didefinisikan 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 di perlukan[3], [4]. Pemilihan suatu metode ataupun algoritma dari data mining ini juga harus tepat karena sangat bergantung pada tujuan dan proses *Knowledge Discovery in Database* (KDD) secara keseluruhan [5]. Secara teknis, data mining dapat disebut sebagai proses untuk menemukan aturan atau pola dari ratusan atau ribuan data dari sebuah relasi basis data yang sangat besar[6].

Berdasarkan hal tersebut, pada penelitian kali ini akan dilakukan perbandingan terhadap metode *naive bayes* dan C4.5 untuk dapat menganalisa dan mengklasifikasikan penerimaan dosen tetap dengan cara mengambil sample data pada penerimaan dosen Tahun Akademik 2015-2016 ganjil yang akan digunakan sebagai data latih. Data input diproses menggunakan algoritma *Naive Bayes Classifier* (NBC) dan C4.5 dalam membentuk tabel probabilitas sebagai dasar proses pengklasifikasian penerimaan dosen. Penerapan teknik data



mining dalam menemukan pola penerimaan dosen baru yang sudah diterima, kemudian diambil untuk menjadi acuan dalam memprediksi penerimaan dan rekrutmen dosen baru di tahun berikutnya.

Percobaan analisis data mining ini menggunakan salah satu *tool* yaitu Weka versi 3.8 yang berfungsi untuk mendapatkan pola dan prediksi tentang aturan dalam kumpulan data. *Tool* ini bertugas dalam hal mendapatkan pola klasifikasi dengan mengenali aturan dan fitur pada data. Diharapkan dengan menggunakan *tool* data mining ini mampu mengetahui pola dalam data dengan input minimal dari user[7].

Berdasarkan penelitian yang telah dilakukan sebelumnya oleh Endang Etriyanti, Dedy Syamsuar dan Yesi Novaria Kunang dengan judul penelitian “Implementasi Data Mining Menerapkan Algoritma *Naive Bayes Classifier* dan C4.5 untuk Memprediksi Kelulusan Terhadap Mahasiswa” memberikan hasil penelitian yang menggambarkan bahwa metode C4.5 lebih unggul dengan nilai selisih persentase akurasi kurang dari 1% dimana metode C4.5 sebesar 79,08% sedangkan *Naive Bayes Classifier* memberikan tingkat hasil akurasi sebesar 78,46% [8].

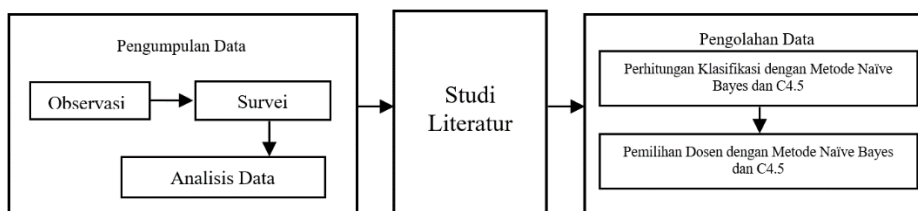
Penelitian lain juga dilakukan oleh Y A Gerhana, et All. dengan judul penelitian “Perbandingan *naive bayes classifier* dan algoritma C4.5 dalam memprediksi masa belajar siswa”, memberikan hasil akurasi dengan menerapkan metode *naive bayes* yang tidak kalah baik yaitu sebesar 88% sedangkan dengan metode C4.5 memberikan akurasi sebesar 87% yang artinya kedua metode tersebut memberikan hasil yang hampir serupa terhadap 109 data sample [9].

Erwina Nurul Azizah, et All juga melakukan penelitian yang berjudul “Kinerja komparatif antara pengklasifikasi C4.5 dan Naive Bayes dalam memprediksi kinerja akademik siswa di Lingkungan Belajar Virtual” memberikan hasil akurasi terhadap Algoritma *naive bayes* lebih baik dari pada algoritma C4.5, dengan nilai akurasi sebesar 63,8%, Hanya sedikit perbedaan dimana 63,6% dalam algoritma C4.5[10].

2. METODOLOGI PENELITIAN

2.1 Tahapan Analisis Penelitian

Adapun metodologi penelitian dari pada Penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Analisis Penelitian

Tahapan awal adalah pengumpulan data meliputi 3 bagian yaitu, observasi, survei dan analisis data yang dilakukan langsung di Universitas Potensi Utama dengan melakukan wawancara kebagian rekrutmen dosen. Tahapan berikutnya studi literatur terhadap metode yang digunakan dengan melakukan pengelolaan data setiap kriteria dalam rekrutmen dosen. Kemudian tahap akhir melakukan pengolahan data dengan menerapkan metode *Naive Bayes* dan C4.5 untuk membuat pohon keputusan.

2.2 Metode *Naive Bayes*

Naive Bayes merupakan suatu algoritma yang dapat bertugas dalam melakukan pengklasifikasian suatu variable dalam kumpulan basis data sesuai dengan perhitungan yang mengacu pada bidang ilmu matematika yaitu menggunakan metode perhitungan probabilitas dan statistik[11]. Salah satu metode klasifikasi yang mengarah pada cabang ilmu statistik atau matematika adalah metode *naive bayes* yang lebih dikenal dengan teori probabilitas dalam mencari peluang dari klasifikasi yang terjadi, dengan melihat frekuensi pola pada tiap klasifikasi pada data training[12]. Keuntungan dalam penggunaan algoritma Naive Bayes adalah algoritma atau metode ini tidak harus memiliki jumlah data latih (Training Data) yang banyak untuk menentukan estimasi parameter atau pola yang diperlukan dalam proses pengklasifikasian dan juga algoritma ini dapat beroperasi bahkan lebih baik dalam kasus dunia nyata yang sangat kompleks[13]. Persamaan dari teorema Bayes adalah[13]:

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

Keterangan:

X : Data dengan class yang belum diketahui

H : Hipotesis data merupakan suatu class spesifik

P(H|X) : Probabilitas hipotesis terhadap H berdasar kondisi X (posteriori probabilitas)



- P(H) : Probabilitas hipotesis terhadap H (prior probabilitas)
- P(X|H) : Probabilitas X berdasarkan kondisi pada hipotesis H
- P(X) : Probabilitas X

Proses klasifikasi dalam menentukan kelas apa yang cocok terhadap sample yang akan dianalisis memerlukan sebuah langkah yang dapat digambarkan pada persamaan 2 berikut.

$$P(C|F_1 \dots F_n) = \frac{P(C)P(F_1 \dots F_n|C)}{P(F_1 \dots F_n)} \tag{2}$$

Variabel C mendeskripsikan terhadap kelas, dan variabel F1 ... Fn menjelaskan karakteristik untuk melakukan pengklasifikasian. Rumus tersebut merepresentasikan karakteristik tertentu dalam kelas C (Posterior) adalah peluang munculnya kelas C yang dikalikan terhadap peluang kemunculan karakteristik sampel yang ada pada kelas C (disebut juga sebagai *likelihood*). Oleh sebab itu, rumus tersebut dapat disederhanakan menjadi sebagai persamaan 3.

$$Posterior = \frac{prior \times likelihood}{evidence} \tag{3}$$

Multinomial Naive Bayes atau yang kita ketahui *Naive Bayes Classifier* merupakan model dari penyederhanaan *Bayes*. Algoritma *Naive Bayes* mengasumsikan efek terhadap nilai variabel disebuah *class* yang ditentukan akan tidak saling terhubung terhadap variabel lain. Lebih lengkap dapat dilakukan dengan melakukan penjabaran nilai $(C|F_1, \dots, F_n)$ dengan menerapkan aturan perkalian yaitu sebagai berikut:

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

Dapat diketahui bahwa hasil dari penjelasan rumus bayes diatas menyimpulkan semakin lengkap kompleksitasnya serta syarat yang dapat mempengaruhi nilai terhadap probabilitas, yang hampir tidak dapat untuk dianalisa satu demi satu data. Akibatnya, menyebabkan nilai probabilitas tersebut menjadi sukar untuk dapat dilakukan. Oleh karena itu, dapat berlaku persamaan berikut:

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

Dimana $i \neq j$, sehingga di dapatkan:

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

Persamaan tersebut adalah model yang akan diterapkan dalam pemrosesan pengklasifikasian. Untuk klasifikasi dengan data yang berlanjut kontinyu dapat menggunakan rumus *Densitas Gauss* :

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

Keterangan:

- P : Peluang
- Xi : Atribut terhadap i
- xi : Nilai atribut berdasarkan i
- Y : Kelas yang akan dicari
- yi : Sub kelas Y yang akan dicari
- μ : mean, mendefenisikan rata – rata terhadap seluruh atribut
- σ : Deviasi standar, menyatakan varian terhadap seluruh atribut.

2.3 Metode C4.5

Algoritma C4.5 merupakan salah satu algoritma yang menerapkan sistem klasifikasi dengan menyerupai pohon keputusan yang memiliki nilai internal node-nya. Setiap node daun (terminal node) pada pohon keputusan menentukan label terhadap class. Node akar adalah node yang paling atas dari sebuah pohon keputusan [14]

Algoritma C4.5 menggunakan konsep dari *information gain* serta *entropy reduction* untuk memilih pembagian yang optimal. Berikut ini tahapan-tahapan yang dilakukan dalam membuat pohon keputusan sebagai berikut[14], [15] :



1. Memperoleh data training yang dikutip melalui data yang pernah ada sebelumnya dan akan dilakukan pengelompokkan dalam beberapa kelas tertentu.
2. Menentukan akar pohon yang nantinya akan dihitung dari nilai gain yang terbesar dari masing – masing atribut atau melalui nilai index *entropy* terkecil yang akan dilakukan perhitungan dengan nilai rumus persamaan berikut ini :

$$Entropy(S) = \sum_{i=0}^n -pi * \log_2 pi \tag{8}$$

Pada persamaan 8 merupakan perhitungan nilai dari *entropy* yang digunakan untuk menentukan seberapa informative dari atribut tersebut[16].

Berikut keterangannya:

- S : Himpunan dari banyaknya kasus
- n : Jumlah dari banyaknya partisi S
- pi : Jumlah kasus dari partisi ke-i

$$Gain(S, A) = Entropy (S) - \sum_{i=1}^n \frac{|Si|}{|S|} * Entropy(Si) \tag{9}$$

Persamaan 9 merupakan rumus yang digunakan dalam perhitungan *gain* setelah melakukan perhitungan nilai *gain* setelah melakukan perhitungan nilai *entropy*[16].

Keterangan :

- S : Himpunan dari banyaknya Kasus
- n : Jumlah dari banyaknya partisi A
- /Si/ : Jumlah kasus dari partisi ke-i
- /S/ : Jumlah dari kasus S

3. HASIL DAN PEMBAHASAN

3.1 Implementasi Metode Naive Bayes

Pengimplementasian metode *Naive Bayes* memberikan asumsi terhadap nilai atribut secara kondisional dimana akan saling terbuka apabila diberikan nilai output. Dimana artinya, diberikan nilai output dan nilai probabilitas yang akan mengamati secara bersama adalah produk dari probabilitas masing-masing data tersebut. Kelebihan menggunakan *Naive Bayes* adalah dapat dipakai untuk data kuantitatif maupun kualitatif, tidak perlu melakukan data training yang banyak dalam arti metode ini hanya membutuhkan *training* data yang sedikit untuk dapat mengamati *rules* dan estimasi parameter yang diperlukan dalam proses pengklasifikasian[17]. Penerapan dari metode NBC adalah:

3.1.1 Training data

Dalam melakukan klasifikasi terlebih dahulu harus mengetahui data apa yang akan digunakan untuk dijadikan data latih dan

Tabel 1. Data Latih

No	IPK	Usia	Jenis Kela min	Kesehatan	Pengalaman Mengajar	Nilai Tes Tertulis	Wawancara	HomeBase	Approve
1	Baik Sekali	Middle	L	Baik	Ada	Sangat Bagus	Bagus	Tidak Ada	Yes
2	Baik Sekali	Middle	L	Baik	Tidak Ada	Sangat Bagus	Sangat Bagus	Tidak Ada	Yes
3	Baik	Middle	L	Baik	Ada	Bagus	Bagus	Ada	No
4	Memuaskan	Muda	L	Baik	Ada	Sangat Bagus	Sangat Bagus	Ada	No
5	Baik	Muda	L	Baik	Ada	Kurang	Bagus	Tidak Ada	No
6	Baik	Tua	P	Baik	Ada	Bagus	Bagus	Tidak Ada	Yes
7	Baik Sekali	Middle	L	Baik	Tidak Ada	Sangat Bagus	Sangat Bagus	Tidak Ada	Yes
8	Baik	Muda	P	Baik	Tidak Ada	Bagus	Bagus	Tidak Ada	Yes
9	Baik	Muda	P	Baik	Ada	Kurang	Sangat Bagus	Tidak Ada	No
10	Memuaskan	Muda	L	Baik	Tidak Ada	Bagus	Sangat Bagus	Tidak Ada	Yes
11	Baik	Tua	L	Baik	Ada	Sangat Bagus	Sangat Bagus	Ada	No
12	Baik	Middle	L	Baik	Tidak Ada	Kurang	Kurang	Tidak Ada	No
13	Baik	Middle	L	Baik	Ada	Bagus	Sangat Bagus	Tidak Ada	Yes
14	Baik	Middle	L	Baik	Ada	Bagus	Sangat Bagus	Ada	No
15	Baik Sekali	Muda	L	Baik	Tidak Ada	Bagus	Sangat Bagus	Tidak Ada	Yes
16	Baik Sekali	Muda	L	Baik	Tidak Ada	Bagus	Sangat Bagus	Tidak Ada	Yes
17	Baik	Muda	L	Baik	Ada	Kurang	Bagus	Tidak Ada	Yes
18	Baik	Tua	P	Baik	Ada	Bagus	Bagus	Tidak Ada	Yes
19	Memuaskan	Muda	L	Baik	Ada	Sangat Bagus	Sangat Bagus	Ada	No
20	Memuaskan	Muda	L	Baik	Tidak Ada	Bagus	Sangat Bagus	Tidak Ada	Yes



No	IPK	Usia	Jenis Kelamin	Kesehatan	Pengalaman Mengajar	Nilai Tes Tertulis	Wawancara	HomeBase	Approve
21	Baik	Tua	L	Baik	Ada	Sangat Bagus	Sangat Bagus	Ada	No
22	Memuaskan	Muda	L	Baik	Tidak Ada	Bagus	Sangat Bagus	Tidak Ada	Yes
23	Baik	Tua	L	Baik	Ada	Sangat Bagus	Sangat Bagus	Ada	No
24	Baik	Middle	L	Baik	Tidak Ada	Kurang	Kurang	Tidak Ada	No
25	Memuaskan	Muda	L	Baik	Tidak Ada	Bagus	Sangat Bagus	Tidak Ada	Yes
26	Memuaskan	Muda	L	Baik	Tidak Ada	Bagus	Sangat Bagus	Tidak Ada	Yes
27	Memuaskan	Muda	L	Baik	Tidak Ada	Bagus	Sangat Bagus	Tidak Ada	Yes
28	Memuaskan	Muda	L	Baik	Tidak Ada	Bagus	Sangat Bagus	Tidak Ada	Yes
29	Baik	Middle	L	Baik	Tidak Ada	Kurang	Kurang	Tidak Ada	No
30	Memuaskan	Muda	L	Baik	Tidak Ada	Bagus	Sangat Bagus	Tidak Ada	Yes
31	Baik	Tua	L	Baik	Ada	Sangat Bagus	Sangat Bagus	Ada	No
32	Baik	Middle	L	Baik	Tidak Ada	Kurang	Kurang	Tidak Ada	No
33	Memuaskan	Muda	L	Baik	Tidak Ada	Bagus	Sangat Bagus	Tidak Ada	Yes
34	Memuaskan	Muda	L	Baik	Tidak Ada	Bagus	Sangat Bagus	Tidak Ada	Yes
35	Memuaskan	Muda	L	Baik	Tidak Ada	Bagus	Sangat Bagus	Tidak Ada	Yes
36	Memuaskan	Muda	L	Baik	Tidak Ada	Bagus	Sangat Bagus	Tidak Ada	Yes
37	Baik	Middle	L	Baik	Tidak Ada	Kurang	Kurang	Tidak Ada	No

3.1.2 Kriteria dan Perhitungan Probabilitas

Adapun nilai *probabilities* dari masing-masing kriteria yang dihitung berdasarkan dari data latih pada tabel 1, dengan menerapkan rumus persamaan 1 dan 2 adalah sebagai berikut:

3.1.2.1 Probabilitas Kriteria IPK

Berdasarkan data latih pelamar pada tabel 1 diketahui data berjumlah 37 data pelamar dosen, dimana dari 37 pelamar terdapat 2 pelamar ditolak dengan nilai IPK memuaskan, 13 pelamar ditolak dengan nilai IPK baik, 12 pelamar diterima dengan nilai IPK memuaskan, 10 pelamar diterima dengan nilai IPK baik dan baik sekali. Probabilitas kriteria terhadap IPK dapat dilihat pada tabel 2.

Tabel 2. Probabilitas IPK

IPK	Approve		Probabilitas	
	Diterima	Ditolak	Diterima	Ditolak
Memuaskan	12	2	0,545	0,133
Baik Sekali	5	0	0,227	0,000
Baik	5	13	0,227	0,867
Jumlah	22	15	1,000	1,000

3.1.2.2 Probabilitas Kriteria Usia

Pada kriteria usia dapat diketahui dari 37 data terdapat 4 pelamar ditolak dengan usia tua, 7 ditolak dengan usia middle, dan 4 ditolak dengan usia muda. Selain itu juga terdapat 2 data pelamar diterima dengan usia tua, 4 data pelamar diterima dengan usia middle, dan 16 data pelamar diterima dengan usia muda. Nilai Probabilitas kriteria terhadap usia lebih lengkapnya dapat dilihat pada tabel 3.

Tabel 3. Probabilitas Usia

Usia	Approve		Probabilitas	
	Diterima	Ditolak	Diterima	Ditolak
Tua	2	4	0,091	0,267
Middle	4	7	0,182	0,467
Muda	16	4	0,727	0,267
Jumlah	22	15	1,000	1,000

3.1.2.3 Probabilitas Kriteria Jenis Kelamin

Diketahui dari 37 data pelamar, terdapat 14 data pelamar yang ditolak dengan jenis kelamin laki-laki sedangkan 1 data pelamar ditolak dengan jenis kelamin perempuan. Namun juga terdapat 19 data pelamar yang diterima dengan jenis kelamin laki-laki dan 3 data pelamar diterima dengan jenis kelamin perempuan. Probabilitas kriteria jenis kelamin dapat dilihat pada tabel 4.

Tabel 4. Probabilitas Jenis Kelamin

Jenis Kelamin	Approve		Probabilitas	
	Diterima	Ditolak	Diterima	Ditolak
Laki-laki	19	14	0,864	0,933



Jenis Kelamin	Approve		Probabilitas	
	Diterima	Ditolak	Diterima	Ditolak
Perempuan	3	1	0,136	0,067
Jumlah	22	15	1,000	1,000

3.1.2.4 Probabilitas Kriteria Pengalaman Mengajar

Berikutnya pada kriteria pengalaman mengajar tampak bahwa yang tidak memiliki pengalaman mengajar lebih banyak diterima dibandingkan yang ditolak. 5 data pelamar diterima dengan memiliki pengalaman mengajar dan 17 pelamar diterima dengan tidak memiliki pengalaman mengajar. Juga terdapat 10 data pelamar yang ditolak dengan memiliki pengalaman mengajar dan 5 pelamar ditolak yang tidak memiliki pengalaman mengajar. Probabilitas pengalaman mengajar dapat dilihat pada tabel 5.

Tabel 5. Probabilitas Pengalaman Mengajar

Pengalaman Mengajar	Approve		Probabilitas	
	Diterima	Ditolak	Diterima	Ditolak
Ada	5	10	0,227	0,667
Tidak Ada	17	5	0,773	0,333
Jumlah	22	15	1,000	1,000

3.1.2.5 Probabilitas Kriteria Nilai Tes Tertulis

Tampak pada kriteria nilai tes tertulis bahwa, nilai tes tertulis yang bagus lebih besar peluang untuk diterima. Dari 37 data pelamar terdapat 3 data pelamar yang diterima dengan nilai tes tertulis sangat bagus, 18 pelamar diterima dengan nilai tes tertulis bagus, dan 1 pelamar diterima dengan nilai tes tertulis kurang. 6 pelamar ditolak dengan nilai tes tertulis sangat bagus, 2 ditolak dengan nilai tes tertulis bagus, dan 7 ditolak dengan nilai tes tertulis kurang. Probabilitas nilai tes tertulis dapat dilihat pada tabel 6.

Tabel 6. Probabilitas Nilai Tes Tertulis

Nilai Tes Tertulis	Approve		Probabilitas	
	Diterima	Ditolak	Diterima	Ditolak
Sangat Bagus	3	6	0,136	0,400
Bagus	18	2	0,818	0,133
Kurang	1	7	0,045	0,467
Jumlah	22	15	1,000	1,000

3.1.2.6 Probabilitas Kriteria Wawancara

Pada Kriteria wawancara, terdapat 17 data pelamar yang diterima dengan wawancara sangat bagus, 5 pelamar diterima dengan wawancara bagus, 8 ditolak dengan wawancara sangat bagus, 2 ditolak dengan wawancara bagus dan 5 ditolak dengan wawancara kurang. Probabilitas Wawancara dapat dilihat pada tabel 7.

Tabel 7. Probabilitas Wawancara

Wawancara	Approve		Probabilitas	
	Diterima	Ditolak	Diterima	Ditolak
Sangat Bagus	17	8	0,773	0,533
Bagus	5	2	0,227	0,133
Kurang	0	5	0,000	0,333
Jumlah	22	15	1,000	1,000

3.1.2.7 Probabilitas Kriteria Homebase

Pada kriteria homebase lebih dominan pelamar yang diterima adalah calon dosen yang belum memiliki homebase. Tampak pada tabel 8 bahwa 22 data pelamar diterima untuk yang belum memiliki homebase, 8 data ditolak dengan calon dosen yang memiliki homebase dan 7 data ditolak dengan calon dosen yang belum memiliki homebase.

Tabel 8. Probabilitas Homebase

Homebase	Approve		Probabilitas	
	Diterima	Ditolak	Diterima	Ditolak
Ada	0	8	0,000	0,533
Tidak Ada	22	7	1,000	0,467
Jumlah	22	15	1,000	1,000



3.2 Perhitungan Metode Naïve Bayes

Perhitungan manual dilakukan dengan bantuan *Microsoft Excel* guna mempercepat untuk mendapatkan hasil perhitungannya. Dari penelitian ini diketahui jumlah data percobaan untuk dilakukannya perhitungan peluang dan probabilitas sebanyak 37 data. Masing-masing pengklasifikasian data dihitung sedemikian rupa untuk mengetahui tingkat keakurasian dari output yang dihasilkan. Proses perhitungan dengan menggunakan rumus persamaan 4 hingga persamaan 6. Untuk lebih jelasnya lihat pada tabel 9.

Tabel 9. Perhitungan Metode Naïve Bayes

Kriteria	X C1	Approve = Yes	Approve = No	X C2	Approve = Yes	Approve = No
IPK	= Baik Sekali	0,227	0,000	= Baik Sekali	0,227	0,000
Usia	= Middle	0,182	0,467	= Middle	0,182	0,467
Jenis Kelamin	= L	0,864	0,933	= L	0,864	0,933
Kesehatan	= Baik	1	1	= Baik	1	1
Peng. Mengajar	= Ada	0,227	0,667	= Tidak Ada	0,773	0,333
Nilai Tes Tulis	= Sangat Bagus	0,136	0,400	= Sangat Bagus	0,136	0,400
Wawancara	= Bagus	0,773	0,533	= Sangat Bagus	0,773	0,533
Homebase	= Tidak Ada	1	0,467	= Tidak Ada	1	0,467
P(X Ci) : P(X Approve)		0,000854	0		0,002905	0
P(X Ci)*P(Ci) = P(X Approve) * P(Approve)		0,000508	0,000		0,001727	0,000

Dimana hasil dari **P(X|Ci) : P(X|Approve)** merupakan perkalian keseluruhan nilai probabilitas kriteria terhadap masing-masing data.

$$P(X|Ci) : P(X|Approve = Yes) = 0,227 \times 0,182 \times 0,864 \times 1 \times 0,227 \times 0,136 \times 0,773 \times 1 = 0,000854649$$

$$P(X|Ci) : P(X|Approve = No) = 0,000 \times 0,467 \times 0,933 \times 1 \times 0,667 \times 0,400 \times 0,533 \times 0,467 = 0$$

Berikutnya hasil dikalikan kembali dengan nilai **P(Approve)** utama sehingga menjadi,

$$P(X|Ci)*P(Ci) = P(X|Approve=Yes) * P(Approve =Yes) = 0,000854 \times 0,595 = 0,000508$$

$$P(X|Ci)*P(Ci) = P(X|Approve=No) * P(Approve =No) = 0 \times 0,405 = 0$$

Hasil diatas menunjukkan bahwa Nilai **P(X|Ci)*P(Ci) = P(X|Approve=Yes) * P(Approve =Yes)** lebih besar dari pada **P(X|Ci)*P(Ci) = P(X|Approve=No) * P(Approve =No)**, artinya output pada data adalah diterima/Yes.

3.3 Perhitungan Metode C4.5

Pada perhitungan metode C4.5 langkah awal dalam perhitungan adalah mencari nilai Entrophy dan nilai Gain. Berikut Perhitungan dari Metode C4.5:

- Mengetahui nilai entrophy dengan menghitung jumlah yang Diterima dan Ditolak melalui keseluruhan data latih dengan menerapkan rumus persamaan 8. Dari jumlah data sebanyak 37 Pelamar diketahui jumlah yang diterima sebanyak 22 pelamar dan 15 Pelamar ditolak. Sehingga dapat dihitung nilai entrophy nya:

$$Entropy(S) = - \sum_{i=1}^n f(i, j) . \log_2 f [(i, j)] = (-22/37 * \log_2(22/37)) + (-15/37 * \log_2(15/37)) = 0,974$$

- Berikutnya menghitung nilai entrophy dan nilai gain berdasarkan masing-masing kriteria yang terdiri dari IPK, usia, jenis kelamin, kesehatan, pengalaman mengajar, nilai tes tertulis, wawancara, dan homebase. Berdasarkan nilai IPK dari 37 data pelamar terdapat 22 data pelamar diterima dengan 12 data pelamar dengan nilai IPK memuaskan, 5 dengan nilai baik sekali, dan 4 dengan nilai baik. Serta terdapat 15 data pelamar ditolak dengan 2 data pelamar memiliki nilai memuaskan, dan 13 dengan nilai baik, sehingga dapat dilihat pada tabel 10.

Tabel 10. Perhitungan nilai Gain

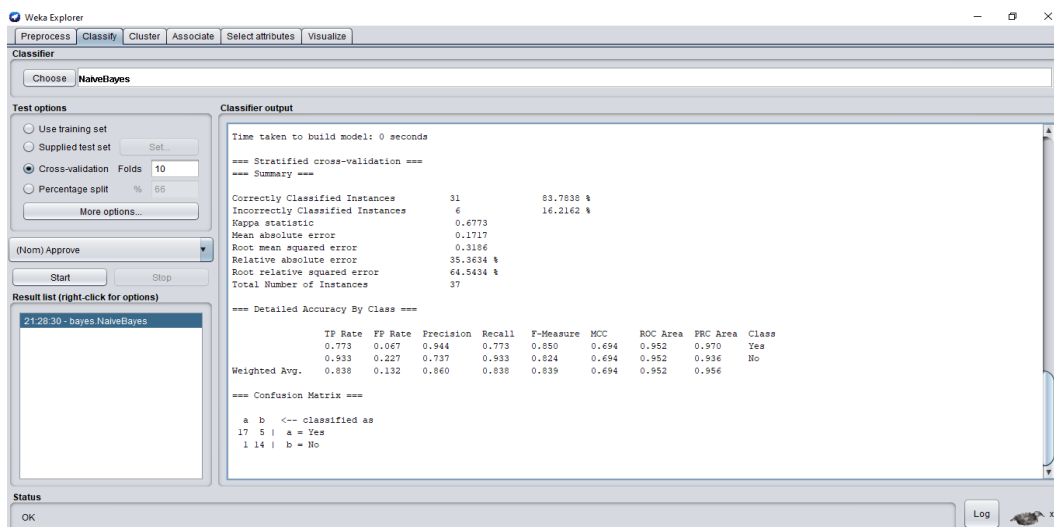
Nilai IPK		log(i)	log(j)	Nilai	Execute Nilai	
Memuaskan	diterima	12	0,1906	0,4011	0,592	0,224
	ditolak	2				
JumlahData		14				
Baik Sekali	diterima	5	0,0000	0,0000	0,000	0,000
	ditolak	0				



Nilai IPK		log(i)	log(j)	Nilai	Execute Nilai
Baik	JumlahData	5			
	diterima	5	0,5133	0,3391	0,852
	ditolak	13			
	JumlahData	18			
Jumlah Nilai Keseluruhan					0,639
Nilai Entropy					0,974
Hasil Nilai Gain					0,335

3.4 Pengujian dengan Metode Naïve Bayes

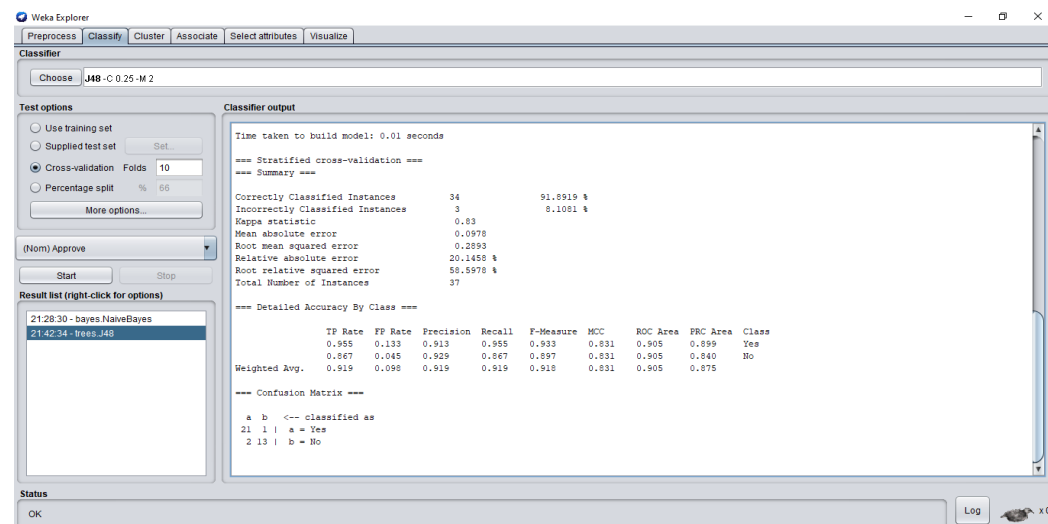
Dari masing-masing nilai perhitungan probabilitas di atas akan dilakukan uji data sebanyak 37 data dan direpresentasikan dengan menggunakan tool weka sehingga dapat menghasilkan klasifikasi penerimaan dosen tetap yang memiliki tingkat akurasi (Correctly Classified Instances) sebesar 83,7838 % sedangkan persentase untuk Incorrectly Classified Instances 16,2162 %. Artinya dari 37 data pelamar terdapat 31 data yang berhasil diklasifikasi dengan benar dan 6 data pelamar yang tidak berhasil diklasifikasikan dengan benar, dapat dilihat pada gambar 2.



Gambar 2. Hasil Klasifikasi Metode Naïve Bayes

3.5. Pengujian dengan Metode C4.5

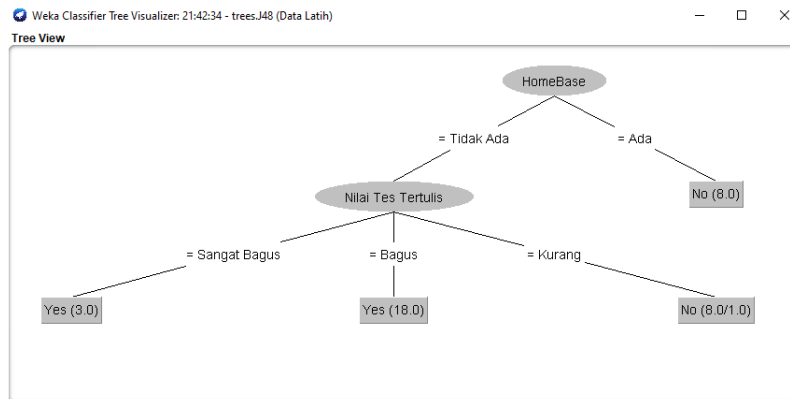
Hasil dari probabilitas diatas dengan jumlah 37 data pelamar akan diuji dengan menggunakan tools weka dengan metode klasifikasi C4.5 sehingga menghasilkan klasifikasi penerimaan dosen tetap dengan tingkat akurasi sebesar 91,8919 % dan *Incorrectly Classified Instances* sebesar 8,1081%. Dimana dari 37 data pelamar yang berhasil diidentifikasi dengan baik yaitu berjumlah 34 data pelamar, sedangkan 3 data pelamar tidak berhasil diklasifikasi. Hasil dari pengklasifikasian dengan tools weka dapat dilihat pada Gambar 3.



Gambar 3. Hasil Klasifikasi Metode C4.5



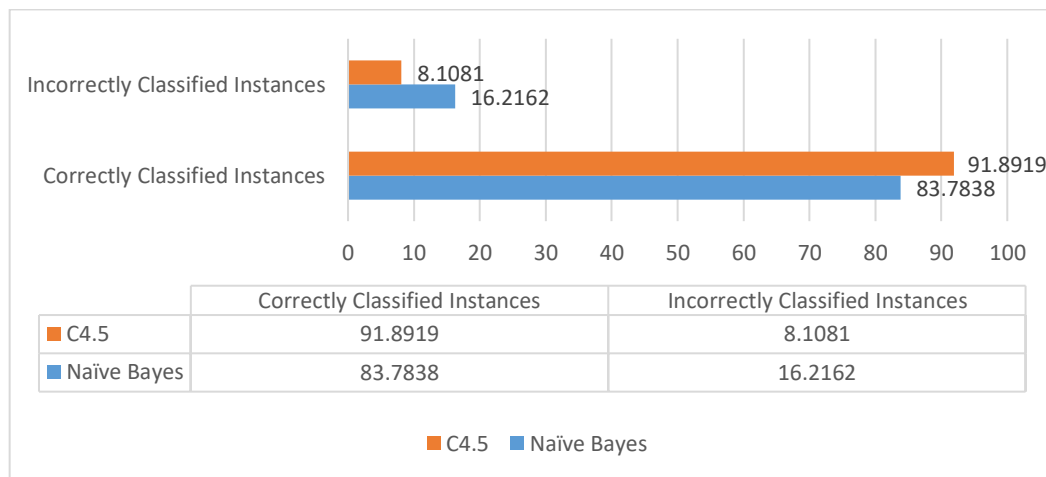
Berdasarkan dari hasil yang telah didapat melalui tools weka maka dapat direpresentasikan dengan sebuah pohon keputusan yang menggambarkan terhadap klasifikasi dengan metode C4.5. Representasi pohon keputusan dapat dilihat pada Gambar 3.



Gambar 4. Representasi Hasil Klasifikasi Metode C4.5

3.6 Hasil Perbandingan Kedua Metode

Berdasarkan hasil implementasi dari kedua metode yang diperoleh dengan menggunakan tools weka 3.8 maka hasil perbandingan akurasi dari kedua metode yang didapat sebagai berikut:



Gambar 4. Hasil Perbandingan Metode

Dari grafik diatas dapat ditarik kesimpulan bahwa metode C4.5 memiliki nilai persentase yang lebih baik dibandingkan metode naive bayes. Artinya metode C4.5 lebih akurat dan lebih efisien dengan adanya pohon keputusan. Perlu diketahui tingkat akurasi akan jauh lebih baik lagi apabila memiliki lebih banyak data.

4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan tentang perbandingan tingkat akurasi dalam Penerimaan Dosen Tetap pada Universitas Potensi Utama dengan menerapkan metode *Naive Bayes Classification* dan C4.5 dapat ditarik kesimpulan bahwa berdasarkan 9 kriteria yang diuji dalam penerimaan dosen tetap yaitu nilai IPK, usia, jenis kelamin, kesehatan, pengalaman mengajar, nilai tes tertulis, wawancara, dan homebase, kriteria homebase merupakan kriteria yang sangat dominan dalam menentukan apakah dosen tersebut dapat diterima atau tidak. Hasil percobaan yang telah dilakukan terhadap data penerimaan dosen tetap yang digunakan sebagai data latih, metode *naive bayes* memiliki tingkat akurasi sebesar 83,7838 % sedangkan C4.5 sebesar 91,891 %. Sehingga dapat diperoleh kesimpulan dalam penelitian ini bahwa metode C4.5 merupakan metode yang memiliki tingkat akurasi yang baik dan efisien dari pada naive bayes.

REFERENCES

[1] E. H. M. E. H. A. Sakinah Mat Zin, N. N. Jaafar, Rosfatihah Che Mat, W. Nurfahizul Ifwah W. Alias, "E-recruitment technology: the effective source of recruitment.," vol. vol 6, p. 84–89., 2016.
 [2] Marwansyah, *Manajemen sumber daya manusia*. KOTA MAKASSAR: Alfabeta, 2016.
 [3] 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.
- [4] E. Buulolo, *Data Mining Untuk Perguruan Tinggi*. Deepublish, 2020.
- [5] Y. Mardi, “Data Mining : Klasifikasi Menggunakan Algoritma C4.5,” *J. Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2017.
- [6] D. Srianto and E. Mulyanto, “Perbandingan K-Nearest Neighbor Dan Naive Bayes,” *Techno.COM*, vol. 15, no. 3, pp. 241–245, 2016.
- [7] S. Yakub, A. Fitri Boy, I. Mariami, W. Stmik, and T. Dharma, “J-SISKO TECH Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD Penerapan Data Mining Pengaturan Pola Tata Letak Barang Pada Berkah Swalayan Untuk Strategi Penjualan Menggunakan Algoritma Apriori,” *□*, vol. 69, no. 1, pp. 69–75, 2019.
- [8] E. Etriyanti, D. Syamsuar, and N. Kunang, “Implementasi Data Mining Menggunakan Algoritme Naive Bayes Classifier dan C4.5 untuk Memprediksi Kelulusan Mahasiswa,” *Telematika*, vol. 13, no. 1, pp. 56–67, 2020.
- [9] Y. A. Gerhana, I. Fallah, W. B. Zulfikar, D. S. Maylawati, and M. A. Ramdhani, “Comparison of naive Bayes classifier and C4.5 algorithms in predicting student study period,” *J. Phys. Conf. Ser.*, vol. 1280, no. 2, 2019.
- [10] E. N. Azizah, U. Pujiyanto, E. Nugraha, and Darusalam, “Comparative performance between C4.5 and Naive Bayes classifiers in predicting student academic performance in a Virtual Learning Environment,” *2018 4th Int. Conf. Educ. Technol. ICET 2018*, no. 1, pp. 18–22, 2018.
- [11] yogiek indra Kurniawan, “PERBANDINGAN ALGORITMA NAIVE BAYES DAN C.45 DALAM KLASIFIKASI DATA MINING,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 4, pp. 455–464, 2018.
- [12] T. Arifin, “Metode Data Mining Untuk Klasifikasi Data Sel Nukleus Dan Sel Radang Berdasarkan Analisa Tekstur,” *Informatika*, vol. II, no. 2, pp. 425–433, 2015.
- [13] A. Saleh, “Implementasi Metode Klasifikasi Naive Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga,” *Creat. Inf. Technol. J.*, vol. 2, no. 3, pp. 207–217, 2015.
- [14] W. D. Septiani, “Komparasi Metode Klasifikasi Data Mining Algoritma C4.5 Dan Naive Bayes Untuk Prediksi Penyakit Hepatitis,” *J. Pilar Nusa Mandiri*, vol. 13, no. 1, pp. 76–84, 2017.
- [15] E. Buulolo, N. Silalahi, Fadlina, and R. Rahim, “C4.5 Algorithm To Predict the Impact of the Earthquake,” *Int. J. Eng. Res. Technol.*, vol. 6, no. 2, 2017.
- [16] 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.
- [17] A. Saleh, “Penerapan Data Mining Dengan Metode Klasifikasi Naive Bayes Untuk Memprediksi Kelulusan Mahasiswa Dalam Mengikuti English Proficiency Test (Studi Kasus : Universitas Potensi Utama),” *Konf. Nas. Sist. Informasi, Univ. Klabat, Manado, Indonesia, Vol. 2015*, no. February 2015, pp. 1–6, 2015.