

# Komparasi Model Klasifikasi *Naïve Bayes* dan *C4.5* pada Data Prestasi Kerja Pegawai Negeri Sipil

Yola Vegita<sup>1</sup>, Cahyo Prianto<sup>\*2</sup>, Syafrial Fachri Pane<sup>3</sup>

<sup>1,2,3</sup>Jurusan Teknik Informatika, Sekolah Vokasi, Universitas Logistik Bisnis Internasional  
Sarijadi, Jl. Sariasih No. 54, Bandung

E-mail: <sup>1</sup>vegityola@gmail.com, <sup>\*2</sup>cahyo@ulbi.ac.id, <sup>3</sup>syafrial.fachri@ulbi.ac.id

**Abstract**—Human resources or HR is important parts to raise success in an organization. Employees who do not complete the work, the organization's targets will not be achieved. So, it is very, if an employee is not optimal and cannot be explained properly, it will affect the development and progress of the company or organization. To evaluate Civil performance Servants of the Transportation Service of West Java Province by utilizing work performance appraisal results, in which the data used is the 2020 assessment. The number of employees makes work performance difficult and neither evaluated nor objective. To carry out a work assessment, you can use supporting methods, one of which is by classifying employee data with data mining. This study uses the performance algorithm *Naïve Bayes* and *C4.5* by comparing the results of modeling using the Confusion Matrix and Classification Report. As a result, *C4.5* has an accuracy of 99.12% while *Naïve Bayes* is only 83%.

**Abstrak**—Salah satu bagian yang terpenting untuk mencapai keberhasilan dalam kemajuan suatu organisasi adalah SDM atau sumber daya manusia. Pegawai yang tidak menuntaskan pekerjaannya, maka target organisasi tidak akan tercapai. Berdasarkan hal tersebut, apabila seorang pegawai tidak maksimal dan tidak dapat bekerja dengan baik, pastinya akan mempengaruhi perkembangan dan kemajuan dari perusahaan atau organisasi. Untuk melakukan evaluasi kinerja PNS Dinas Perhubungan Provinsi Jawa Barat dengan memanfaatkan hasil penilaian prestasi kerja, yang mana data yang digunakan adalah penilaian pada tahun 2020. Banyaknya pegawai membuat penilaian Prestasi Kerja menjadi sulit dan tidak dipungkiri penilaian juga terkadang dilakukan tidak objektif. Untuk melakukan suatu penilaian kerja dapat menggunakan metode pendukung, salah satunya dengan melakukan klasifikasi data pegawai dengan data mining. Penelitian ini membandingkan algoritma performance algoritma *Naïve Bayes* dan *C4.5* dengan mengevaluasi hasil pemodelan dengan Confusion Matrix dan Classification Report. Hasilnya, *C4.5* memiliki akurasi 99.12% sedangkan *Naïve Bayes* hanya 83%.

**Kata Kunci**—*C4.5*, Classification Report, Confusion Matrix, Evaluasi Model, *Naïve Bayes*, Perbandingan Model

## I. PENDAHULUAN

SDM merupakan elemen yang penting dalam proses menjalankan bisnis organisasi. Sumber daya manusia adalah tenaga yang penting untuk mencapai keberhasilan, target untuk mencapai kemajuan suatu organisasi [1]. Sudah sangat jelas, apabila seorang pegawai tidak maksimal dan tidak dapat bekerja dengan baik, pastinya akan mempengaruhi perkembangan dan kemajuan dari perusahaan atau organisasi [2]

PNS atau Pegawai Negeri Sipil adalah seorang pegawai yang bekerja untuk pemerintah. Setiap pegawai memiliki sasaran kinerja yang harus direalisasikan sebelum akhir tahun, yang mana target sasaran tersebut telah ditetapkan pada awal tahun pada awal tahun, yang kemudian realisasi ini akan menjadi penilain prestasi kerja PNS.

Seorang PNS juga akan dinilai berdasarkan nilai perilaku kerja yang merujuk ke profesionalan sikap dan tindakan. Semua penilaian tersebut akan menjadi Penilaian Prestasi kerja PNS yang di proses dan dinilai oleh Pejabat Penilai. Penjelasan tersebut berdasarkan Perka BKN No 35 Tahun 2011 tentang Pedoman Penyusunan Karir PNS [3].

Banyaknya pegawai membuat penilaian Prestasi Kerja menjadi sulit dan tidak dipungkiri penilaian juga terkadang dilakukan tidak objektif. Untuk melakukan suatu penilain kerja dapat menggunakan metode pendukung, salah satunya dengan melakukan klasifikasi data pegawai dengan data mining.

Pada penelitian ini, akan membandingkan metode *Naïve Bayes* dan metode *C4.5* dalam melakukan klasifikasi data prestasi kerja PNS. Hal ini dilakukan untuk mengetahui klasifikasi dengan tingkat akurasi yang lebih tinggi

Dengan menggunakan data Prestasi Kerja PNS di salah satu Instansi Pemerintah di Indonesia yang berjumlah 155 data, dan memiliki 6 atribut yaitu Nama, NIK, nilai SKP, nilai Perilaku dan nilai Akhir. Data tersebut akan diklasifikasikan dengan 5 label *class* yaitu A (Sangat Baik), B (Baik), C (Cukup), D (Kurang) dan E (Sangat Kurang).

Beberapa penelitian sebelumnya yang membahas mengenai penilain kerja pegawai ataupun karyawan sebagai berikut ; Penelitian yang dilakukan oleh Putu Wirya Kastawan dkk dengan 184 data pegawai menggunakan metode *C5.0* dan mendapatkan akurasi yang tinggi sampai 96.08% [4] Kemudian penelitian yang dilakukan oleh

Darmanto dengan menggunakan metode C4.5 dan mendapatkan akurasi sebesar 91.21% [5]

II. METODE PENELITIAN



Gambar 1. Alur Metodologi Penelitian

Gambar 1 adalah alur dari penelitian ini, dan berikut penjelasan setiap alur yang dilaksanakan.

A. Studi Literatur

Tahapan studi *literature* dilakukan untuk memperkuat landasan penelitian. Dengan mempelajari penelitian terdahulu dengan pembahasan yang serupa dapat meningkatkan kualitas dari suatu penelitian

B. Pengumpulan Data

Teknik pengumpulan data yang digunakan pada penelitian ini adalah dengan cara observasi, studi dokumen dan wawancara.

C. Data Preprocessing

*Preprocessing* adalah tahap pra-pemrosesan data untuk dapat mengkoreksi *scattering*, kebisingan, nilai-nilai yang hilang dan lainnya sehingga struktur data menjadi lebih relevan [6]

D. Encode Data

Melakukan pengkodean data *categorical* ke data *numerical*. Hal ini dikarenakan algoritma pembelajaran mesin dan hampir semua arsitektur *deep learning* tidak mampu memproses teks biasa dalam bentuk mentahnya [7]

E. Mengecek Missing Value

*Missing Value* adalah kumpulan data dengan nilai yang hilang [8]

F. Melihat Korelasi antar Atribut

Nilai korelasi yang baik ditentukan apabila suatu nilai korelasi mencapai +1. Nilai tersebut menandakan hubungan searah antara atribut x dan atribut y, dimana bila atribut x naik maka atribut y juga akan naik (Widayanti, 2014).

G. Mengecek Outlier

Data *Outlier* adalah data yang memiliki karakteristik unik. Data *Outlier* atau data pencilan ini sangat berbeda jauh dengan data lainnya yang memiliki bentuk nilai ekstrim baik sebagai *variable* tunggal maupun kombinasi (Ghozali, 2011 : 41).

H. Mengatasi Imbalanced Data

Pada proses ini ketidak-seimbangan data diatasi dengan *Randomoversampling* yang mengubah total data setiap atribut kelas menjadi sama dengan jumlah atribut kelas mayoritas.

I. Imbalance data

Klasifikasi data yang tidak seimbang dapat menyebabkan masalah sehingga data memiliki nilai yang tidak proporsional. Terdapat kelas yang jumlahnya mendominasi dari kelas lain, kasus ini disebut data mayoritas, sedangkan data yang lebih kecil adalah data minoritas. [10]

J. Pemodelan

Pada penelitian ini akan dilakukan pemodelan klasifikasi menggunakan metode *Naïve Bayes* dan C4.5

K. Formulir Copyright

Artikel yang siap untuk dipublikasikan harus dilengkapi dengan sebuah formulir *copyright*. Setiap penulis bertanggung jawab secara pribadi untuk mendapatkan *security clearances* apapun yang dibutuhkan.

L. Naïve Bayes

*Naïve Bayes* merupakan metode klasifikasi *statistic*. Metode ini dapat melakukan prediksi atau peluang keanggotaan kelas sebuah data [11]. Berdasarkan *history* penilaian yang telah memiliki lable *class Naïve Bayes* dapat memprediksi dengan menghitung probabilitas data. Metode ini memanfaatkan perhitungan probabilitas dan statistika yang dapat memprediksi peluang di masa depan berdasarkan *history* data [12] Algoritma *Naïve Bayes* memiliki klasifikasi *probabilistic* yang memperhitungkan sekumpulan probabilitas dengan mengkombinasikan nilai dataset. (Afdhaluzzikri, 2021)

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

Sehingga:

$$P(F_i/C, F_j) = P(F_i/C)$$

Persamaan *Naïve Bayes* diatas akan digunakan untuk proses klasifikasi. Kemudian untuk Rumus *Densitas Gauss* akan digunakan sebagai klasifikasi data kontinyu. Berikut persamaan *Densitas Gauss*:

$$P = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(xi-\mu Xi)^2}{2^2\sigma}}$$

Simbol	Keterangan
P	: Peluang
Xi	: Atribut Ke-I data training
xi	: Nilai atribut Ke-I data testing
μ	: Mean
σ	: Deviasi standar : Varian dari seluruh atribut

Berikut rumus untuk menghitung nilai *Mean* :

$$\mu = \frac{x_1 + x_2 + x_3 + \dots + X_n}{n}$$

Simbol	Keterangan
$\mu$	: Mean
$x_i$	: Nilai sample Ke - i
$n$	: Jumlah keseluruhan sample (Banyaknya)

Berikut adalah rumus perhitungan standar deviasi :

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n - 1}}$$

Simbol	Keterangan
$\sigma$	: Standar Deviasi
$x_i$	: Nilai sample Ke - i
$n$	: Jumlah keseluruhan sample (Banyaknya)
$\mu$	: Mean

**M. Algoritma C4.5**

Algoritma C4.5 adalah salah satu algoritma *Decision Tree* yang memiliki kecepatan dalam pengklasifikasian pohon keputusan disampling yang dapat mengkonstruksi pengklasifikasian [14]. Untuk melakukan pembagian dan pemilihan data dengan hasil yang baik menggunakan gain atau entropy [15]

Berikut adalah tahapan dalam melakukan klasifikasi dengan algoritma C4.5

1. Mempersiapkan Data Training  
Data training yang digunakan adalah sebagai data pembelajaran metode. Data yang digunakan adalah data yang sudah ada sebelumnya dan sudah memiliki kelompok dalam kelas tertentu.
2. Menghitung entropy (Menentukan akar dari pohon)  
Untuk menentukan akar dari pohon dilakukan dengan mencari nilai gain dari setiap atribut. Atribut yang memiliki nilai paling tinggi akan menjadi akar yang pertama. Untuk menghitung gain, terlebih dahulu menghitung nilai entropy.
3. Menghitung nilai gain ratio
4. Melakukan pengulangan terhadap point 2 hingga semua record terpartisi.
5. Perulangan proses akan berhenti ketika semua record dalam simpul N memiliki kelas yang sama, yang ditandai dengan tidak adanya record yang dipartisi dan record kosong dalam cabang.

Berikut adalah persamaan untuk menghitung nilai entropy dan nilai gain

1. Entropy

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

Simbol	Keterangan
S	: Himpunan dari kasus
n	: Jumlah dari partisi s
$p_i$	: Probabilitas (jumlah kelas : total kasus)

2. Gain

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

Simbol	Keterangan
S	: Himpunan dari kasus
A	: Atribut
n	: Jumlah atribut
$ S_i $	: Jumlah dari partisi ke-i
$ S $	: Jumlah dari kasus yang ada pada S

**N. Evaluasi Model**

Evaluasi model dilakukan untuk mengetahui tingkat keberhasilan, ketepatan dan kecepatan klasifikasi menggunakan *Confusion Matrix* dan *Classification Report*

**O. Confusion Matrix**

Confusion Matrix merupakan pengukuran performa dalam pemodelan klasifikasi *machine learning*. *Confusion Matrix* merupakan table dengan 4 kombinasi berbeda dari penilaian prediksi dan penilaian actual. Istilah yang merepresentasikan hasil proses klasifikasi confusion matrix adalah *True Positif*, *True Negatif*, *False Positif* dan *False Negatif*.

- *True Positif* (TP) adalah ketika klasifikasi memprediksi data positif dan data tersebut memang positif
- *True Negative* (TN) adalah ketika klasifikasi memprediksi data negative dan data tersebut memang *negative*
- *Flase Positif* (FP) adalah ketika klasifikasi memprediksi data positif namun prediksi tersebut salah (*negative*)
- *False Negative* (FN) adalah ketika klasifikasi memprediksi data negative namun prediksi tersebut salah (positif) [16]

Evaluasi model dengan confision matrix dapat mengetahui Akurasi, Presisi dan Recall, yang mana pengertian dari ketiga hal tersebut adalah sbagai berikut:

- Akurasi : kedekatan diantara penilaian prediksi dengan penilaian actual
- Presisi : ketepatan diantara informasi yang telah diminta dengan jawaban yang diberikan system
- Recall : keberhasilan dari system dalam melakukan proses menemukan sebuah informasi.[17]
- F1-Score menekankan pada bahaya atau resiko dari false positif dan false negative

**III. HASIL DAN PEMBAHASAN**

**A. Analisis Pembahasan**

Pada penelitian kali ini, penulis akan mengklasifikasikan data prestasi kerja PNS dengan menggunakan algoritma *Naive Bayes* dan C4.5 yang banyak digunakan oleh penelitian-penelitian sebelumnya.

Data yang digunakan adalah data prestasi kerja PNS dari salah satu instansi pemerintah tahun 2020 yang berjumlah 155 data. Terdapat 6 atribut diantaranya Nama, NIK, nilai

Nama	NIK	SKP	Perilaku	Akhir	K
PNS 1	NIK	41.814	27.876	69.69	C
PNS 2	NIK	50.004	33.336	83.34	B
PNS 3	NIK	50.304	33.536	83.84	B
PNS 4	NIK	38.88	25.92	64.8	C
PNS 5	NIK	49.92	33.28	83.2	B
PNS 6	NIK	39.36	26.24	65.6	C
PNS 7	NIK	54.054	36.036	90.09	A
PNS 8	NIK	29.076	19.384	48.46	E
PNS 9	NIK	34.812	23.208	58.02	D
PNS 10	NIK	35.208	23.472	58.68	D

Berikut adalah penjelasan detail data yang digunakan

SKP, nilai Perilaku, nilai Akhir dan K atau keterangan. Contoh sample data yang digunakan dapat dilihat pada table 1 sample data.

Table 1. Sample Data

Data : Prestasi Kerja PNS (Dishub Prov Jabar)  
 Waktu : Tahun 2020  
 Jumlah : 155 data  
 Atribut : Nama, NIK, SKP, Perilaku, Akhir, dan Kategori  
 Atribut yang digunakan : Nilai SKP (Sasaran Kerja Pegawai) dan nilai Perilaku sebagai variable x K (Kategori) sebagai variable y

Type atribut dari data prestasi kerja PNS dapat dilihat seperti pada gambar 2. Perlu dilakukan encode data atau melakukan pengkodean terhadap data yang bertipe category. Atribut yang akan di encode adalah atribut K atau label dari data ini. Sehingga 0 = A, 1 = B, 2 = C, 3 = D dan 4 = E. Data prestasi kerja PNS yang akan digunakan pada penelitian ini, tidak memiliki nilai yang kosong, dan juga tidak memiliki *Outlier*.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 155 entries, 0 to 154
Data columns (total 6 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   Nama        155 non-null   object
1   NIK         155 non-null   object
2   SKP         155 non-null   float64
3   Perilaku    155 non-null   float64
4   Akhir       155 non-null   float64
5   K           155 non-null   object
dtypes: float64(3), object(3)
memory usage: 7.4+ KB
```

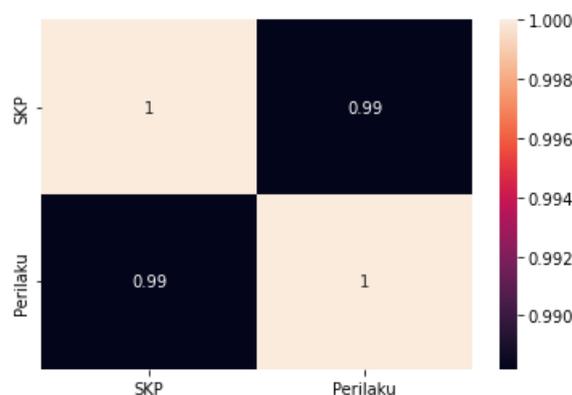
Gambar 2. Type Atribut

Atribut yang akan digunakan seperti table 2, sebagai atribut penyebab (independent) adalah atribut penilaian SKP dan penilaian Perilaku, yang mana untuk label *class* nya,

atau atribut y nya akan digunakan atribut K atau keterangan yang memiliki 5 label yaitu SB (Sangat Baik), B (Baik), C (Cukup), D (Kurang) dan E (Sangat Kurang). Berikut adalah 10 sample data yang akan digunakan dari total keseluruhan data 155 data.

Tabel 1. Atribut yang digunakan

SKP	Perilaku	K
41.814	27.876	C
50.004	33.336	B
50.304	33.536	B
38.88	25.92	C
49.92	33.28	B
39.36	26.24	C
54.054	36.036	A
29.076	19.384	E
34.812	23.208	D
35.208	23.472	D



Gambar 3. Visualisasi Korelasi SKP & Perilaku

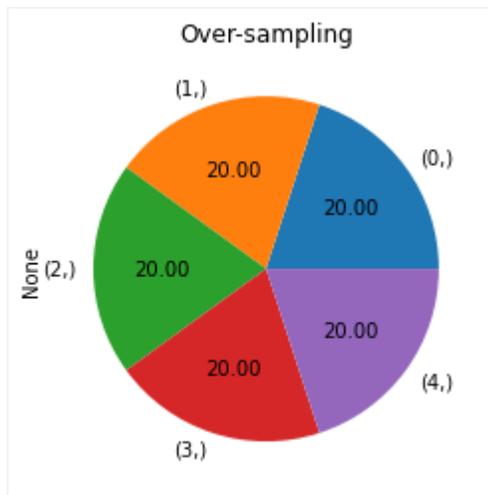
Pada gambar 3 dapat di Tarik kesimpulan, bahwa atribut SKP dan Perilaku memiliki korelasi yang bagus, nilai korelasi kedua atribut tersebut mencapai 0.990. Seperti yang sudah dijelaskan sebelumnya, penelitian ini akan menggunakan atribut SKP dan Perilaku. Hal ini berlandaskan (PP) Nomor 30 Tahun 2019 tentang Penilaian Kerja Pegawai Negeri Sipsil (PNS) BAB I Pasal 1 Ayat 3 “Prestasi kerja adalah hasil kerja yang dicapai oleh setiap PNS pada satuan organisasi sesuai dengan sasaran kerja pegawai dan perilaku kerja.”(Presiden Republik Indonesia, 2019) dan Perka BKN No 35 Tahun 2011 tentang Pedoman Penyusunan Karir PNS [3].

Selanjutnya untuk mengatasi *imbalanced* data, penelitian ini menggunakan *RandomOverSampling*. Gambar 4 adalah total data berdasarkan kelas sebelum di seimbangkan

0	75
1	28
2	21
3	21
4	10
Name: K, dtype: int64	

Gambar 4 Jumlah data /Atribut kelas

Dapat dilihat bahwa data antar kelas tidak seimbang, untuk kelas A=0 berjumlah 75 data, sedangkan kelas E=4 hanya 10 data. Berikut penanganan data *imbalance* :



Gambar 5. Visualisasi RandomOverSampling

Pada visualisasi gambar 5, data dibagi menjadi 20% sama rata antara 5 kelas yang ada pada atribut K sebagai atribut lable data. Setelah dilakukan RandomOverSampling pada data prestasi kerja PNS, jumlah data di masing-masing kelas sudah berubah menyamai jumlah data mayoritas pada atribut kelas (gambar 6), yaitu kelas A dengan jumlah 75 data. Sehingga, jika ditototal, data yang akan digunakan untuk proses learning sebanyak 375 data.

K	
0	75
1	75
2	75
3	75
4	75
dtype: int64	

Gambar 6. Jumlah data setelah dilakukan ROS

Setelah melalui tahap *Preprocessing*, langkah selanjutnya adalah melakukan pemodelan klasifikasi dengan *Naive Bayes* dan C4.5.

375 data akan dibagi 70% untuk data training dan 30% untuk data testing. Pembagian jumlah data testing dan training seperti pada gambar 7.

(262, 2)
(262, 1)
(113, 2)
(113, 1)

Gambar 7. Train Test Split

Data training berjumlah 262 data, dan data testing 113 data.

1. *Naive Bayes*

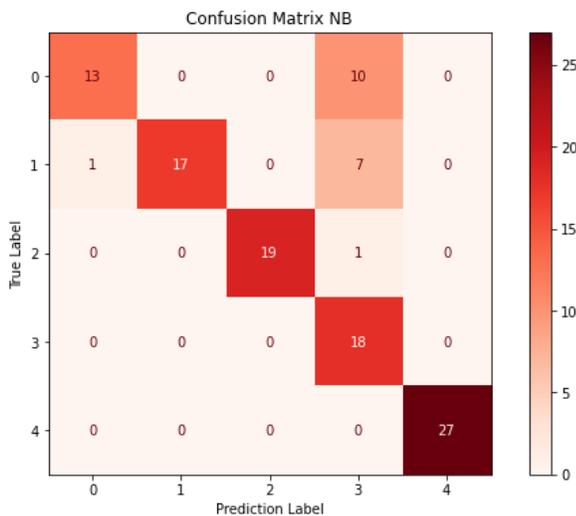
Percobaan pengklasifikasian data yang pertama menggunakan algoritma *Naive Bayes*. Klasifikasi dan prediksi menggunakan *Naive Bayes* menghasilkan prediksi data testing A sebanyak 14 data, B sebanyak 17 data, C sebanyak 19 data, D sebanyak 36 data, dan yang masuk ke kelas E sebanyak 27 data. Kemudian keakuratan prediksi mencapai 0.83 atau 83%. Berikut adalah jumlah data yang diklasifikasikan ke dalam kelas sangat baik, baik, cukup, kurang dan sangat kurang dengan klasifikasi *Naive Bayes*.

2. C4.5

Klasifikasi dan prediksi menggunakan C4.5 menghasilkan data testing yang berjumlah 113 data yang masuk ke dalam kelas A sebanyak 22 data, B sebanyak 26 data, C sebanyak 20 data, D sebanyak 18 data, dan yang masuk ke kelas E sebanyak 27 data. Kemudian keakuratan prediksi mencapai 99.12 atau 99.12%.

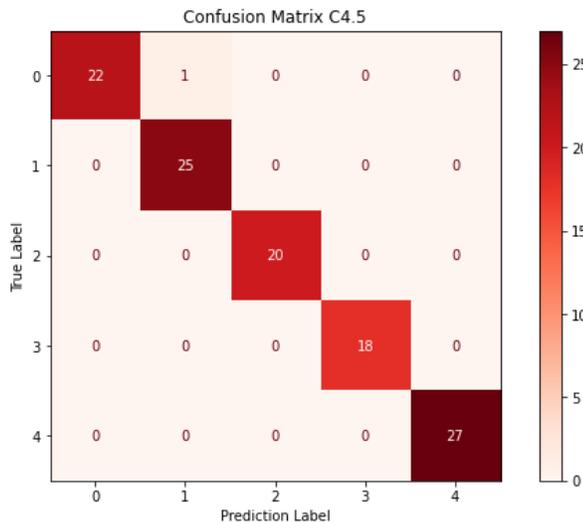
B. Hasil Pembahasan

Dari percobaan pengklasifikasian data prestasi kerja PNS menggunakan metode *Naive Bayes* dan C4.5 akan dibandingkan dengan melihat evaluasi masing-masing model menggunakan confusion matrix. Untuk mengevaluasi model dengan visualisasi *Confusion Matrix* dapat memperhatikan garis diagonal antara sumbu x dan sumbu y yang menyatakan bahwa prediksi data benar. Sedangkan untuk nilai yang ada pada kolom lain dinyatakan prediksi data yang salah. Jumlah nilai atau data yang diprediksi dalam kelas tertentu semakin tinggi atau banyak digambarkan dengan warna merah yang lebih gelap, dan . Jumlah nilai atau data yang diprediksi dalam kelas tertentu semakin kecil atau dikit digambarkan dengan warna yang lebih terang.



Gambar 8. Visualisasi CM Naive-Bayes

Dari visualisasi *Confusion Matrix* pada gambar 8 pemodelan dengan *Naive Bayes* memiliki kesalahan prediksi yang cukup banyak. Terdapat 1 data dengan true label baik atau 1 diprediksikan sebagai 0 atau sangat baik. Kemudian 10 data dengan true label 0 diprediksikan ke dalam kelas 3, lalu ada 1 data dengan true label 2 di prediksikan ke kelas 3, dan yang terakhir 7 data dengan true label 1, di prediksikan sebagai kelas 3. Total kesalahan prediksi adalah 19 data.



Gambar 9. Visualisasi CM C4.5

Visualisasi pada gambar 9 adalah *Confusion Matrix* model C4.5 menggambarkan bahwa prediksi dilakukan dengan lebih baik dibandingkan dengan *Naive Bayes*. Setiap data hamper di prediksi ke dalam kelas dengan baik. Hanya ada 1 kesalahan dimana terlihat pada kolom dengan nilai 1. 1 data ini adalah data yang seharusnya terklasifikasikan ke dalam kelas 0=A. Namun data tersebut termasuk ke dalam kelas 1=B.

1. Evaluasi dengan *Classification Report*

*Classification Report* dapat mengetahui nilai precision, recall, f1-score, dan support dari setiap kelas.

Kemudian akan diketahui pula nilai akurasi dari prediksi yang sudah dilakukan. Berikut table 3 perbandingan evaluasi *Classification Report* antara *Naive Bayes* dan C4.5.

Tabel 2. *Classification Report*

Model	Kelas	Precision	Recall	F1-Score	Support	Akurasi
Naive Bayes	0	0.93	0.57	0.70	23	0.83
	1	1.00	0.67	0.81	25	
	2	1.00	0.95	0.97	20	
	3	0.50	1.00	0.67	18	
	4	1.00	1.00	1.00	27	
C4.5	0	1.00	0.96	0.98	23	0.99
	1	0.96	1.00	0.98	25	
	2	1.00	1.00	1.00	20	
	3	1.00	1.00	1.00	18	
	4	1.00	1.00	1.00	27	

Dari perbandingan *Classification Report* tersebut, algoritma *Naive Bayes* memiliki precision yang tinggi. Kecuali untuk kelas 3 yang hanya memiliki precision 50%. Untuk recall, prediksi kelas 0 dan 1 masih kurang baik karena recall keduanya hanya 57% dan 67%. Sedangkan untuk f1-score prediksi untuk kelas 3 hanya memiliki f1score 67%.

Sedangkan untuk evaluasi *classification matrix* algoritma C4.5 memiliki nilai precision, nilai recall, nilai f1-score yang sangat baik. Setiap kelas memiliki nilai precision, recall, dan f1-score lebih dari 95%. Klasifikasi data prestasi kerja PNS dengan C4.5 memiliki nilai ketepatan prediksi mencapai 99.12%. Sedangkan ketepatan klasifikasi menggunakan *Naive Bayes* hanya 83%.

2. Perbandingan dengan penelitian sebelumnya

Untuk melihat apakah penelitian ini memiliki perkembangan dari penelitian sebelumnya, berikut perbandingan akurasi yang dihasilkan dari penelitian sebelumnya di table 5 dengan penelitian yang penulis lakukan di table 4.

Tabel 3. Akurasi Penelitian Penulis

Judul Penelitian Saat Ini	Model	Akurasi
Komparasi Model Klasifikasi <i>Naive Bayes</i> Dan C4.5 Pada Data Prestasi Kerja Pns	C4.5	99,12%
	<i>Naive Bayes</i>	83,92%

Tabel 4. Akurasi Penelitian Sebelumnya

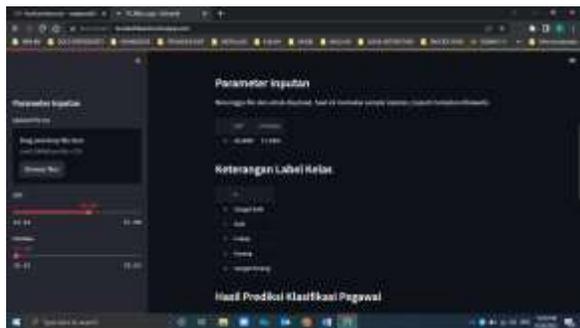
Judul Penelitian Saat Ini	Model	Akurasi

Analysis Of <i>Classic</i> Assumption Test And Multiple Linear Regression Coefficient Test For Employee Structural Office Recommendation [19]	Regresi Linier Berganda	80,66%
Application Of <i>Classification</i> Technique Of Data Mining For Employee Management System [20]	ID3,C4.5 Dan K-Nearest-Neighbor	C4.5=92,69% KNN=73,98% ID3=51,66%
Decision Support Model For Employee Recruitment Using Data Mining <i>Classification</i> [21]	C4.5	88,24%

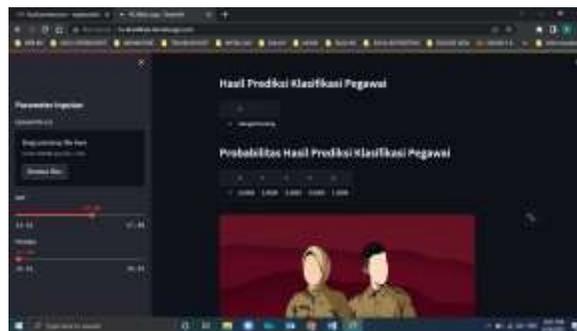
Pada penelitian ini menghasilkan klasifikasi dan prediksi data prestasi kerja PNS dengan nilai akurasi yang lebih tinggi dari penelitian sebelumnya, mencapai 99.12% dengan lagoritma C4.5.

3. *Dashboard Klasifikasi*

Pada dashboard ini, user dapat melakukan prediksi pada data prestasi kerja PNS untuk melihat seorang pegawai masuk dalam kategori Sangat Baik, Baik, Cukup, Kurang atau Sangat Kurang. Terdapat 2 opsi untuk dapat melakukan klasifikasi. Pertama menggunakan fitur slider untuk input parameter nilai SKP dan nilai Perilaku. Kemudian, nilai akan muncul pada table Parameter Inputan seperti pada gambar 10. Lalu, prediksi data akan tampil dalam table Hasil Prediksi Klasifikasi Pegawai seperti pada gambar 11.

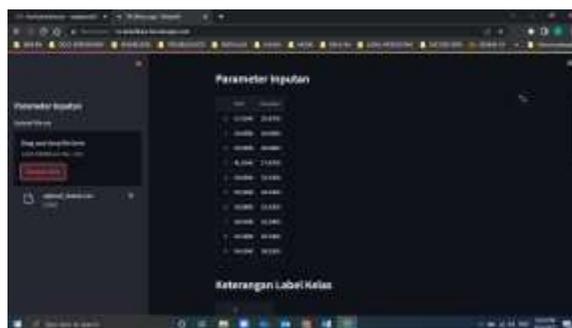


Gambar 10. Parameter Inputan (opsi 1).

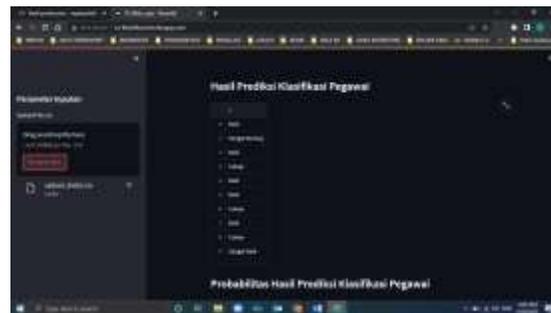


Gambar 11. Hasil Prediksi Klasifikasi Pegawai (opsi 1).

Opsi yang kedua adalah dengan upload file csv. Isi file tersebut adalah 2 kolom inputan parameter dengan urutan SKP kemudian Perilaku. Baris pertama memuat nama atribut yaitu SKP dan perilaku, kemudian baris selanjutnya berisi nilai dari SKP dan Perilaku. Upload file csv pada form upload, kemudian data akan terprediksi dan akan tampil pada table Hasil Prediksi Klasifikasi Pegawai. Dashboard tampilan fungsi fitur tersebut dapat dilihat pada gambar 12 dan gambar 13.



Gambar 12. Inputan Parameter (opsi 2)



Gambar 13. Hasil Prediksi Klasifikasi Pegawai (opsi 2)

Dengan menggunakan dashboard ini, user dapat mengetahui kinerja pegawai berdasarkan nilai SKP dan Nilai perilaku. Dashboard ini mempermudah proses pegawai penilai untuk mengetahui pegawai yang berkinerja mulai dari sangat baik sampai pegawai dengan kinerja sangat buruk.

Pada penelitian ini menghasilkan klasifikasi dan prediksi data prestasi kerja PNS dengan nilai akurasi yang lebih tinggi dari penelitian sebelumnya, mencapai 99.12% dengan lagoritma C4.5.

Data yang digunakan berupa data numerik, yang hasil akhirnya berkaitan terhadap atribut yang digunakan. Tidak

ada spesifitas terhadap nilai dalam atribut tertentu, keduanya saling berhubungan dengan penentuan nilai akhir.

Proses pengujiannya klasifikasi dengan C4.5 dilakukan menyerupai struktur flowchart dan dilakukan berulang-ulang untuk menentukan lable class. C4.5 menggunakan konsep information gain atau entropy reduction yang digunakan untuk pemilihan pembagian yang optimal. Klasifikasi dengan *Naive Bayes* dilakukan dengan melihat nilai probabilitas setiap atribut terhadap lable kelas. *Naive Bayes* yang focus terhadap probabilitas setiap atribut, kurang cocok dengan data yang digunakan dalam penelitian ini. Dikarenakan setiap nilai pada atribut sangat random dan tidak memiliki spesifitas nilai tertentu.

#### IV. KESIMPULAN

Pada proses penelitian yang telah melalui proses perancangan dan juga analisis, untuk menjawab identifikasi permasalahan dan hasil penelitian, maka pada penelitian ini, dapat disimpulkan bahwa:

1. Untuk klasifikasi data prestasi kerja PNS, Algoritma C4.5 lebih baik dari pada algoritma *Naive Bayes*. Hal ini ditunjukkan dengan pengujian Confussion Matrix, yang mana untuk metode C4.5 hanya melakukan 1 kesalahan prediksi data. Sedangkan untuk *Naive Bayes*, terdapat 19 data yang diprediksikan ke dalam kategori yang kurang tepat (tidak sesuai dengan true label data testing). Kemudian dilakukan juga evaluasi classification report, yang mana, nilai precision, recall, dan f1-score dari C4.5 lebih besar dari pada nilai *Naive Bayes* yang dapat dilihat pada BAB IV sub nomor 4.2.2 Untuk klasifikasi data prestasi kerja PNS, Algoritma C4.5 memiliki tingkat akurasi mencapai 99%, sedangkan untuk *Naive Bayes* hanya 83%. Maka C4.5 memiliki nilai hasil kedekatan prediksi data yang lebih baik dibandingkan *Naive Bayes*. Cara untuk mengevaluasi pemodelan C4.5 dan *Naive Bayes* yang dilakukan pada penelitian ini adalah dengan menguji hasil pemodelan menggunakan *Confusion Matrix* dan *Classification Report*, kedua evaluasi ini dapat melihat presisi yaitu ketepatan prediksi, *recall* yaitu keberhasilan prediksi, *f1-score* yaitu error atau resiko prediksi dan akurasi atau kedekatan prediksi dengan nilai dari hasil pemodelan. Evaluasi yang dikatakan baik, adalah nilai precision, recall, f1-score dan akurasi yang tinggi.
2. Klasifikasi data prestasi kerja PNS dapat di prediksi dengan menggunakan dashboard klasifikasi dengan metode C4.5. Data dalam jumlah yang banyak, dapat secara langsung di klasifikasikan ke dalam kategori Sangat Baik, Baik, Cukup, Kurang dan Sangat Kurang.

#### DAFTAR PUSTAKA

- [1] T. Tania, D. Haryadi, W. W. Mirza, and A. M. Khairusy, "Improving employe performance with structural empowerment and transformational leadership through job satisfaction , organizational citizenship behavior and interpersonal trust (study at PT . BPRS Cilegon Mandiri)," Am. J. Humanit. Soc. Sci. Res., vol. 5, no. 11, pp. 91–102, 2021.
- [2] M. T. Lesmana, "Pengaruh Kompetensi Dan Disiplin Kerja Terhadap Kinerja Pegawai," vol. 6681, pp. 665–670, 2017.
- [3] Kepala Badan kepegawaian Negara, "Peraturan Kepala Badan Kepegawaian Negara Nomor 1 Tahun 2013 Tentang Ketentuan Pelaksanaan Peraturan Pemerintah Nomor 45 Tahun 2011 Tentang Penilaian Prestasi Kerja Pegawai Negeri Sipil," pp. 1–135, 2013, [Online]. Available: <https://www.bkn.go.id/wp-content/uploads/2015/02/Perka-Bkn-Nomor-1-Tahun-2013-Ketentuan-Pelaksanaan-Pp-Nomor-46-Tahun-2011-Tentang-Penilaian-Prestasi-Kerja-Pns.pdf>
- [4] P. W. Kastawan, D. M. Wiharta, and M. Sudarma, "Implementasi Algoritma C5.0 pada Penilaian Kinerja Pegawai Negeri Sipil." Maj. Ilm. Teknol. Elektro, vol. 17, no. 3, p. 371, 2018, doi: 10.24843/mite.2018.v17i03.p11.
- [5] D. Suer, "Analisa Penentuan Karyawan Terbaik Menggunakan Data Mining Dengan Metode Algoritma C4.5 Di Pt.Shei Tai Industrial," World Dev., vol. 1, no. 1, pp. 1–15, 2018, [Online]. Available: <http://www.fao.org/3/I8739EN/i8739en.pdf%0Ahttp://dx.doi.org/10.1016/j.adolescence.2017.01.003%0Ahttp://dx.doi.org/10.1016/j.chil dyouth.2011.10.007%0Ahttps://www.tandfonline.com/doi/full/10.1080/23288604.2016.1224023%0Ahttp://pdx.sagepub.com/lookup/doi/10>
- [6] P. Mishra, A. Biancolillo, J. M. Roger, F. Marini, and N. Rutledge, "New data preprocessing trends based on ensemble of multiple preprocessing techniques," Trends Anal. Chem., p. 116045, 2020, doi: 10.1016/j.trac.2020.116045.
- [7] M. K. Dahouda and I. Joe, "A Deep-Learned Embedding Technique for Categorical Features Encoding," IEEE Access, vol. 9, pp. 114381–114391, 2021, doi: 10.1109/Access.2021.3104357.
- [8] A. Elhassan, S. M. Abu-soud, F. Alghanim, and W. Salameh, "ILA4 : Overcoming missing values in machine learning datasets – An inductive learning approach," J. King Saud Univ. - Comput. Inf. Sci., vol. 34, no. 7, pp. 4284–4295, 2022, doi: 10.1016/j.jksuci.2021.02.011.
- [9] K. Surabaya et al., "Antara Kejadian Demam Berdarah Dengue Dengan Kepadatan Penduduk Di Kota Surabaya Pada Tahun 2012 - 2014 Pearson Correlation Analysis to Determine The Relationship Between City Population Density with Incident Dengue Fever of Surabaya in The Year 2012-2014 Widayanti Ratna Safitri Program Studi S1 Ilmu Kesehatan Masyarakat Universitas Airlangga Surabaya," 2014.
- [10] H. Kaur, H. S. Pannu, and A. K. Malhi, "A Systematic Review on Imbalanced Data Challenges in Machine Learning : Applications and Solutions," vol. 52, no. 4, 2019.
- [11] S. Widaningsih, "Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4.5, Naive Bayes, Knn Dan Svm," J. Tekno Insentif, vol. 13, no. 1, pp. 16–25, 2019, doi: 10.36787/jti.v13i1.78.
- [12] A. Y. Simanjuntak, I. S. E. S. Simatupang, and A. Anita, "Implementasi Data Mining Menggunakan Metode Naï• Ve Bayes Classifier Untuk Data Kenaikan Pangkat Dinas Ketenagakerjaan Kota Medan," J. Sci. Soc. Res., vol. 5, no. 1, p. 85, 2022, doi: 10.54314/jssr.v5i1.804.
- [13] Afdhaluzzikri, "Analisa Kinerja Metode Naive Bayes Dengan Pembobotan Data," 2021.
- [14] F. Elfaladonna and A. Rahmadani, "Analisa Metode Classification-Decision Tree Dan Algoritma C.45 Untuk Memprediksi Penyakit Diabetes Dengan Menggunakan Aplikasi Rapid Miner," SINTECH (Science Inf. Technol. J., vol. 2, no. 1, pp. 10–17, 2019, doi: 10.31598/sintechjournal.v2i1.293.
- [15] F. Marisa, A. L. Maukar, I. Khalim, and M. R. Putra, "Jurnal Teknologi dan Manajemen Informatika Analisa Prediksi Varietas Buah Salak yang Sesuai dengan Lahan Daerah," vol. 8, no. 1, pp. 20–25, 2022.
- [16] Binus, "Clustering." [Online]. Available: <https://socs.binus.ac.id/2017/03/09/clustering/>
- [17] M. Azhari, Z. Situmorang, and R. Rosnelly, "Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM dan Naive Bayes," J. Media Inform. Budidarma, vol. 5, no. 2, p. 640, 2021, doi: 10.30865/mib.v5i2.2937.
- [18] Presiden Republik Indonesia, "Peraturan Pemerintah Republik Indonesia Nomor 30 Tahun 2019 Tentang Penilaian Kinerja Pegawai

[1] T. Tania, D. Haryadi, W. W. Mirza, and A. M. Khairusy, "Improving employe performance with structural empowerment and transformational leadership through job satisfaction , organizational

- Negeri Sipil,” Kementeri. Sekr. Negara Republik Indones., pp. 1–52, 2019.
- [19] D. Alita, A. D. Putra, and D. Darwis, “Analysis of Classic assumption test and multiple linear regression coefficient test for employee structural office recommendation,” vol. 15, no. 3, pp. 295–306, 2021.
- [20] S. J. Kamatkar, A. Tayade, A. Viloría, and A. Hernández-Chacín, “Application of classification technique of data mining for employee management system,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10943 LNCS, pp. 434–444, 2018, doi: 10.1007/978-3-319-93803-5\_41.
- [21] C. Elfira, A. Pah, and I. Journal, “Decision Support Model for Employee Recruitment Using Data Mining Classification”.