

Perbandingan Algoritma *Naive Bayes* dan *Decision Tree(C4.5)* dalam Klasifikasi Dosen Berprestasi

Andy Supriyadi¹

¹Sekolah Vokasi, Universitas Sebelas Maret

E-mail: ¹andy.supriyadi@staff.uns.ac.id

Diterima Redaksi: 03 Maret 2023 Revisi Akhir: 09 Maret 2023 Diterbitkan Online: 11 Maret 2023

Abstrak – Salah satu upaya perguruan tinggi untuk menjadi unggul adalah dengan mendorong peningkatan pelaksanaan tridharma perguruan tinggi bagi dosen. Pimpinan perguruan tinggi perlu melakukan upaya agar dapat terpenuhi. Bentuk pemenuhan peningkatan tridharma perguruan tinggi dengan memberikan penghargaan bagi dosen yang berprestasi secara terprogram. Proses pengumpulan data dengan melakukan wawancara dan konsultasi terhadap Kepala Bagian Kepegawaian Universitas Sebelas Maret. Tujuan dari penelitian adalah melakukan perbandingan hasil klasifikasi dengan kelas dosen berprestasi dan dosen tidak berprestasi. Fitur-fitur yang dipergunakan dalam klasifikasi yaitu publikasi internasional, jenjang pendidikan, lama memperoleh serdos dan riwayat jabatan fungsional. Metode yang digunakan dalam penelitian ini yaitu *Naive Bayes* dan *Decision Tree*. Penelitian dilakukan pada Universitas Sebelas Maret pada tahun 2018-2021 dengan menggunakan dataset sebanyak 500 *record* terbagi data latih 350 *record* dan data uji sebanyak 130 *record*. Hasil pengujian menggunakan *10-fold cross validation* metode *Naive Bayes* diperoleh akurasi 94,80%, metode *Decision Tree* sebesar 95,80%.

Kata Kunci — berprestasi, dosen, *decision tree*, klasifikasi, *k-fold cross validation*, *naive bayes*

Abstract – Enhancing the execution of Tri Dharma for lecturers is one of the factors in obtaining and sustaining the level of universities with good institution achievement. The Rectorate should exercise consideration while making a decision to reward lecturers who do very well. The information was gathered through speaking with members of the rectorate staff to classify lectures at Sebelas Maret University. In this study, accuracy results in the classification based on lecturers' accomplishments will be compared. International and national publications, education level, the length of doctoral studies, becoming an associate professor, and the length of certification as a lecturer are the features considered in the classification. To categorize lecturers according to their accomplishment, the algorithms *Naive Bayes* and *Decision Tree* were applied. 350 records of training data and 130 records of testing data total 500 records in this study. From 2018 to 2021, the study was carried out at Sebelas Maret University. The accuracy value obtained from *10-fold cross validation* the testing using the *Naive Bayes* method is 94,80%, while the accuracy value obtained from the testing using the *Decision Tree* is 95,80%.

Keywords — categorization, *decision tree*, *k-fold cross validation*, lecturers, *naive bayes*

1. PENDAHULUAN

Dosen atau pendidik adalah tenaga akademik yang memiliki tugas melakukan proses perencanaan, pengimplementasian pembelajaran, pembimbingan, penelitian dan pengabdian serta melakukan evaluasi dari proses yang telah dilakukan sebagai bentuk pelayanan kepada masyarakat. Tenaga pendidik atau dosen yang telah melakukan kinerja yang baik dalam melaksanakan tridharma pendidikan berhak memperoleh penghargaan atau promosi [1].

Dosen berprestasi adalah seorang pendidik yang telah melaksanakan tridharma perguruan tinggi dalam kurun lima tahun terakhir dan memiliki prestasi yang sangat membanggakan bagi perguruan tinggi tempat yang bersangkutan berhomebase baik dalam tingkat nasional dan internasional [2]. Pimpinan perguruan tinggi sudah selayaknya memberikan penghargaan kepada dosen yang berprestasi. Pemberian penghargaan akan dapat mendorong para dosen untuk lebih produktif dalam menghasilkan karya-karya demi kemajuan pendidikan dan mencapai tujuan nasional yaitu mencerdaskan kehidupan bangsa.

Klasifikasi adalah suatu proses pengelompokkan dengan menggunakan data pelatihan atau pembelajaran dapat berupa data kategori atau *continue* dengan dilakukan pelabelan terhadap atribut yang menjadi luaran atau kelas *record* [3].

Metode *Naïve Bayes* adalah metode klasifikasi dari *machine learning* yang memiliki keunggulan yaitu menggunakan sampel data pelatihan untuk mengestimasi *parameter* yang terlibat dalam proses klasifikasi dapat tersaji secara cepat, serta memperoleh akurasi tinggi. Beragam metode dan algoritma telah diusulkan dari berbagai penelitian terkait klasifikasi dengan metode *Naïve Bayes*, antara lain klasifikasi potensi kinerja *security* atau keamanan, dihasilkan prosentase berkinerja baik, berkinerja cukup dan berkinerja buruk dengan melakukan implementasi *Computer Based Test* untuk mengetahui kemampuan, kepribadian dan ketrampilan *security* [4].

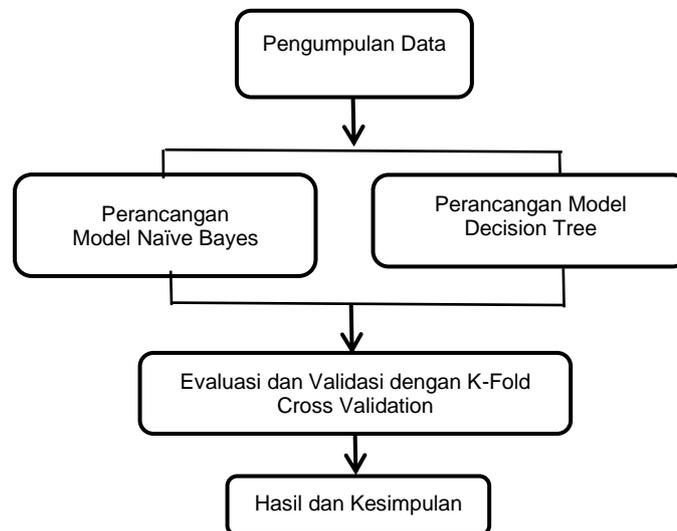
Metode pengujian *K-Fold Cross Validation* adalah contoh jenis pengujian metode validasi untuk mengukur kinerja algoritma dengan melakukan sampel data secara acak serta melakukan pengelompokkan data berdasarkan nilai *k-fold*. Dimana salah satu *fold* dijadikan sebagai data *training* dan sisanya sebagai data uji [5].

Metode perbandingan antara *Naïve Bayes* dan Algoritma *Decision Tree(C.45)* untuk mendeteksi dini gangguan autism pada anak diperoleh hasil akurasi Algoritma *Decision Tree(C.45)* sebesar 72% dan *Naïve Bayes* dihasilkan akurasi 73,33% [6]. Kemudian penelitian berikutnya dilakukan perbandingan *Naïve Bayes* dan Algoritma *C.45* pada kasus klasifikasi data mining untuk memprediksi kelulusan mahasiswa menghasilkan akurasi sebesar 94% untuk metode *Naive Bayes* dan memperoleh akurasi 92,60% untuk metode Algoritma *C.45* [7].

Penelitian ini menitikberatkan pada perbandingan Algoritma *Naive Bayes* dan *Decision Tree (C.45)* dalam melakukan Klasifikasi Dosen Berprestasi dengan jabatan fungsional lektor dan lektor kepala dengan mengolah data riwayat publikasi internasional, jenjang studi, sertifikasi dosen dan riwayat jabatan fungsional, dengan harapan dapat memberikan informasi yang bermanfaat, yaitu dengan melakukan pelabelan data menjadi dua kelas: (i) berprestasi; (ii) tidak berprestasi.

2. METODE PENELITIAN

Tahapan penelitian diawali dengan pengumpulan data dosen yang diambil dari wawancara dan *database* sistem kepegawaian Universitas Sebelas Maret dalam bentuk excel. Kemudian dilakukan pengolahan awal data dengan menghapus data yang tidak lengkap dan memberikan label pada data *training*, pertama pemodelan dengan *Naive Bayes*, kedua dengan *Decision Tree*, kemudian dilakukan pengujian dan evaluasi model dengan *tool Rapidminer Studio versi 9.1*. Alur metode penelitian dapat dilihat pada Gambar 1.

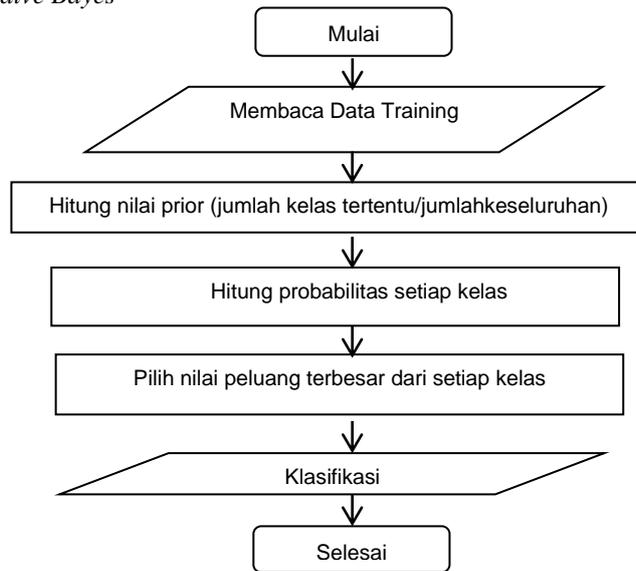


Gambar 1. Metode Tahapan Penelitian

2.1. Pengumpulan Data

Tahap pengumpulan data dilakukan dengan mengumpulkan data mengenai profil dosen yang diperoleh dari bagian sumberdaya manusia Universitas Sebelas Maret dan wawancara yang akan digunakan sebagai dataset. Data yang diperoleh berisi tentang profil dosen yaitu idpegawai, publikasi internasional, jenjang pendidikan, sertifikasi dosen, riwayat jabatan dalam format excel. Data yang diambil pada rentang tahun 2018 sampai dengan bulan Juni tahun 2021 sebanyak mungkin kemudian dipilih 500 *record* yang relevan untuk digunakan sebagai data training sejumlah 350 *record* dan 150 *record* data uji.

2.2. Model Klasifikasi dengan Naïve Bayes



Gambar 2. Diagram Alir Algoritma Naïve Bayes

Gambar 2. menunjukkan proses diagram alir dari algoritma *Naïve Bayes*. *Teorema Bayes* merupakan teorema yang berpedoman pada konsep probabilitas bersyarat. *Naïve Bayes* merupakan metode pengklasifikasian berdasarkan *Teorema Bayes*, yaitu probabilitas dan statistik sebagaimana dikemukakan oleh ilmuwan Inggris *Thomas Bayes*. Metode algoritma ini mengasumsikan bahwa fitur pada data adalah bersifat independen, dengan pengertian sebuah fitur data tidak berkaitan dengan fitur lain dalam data yang sama [8].

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

Keterangan:

- H* : Kelas dari suatu data yang belum teridentifikasi
- X* : Anggapan dasar(Hipotesis) dari kelas yang spesifik
- P(H|X)* : *Conditional probability H* berdasar nilai *X* (*posteriori*)
- P(X|H)* : *Conditional probability X* berdasar nilai hipotesis nilai *H*
- P(H)* : Anggapan dasar(Hipotesis) dari nilai *H* (*priori*)
- P(X)* : Anggapan dasar(Hipotesis) dari nilai *X*

Adapun tahapan menghitung probabilitas adalah

1. Membaca data *training* yang telah dimasukkan ke *database*
2. Melakukan perhitungan perhitungan nilai *probabilitas prior*
3. Menghitung probabilitas setiap kelas
4. Memilih nilai peluang terbesar dari setiap kelas
5. Memperoleh hasil klasifikasi

2.3. Model Klasifikasi dengan Algoritma *Decision Tree* (C.45)

Model *Decision Tree* merupakan salah satu metode klasifikasi dalam data mining yang menghasilkan klasifikasi berupa pohon keputusan berdasarkan sampel data dapat berupa data *predictor* univariat ataupun data *multivariat* [9], [10]. Metode *Decision Tree* ini dapat menyelesaikan permasalahan secara sistematis yang tertuang dalam bentuk pohon keputusan dengan beberapa langkah seperti berikut:

2.3.1. Menghitung Nilai *Gain*

Nilai atau Informasi *Gain* adalah ketergantungan antara dua *variable* acak *X* dan *Y* yang merupakan suatu ukuran dalam korelasi pada sebuah model parametrik. Persamaan Nilai *Gain* dapat dirumuskan:

$$Gain(A) = I(S_1, S_2, \dots, S_m) - E(A) \tag{2}$$

Dimana:

$I(S_1, S_2, \dots, S_m)$ adalah *split* info atau informasi harapan yang dapat dirumuskan:

$$I(S_1, S_2, \dots, S_m) = - \sum_{i=1}^m P_i \log_2(P_i) \quad (3)$$

Keterangan :

m : Banyaknya nilai yang berbeda dari atribut label kelas yang akan melakukan pendefinisian kelas berbeda, $C_i (i = 1, 2, \dots, m)$

S_i : Merupakan jumlah sampel dalam himpunan sampel S (berisi s sampel) yang masuk kedalam kelas C_i

P_i : Merupakan peluang suatu sampel yang akan masuk ke dalam kelas C_i dan di estimasi dengan nilai $\frac{S_i}{s}$

2.3.2. Menghitung Nilai Entropy(A)

Nilai *Entropy* merupakan ukuran yang mempresentasikan ketidakpastian yang tersaji secara probabilistik. Persamaan *Entropy* A sebagai berikut:

$$E(A) = \sum_{j=1}^v \frac{S_{1j} + S_{2j} + \dots + S_{mj}}{s} I(S_1, S_2, \dots, S_m) \quad (4)$$

Dimana

v : merupakan banyaknya nilai dibagi kategori yang berbeda yang dimiliki atribut A

S_{ij} : merupakan banyaknya sampel pada atribut A yang termasuk kategori ke j dan kelas C_i

$\frac{S_{1j} + S_{2j} + \dots + S_{mj}}{s}$: merupakan proporsi jumlah antara sampel atribut A dengan kategori j terhadap jumlah sampel total

2.3.3. Menghitung Gain Ratio

Tahap ke-3 menghitung *Gain Ratio* ini merupakan modifikasi dari *information gain* dengan mengurangi bias yang memiliki banyak cabang. Persamaan *gain ratio* sebagai berikut:

$$\text{Gain Ratio}(S, A) = \frac{\text{Gain}(S, A)}{\text{Split Information}(S, A)} \quad (5)$$

$$\text{Split Information}(S, A) = - \sum_{i=1}^c \frac{S_i}{s} \log_2 \frac{S_i}{s} \quad (6)$$

Nilai c subset adalah nilai S_1 sampai dengan S_c yang dihasilkan dari proses pemecahan S dengan memakai atribut A yang mempunyai sebanyak c nilai.

2.4. Pengujian dengan K-Fold Cross Validation dan Confusion Matrix

Menggunakan metode *K-Fold Cross validation* dan *Confusion Matrix* sebagai metode evaluasi untuk mengetahui nilai *Accuracy*, *Precision* dan *Recall* dari masing-masing metode klasifikasi. Proses evaluasi dan validasi metode dilakukan menggunakan *tool rapidminer versi 9*. Beberapa komponen evaluasi dari klasifikasi antara lain [11]:

1. Banyaknya kelas positif yang diklasifikasi secara benar (TP).
2. Banyaknya kelas negatif yang diklasifikasi secara positif kelas (FP).
3. Banyaknya kelas positif yang diklasifikasi secara negatif (FN).
4. Banyaknya kelas negatif yang diklasifikasi secara benar (TN).

Dalam menghitung *presicion*, *recall* menggunakan komponen pada Tabel 1

| Predictiv class | Classified Actual Class | |
|-----------------|-------------------------|---------------------|
| | Class Positive | Class Negative |
| Positive | TP (True Positive) | FN (False Negative) |

Negative FP (False Positive) TN (True Negative)

Tabel 1. menyajikan evaluasi hasil klasifikasi dengan menggunakan dua kelas yang kemudian dapat dicari nilai *precision*, *recall*, dan akurasi [12]. Definisi dari *precision*, *recall*, dan akurasi sebagai berikut:

1. *Precision*

Precision adalah suatu proses perhitungan banyaknya record atau data dengan nilai *true* positif dibagi dengan banyaknya *record*, seperti yang ditunjukkan pada Persamaan (7).

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

2. *Recall*

Recall adalah banyaknya kelas positif yang diklasifikasi secara benar dibagi dengan penjumlahan banyaknya kelas positif yang diklasifikasi secara benar dan kelas klasifikasi yang diklasifikasi secara *negative*, ditunjukkan pada Persamaan (8).

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

3. *Akurasi*

Akurasi klasifikasi adalah hasil banyaknya kelas yang diklasifikasi secara benar dibagi dengan banyaknya keseluruhan data uji ditunjukkan pada Persamaan (9)

$$Accuracy = \frac{\text{jumlah klasifikasi benar}}{\text{jumlah total data uji}} \tag{9}$$

Klasifikasi bernilai benar, jika model yang diimplementasikan telah dilatih untuk menentukan kelas yang sama dengan kelas *actual* atau asli pada data uji.

3. HASIL DAN PEMBAHASAN

Model klasifikasi dosen berprestasi terbagi menjadi lima komponen utama yaitu pelabelan, pemilihan atribut, pra pemrosesan data, model klasifikasi *naive bayes*, model klasifikasi *decision tree* dan pengujian akurasi, *precision*, *recall*.

3.1. Pelabelan Data

Data *training* yang diperoleh dari *database* bagian sumber daya manusia kemudian dilabeli manual dengan dua label yaitu berprestasi dan tidak berprestasi. Kelas berprestasi berdasarkan publikasi internasional, serdos dan jabatan fungsional dan jenjang pendidikan yang diperoleh.

Tabel 2. Kutipan Dataset

| id | idpeg | publikasi_ internasional | jenjang pendidikan | lama serdos | jabatan fungsional | Golongan ruang | kelas_dospres |
|------|-------|--------------------------|--------------------|-------------|--------------------|----------------|-------------------|
| 1198 | peg60 | tidak_ada | S3 | sedang | Lektor | IV/c | berprestasi |
| 1203 | peg62 | tidak_ada | S3 | sedang | Lektor | IV/c | berprestasi |
| 1204 | peg63 | tidak_ada | S2 | sedang | Lektor | IV/c | tidak_berprestasi |
| 1209 | peg64 | tidak_ada | S3 | cepat | Lektor | IV/c | berprestasi |
| 1210 | peg65 | ada | S2 | sedang | Lektor | IV/b | tidak_berprestasi |
| 1211 | peg66 | tidak_ada | S2 | lambat | Lektor | IV/a | tidak_berprestasi |
| 1212 | peg67 | tidak_ada | S2 | sedang | Lektor | IV/a | tidak_berprestasi |

| | | | | | | | |
|------|-------|-----------|----|--------|--------|------|-------------------|
| 1214 | peg68 | ada | S3 | lambat | LK | IV/a | berprestasi |
| 1215 | peg69 | tidak_ada | S3 | lambat | Lektor | IV/b | tidak_berprestasi |
| 1104 | peg7 | tidak_ada | S3 | sedang | Lektor | IV/c | berprestasi |
| 1216 | peg70 | ada | S3 | sedang | Lektor | IV/b | berprestasi |

3.2. Pemilihan Atribut

Proses pemilihan atribut yang dianggap relevan dalam data mining adalah seleksi fitur. Banyaknya atribut akan mempengaruhi proses komputasi dan jika banyak atribut yang tidak relevan digunakan dalam proses klasifikasi dapat mempengaruhi akurasi. Proses seleksi fitur dengan cara menghilangkan atribut yang tidak relevan sehingga dapat meningkatkan akurasi klasifikasi. Atribut atau fitur yang digunakan yaitu publikasi internasional, jenjang pendidikan, sertifikasi dosen, riwayat jabatan.

3.3. Pra Pemrosesan Data

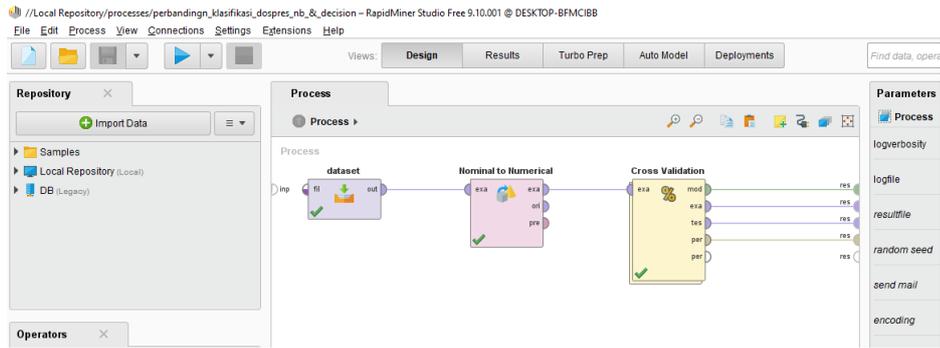
Pra pemrosesan data dilakukan sebagai langkah mendapatkan data *training* yang relevan sehingga siap dipergunakan. Tahapan yang dilakukan dalam pra pemrosesan data yaitu proses pembersihan data, melakukan kombinasi data (*integration*) data menghilangkan data yang ganda. Atribut atau *variable* golongan ruang mengalami proses *reduction*. Sedangkan *variabel* publikasi internasional, jenjang pendidikan, sertifikasi dosen, riwayat jabatan fungsional dilakukan proses diskretisasi.

Tabel 3. Kutipan Data Hasil Preprocessing

| id | idpeg | publikasi_internasional | jenjang pendidikan | lama serdos | jabatan fungsional | kelas_dospres |
|------|-------|-------------------------|--------------------|-------------|--------------------|-------------------|
| 1198 | peg60 | tidak_ada | S3 | sedang | Lektor | berprestasi |
| 1203 | peg62 | tidak_ada | S3 | sedang | Lektor | berprestasi |
| 1204 | peg63 | tidak_ada | S2 | sedang | Lektor | tidak_berprestasi |
| 1209 | peg64 | tidak_ada | S3 | cepat | Lektor | berprestasi |
| 1210 | peg65 | ada | S2 | sedang | Lektor | tidak_berprestasi |
| 1211 | peg66 | tidak_ada | S2 | lambat | Lektor | tidak_berprestasi |
| 1212 | peg67 | tidak_ada | S2 | sedang | Lektor | tidak_berprestasi |
| 1214 | peg68 | ada | S3 | lambat | LK | berprestasi |
| 1215 | peg69 | tidak_ada | S3 | lambat | Lektor | tidak_berprestasi |
| 1104 | peg7 | tidak_ada | S3 | sedang | Lektor | berprestasi |
| 1216 | peg70 | ada | S3 | sedang | Lektor | berprestasi |

3.4. Klasifikasi dengan Algoritma Naïve Bayes

Algoritma *Naïve Bayes* diawali dengan proses menghitung probabilitas setiap kata. Setelah itu dilakukan perhitungan nilai *probabilitas prior* yaitu nilai kemungkinan munculnya kelas tertentu dalam sekumpulan data. Berikut ini tersaji klasifikasi metode *naïve bayes* menggunakan *tool rapidminer versi 9*.



Gambar 1. Implementasi Metode *Naive Bayes* Tahap *Read Dataset*

Proses implementasi metode *Naive Bayes* diawali dengan membaca dataset yang telah melalui proses *preprocessing*, kemudian dilakukan perubahan tipe data yang berupa nominal menjadi tipe data numerik. Selanjutnya dilakukan klasifikasi data seperti pada Gambar 2.

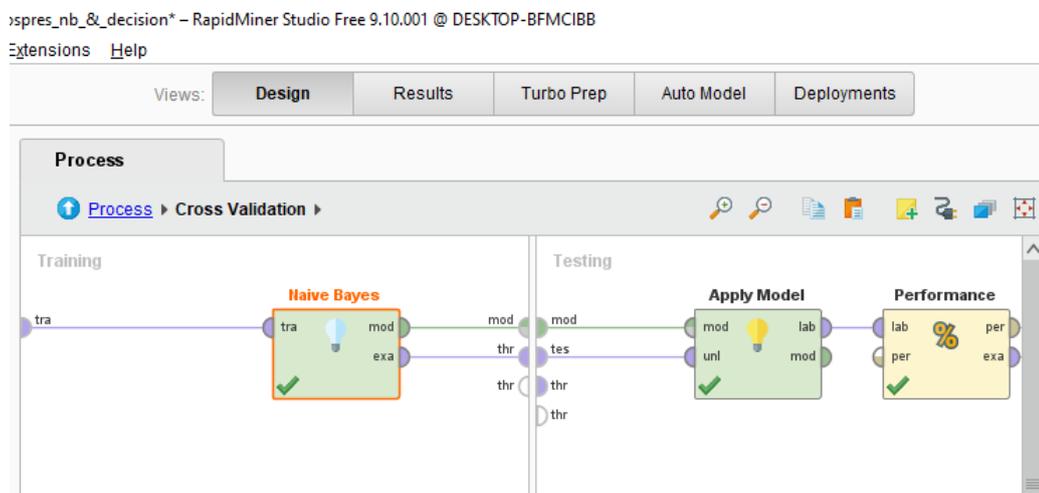
| Row... | kelas_asli(Y) | prediction(kelas... | confidence(... | confidence(t... |
|--------|-------------------|---------------------|----------------|-----------------|
| 1 | berprestasi | berprestasi | 1.000 | 0.000 |
| 2 | tidak_berprestasi | tidak_berprestasi | 0.000 | 1.000 |
| 3 | tidak_berprestasi | tidak_berprestasi | 0.000 | 1.000 |
| 4 | berprestasi | berprestasi | 1 | 0 |
| 5 | tidak_berprestasi | tidak_berprestasi | 0.000 | 1.000 |
| 6 | tidak_berprestasi | tidak_berprestasi | 0.000 | 1.000 |
| 7 | tidak_berprestasi | tidak_berprestasi | 0.000 | 1.000 |
| 8 | berprestasi | berprestasi | 1 | 0 |
| 9 | tidak_berprestasi | tidak_berprestasi | 0.000 | 1.000 |
| 10 | berprestasi | berprestasi | 1.000 | 0.000 |
| 11 | tidak_berprestasi | tidak_berprestasi | 0.000 | 1.000 |
| 12 | berprestasi | berprestasi | 1.000 | 0.000 |
| 13 | berprestasi | berprestasi | 1 | 0 |

ExampleSet (500 examples, 4 special attributes, 10 regular attributes)

Gambar 2. Kutipan Hasil Klasifikasi *Naive Bayes*

Berdasarkan Gambar 2. dengan menggunakan *dataset* sejumlah 500 data terbagi data latih 350 data dan data uji 150 data dilakukan klasifikasi dengan metode *Naive Bayes* diperoleh hasil klasifikasi benar 474 data dan 26 data menunjukkan hasil yang salah. Langkah selanjutnya melakukan pengujian metode.

3.5. Pengujian Akurasi, Precision, Recall pada metode *Naive Bayes*



Gambar 3. Pengujian metode *Naive Bayes* dengan 10-Fold Cross Validation

Pada tahapan pengujian metode *Naïve Bayes* menggunakan *k-fold crossvalidation* dengan nilai *10-fold* sebagaimana tersaji pada Gambar 3. Dari proses tersebut dihasilkan prediksi dosen berprestasi dan dosen tidak berprestasi dengan nilai *true* (diprediksi benar) dan nilai *false* (diprediksi salah) sebagai berikut:

Tabel 4. Hasil Prediksi Metode *Naïve Bayes*

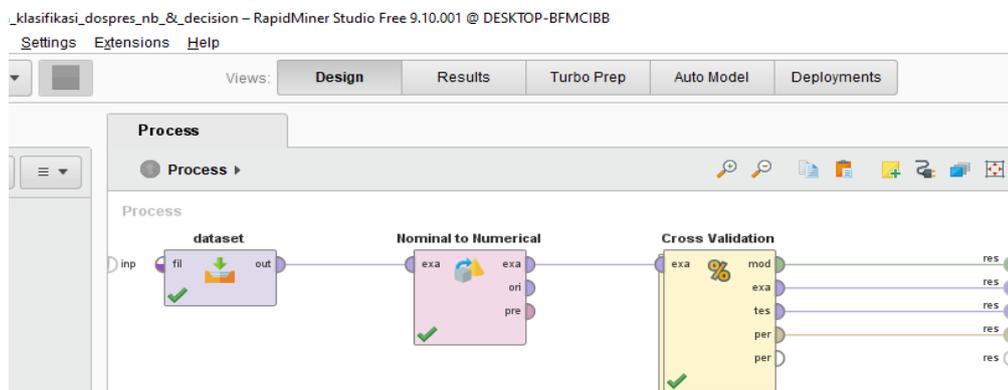
| | Aktul Dosen Berprestasi | Aktual Dosen Tidak Berprestasi | Jumlah |
|----------------------------------|-------------------------|--------------------------------|--------|
| Prediksi Dosen Berprestasi | 281 | 14 | 295 |
| Prediksi Dosen Tidak Berprestasi | 12 | 193 | 205 |
| Jumlah | 293 | 207 | 500 |

Berdasarkan pada Tabel 4. diperoleh prediksi dosen berprestasi 295 *record*, dan prediksi dosen tidak berprestasi 205 *record*, kelas yang berhasil melakukan klasifikasi secara benar (*true*) dosen berprestasi sebanyak 281 *record*, sedang kelas yang salah diprediksi (*false*) sebanyak 14 *record* dengan hasil klasifikasi dosen berprestasi, sedangkan hasil pemodelan dosen tidak berprestasi. Kemudian untuk hasil prediksi dosen tidak berprestasi berhasil melakukan klasifikasi benar (*true*) 193 *record* dan 12 *record* terprediksi salah (*false*).

Tahap selanjutnya melakukan evaluasi model dengan *confusion matrix*. Proses ini melakukan perhitungan akurasi, *precision*, *recall* dan diperoleh akurasi 94,80%, nilai *recall* 95,90% untuk kelas dosen berprestasi, kelas tidak berprestasi diperoleh *recall* 93,24%, kemudian kelas dosen berprestasi diperoleh *precision* 95,25% dan kelas dosen tidak berprestasi diperoleh *precision* 94,15%.

3.6. Klasifikasi dengan Algoritma *Decision Tree* (C.45)

Metode *Decision Tree* adalah bagian dari metode *machine learning* yang menghasilkan klasifikasi berupa pohon keputusan berdasarkan *dataset* yang diujikan. Selanjutnya dilakukan klasifikasi *dataset* yang diperoleh, dengan komposisi 70% untuk data latih dan 30% data uji. Tahap selanjutnya dilakukan klasifikasi menggunakan tool *rapidminer* seperti pada Gambar 4.



Gambar 4. Implementasi Metode *Decision Tree*

Proses klasifikasi dengan metode *Decision Tree* diawali dengan membaca *dataset*, kemudian proses *preprocessing* dan *apply model*. Hasil klasifikasi yang diperoleh seperti tersaji pada Gambar 5.

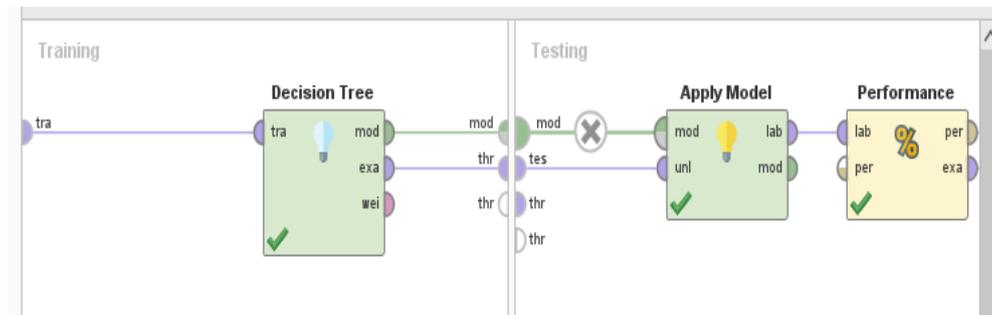
| Row No. | kelas_asli(Y) | prediction(k... | confidence(... | confidence |
|---------|-----------------|-----------------|----------------|------------|
| 2 | tidak_berpre... | tidak_berpre... | 0.058 | 0.942 |
| 3 | tidak_berpre... | tidak_berpre... | 0.058 | 0.942 |
| 4 | berprestasi | berprestasi | 0.972 | 0.028 |
| 5 | tidak_berpre... | tidak_berpre... | 0.058 | 0.942 |
| 6 | tidak_berpre... | tidak_berpre... | 0.058 | 0.942 |
| 7 | tidak_berpre... | tidak_berpre... | 0.058 | 0.942 |
| 8 | berprestasi | berprestasi | 0.972 | 0.028 |
| 9 | tidak_berpre... | tidak_berpre... | 0.058 | 0.942 |
| 10 | berprestasi | berprestasi | 0.972 | 0.028 |
| 11 | tidak_berpre... | tidak_berpre... | 0.058 | 0.942 |
| 12 | berprestasi | berprestasi | 0.972 | 0.028 |
| 13 | berprestasi | berprestasi | 0.972 | 0.028 |
| 14 | tidak_berpre... | tidak_berpre... | 0.058 | 0.942 |

ExampleSet (500 examples, 4 special attributes, 10 regular attributes)

Gambar 5. Kutipan Hasil Klasifikasi *Decision Tree*

Berdasarkan Gambar 5. dengan menggunakan *dataset* sejumlah 500 data terbagi atas data latih 350 data kemudian data uji 150 data selanjutnya dilakukan klasifikasi dengan metode *Decision Tree* diperoleh hasil klasifikasi benar 479 data, dan 21 data menunjukkan hasil yang salah

3.7. Pengujian Akurasi, Precision, Recall pada metode *Decision Tree*



Gambar 6. Pengujian metode *Decision Tree* dengan 10-Fold Cross Validation

Tahap selanjutnya melakukan pengujian metode *Decision Tree* menggunakan *k-fold crossvalidation* dengan rapidminer dengan nilai *10-fold* seperti ditunjukkan pada Gambar 6. Dari proses evaluasi model dihasilkan prediksi dosen berprestasi dan dosen tidak berprestasi sebagai tersaji pada Tabel 5.

Tabel 5. Hasil Prediksi Metode *Decision Tree*

| | Aktul Dosen Berprestasi | Aktual Dosen Tidak Berprestasi | Jumlah |
|----------------------------------|-------------------------|--------------------------------|--------|
| Prediksi Dosen Berprestasi | 281 | 9 | 290 |
| Prediksi Dosen Tidak Berprestasi | 12 | 198 | 210 |
| Jumlah | 293 | 207 | 500 |

Berdasarkan pada Tabel 5. dapat disampaikan prediksi dosen berprestasi 290 *record*, dan prediksi dosen tidak berprestasi 210 *record*, kelas yang benar (*true*) diklasifikasi berprestasi 281 dan kelas yang salah diprediksi 9 *record* dengan hasil klasifikasi dosen berprestasi, sedangkan pada pemodelan dosen tidak berprestasi. Kemudian hasil prediksi dosen tidak berprestasi diprediksi secara benar (*true*) 198 *record* dan terdapat kesalahan prediksi dosen tidak berprestasi (*false*) oleh pemodelan diklasifikasi menjadi dosen berprestasi 12 *record*.

Selanjutnya sesuai proses implementasi metode *Decision Tree* yang ditunjukkan pada Gambar 4. dan Tabel 5. diperoleh hasil akurasi 95,80%, dengan nilai akurasi *recall* 95,90% untuk kelas dosen berprestasi dan kelas dosen tidak berprestasi diperoleh 95,65%, kemudian akurasi *precision* kelas dosen berprestasi diperoleh 96,90% dan dosen tidak berprestasi 94,29%.

Berdasarkan percobaan yang telah dilakukan peneliti dalam melakukan implementasi metode klasifikasi *naïve bayes* dan *decision tree* didapatkan hasil *performace* atau akurasi klasifikasi seperti ditunjukkan dalam Tabel 6.

Tabel 6. Hasil Perbandingan Algoritma *Naive Bayes* dan *Decision Tree* (C4.5)

| No | Metode | Akurasi | standar deviasi (+/-) | Precision | | Recall | |
|----|-------------------------|---------|-----------------------|-------------|-------------------|-------------|-------------------|
| | | | | Berprestasi | Tidak Berprestasi | Berprestasi | Tidak Berprestasi |
| 1 | Algoritma Naïve Bayes | 94,80% | 3,79% | 95,25% | 94,15% | 95,90% | 93,24% |
| 2 | Algoritma Decision Tree | 95,80% | 2,74% | 96,90% | 94,29% | 95,90% | 95,65% |

4. SIMPULAN DAN SARAN

Setelah penelitian ini dilakukan, diperoleh hasil penelitian dengan kesimpulan bahwa hasil perbandingan klasifikasi dosen berprestasi menggunakan metode *Naïve Bayes* mendapatkan nilai akurasi sebesar 94,80 %, dengan nilai standar deviasi sebesar 3,79%, nilai presisi kelas berprestasi sebesar 95,25% dan nilai presisi kelas tidak berprestasi sebesar 94,15%, selanjutnya nilai *recall* kelas berprestasi sebesar 95,90%, *recall* kelas tidak berprestasi 93,24% dan metode *Decision Tree* mendapatkan nilai akurasi sebesar 95,80%, nilai standar deviasi sebesar 2,74% dengan nilai presisi berprestasi sebesar 96,90%, nilai kelas tidak berprestasi 94,29% dan nilai *recall* berprestasi 95,90% dan kelas tidak berprestasi 95,65%. Sehingga dapat disimpulkan kedua metode melakukan klasifikasi dosen berprestasi dan tidak berprestasi dengan baik. Melalui penelitian ini diharapkan dapat membantu pimpinan perguruan tinggi dalam mengambil keputusan untuk memberikan pembinaan bagi dosen.

DAFTAR PUSTAKA

- [1] Menteri Pendidikan dan Kebudayaan, “Peraturan Menteri Pendidikan dan Kebudayaan Nomor 3 Tahun 2020 Tentang Standar Nasional Pendidikan Tinggi”.
- [2] Menteri Riset Teknologi dan Pendidikan Tinggi, “Pedoman Dosen Berprestasi Tahun 2019”.
- [3] A. Muslim, M., A., Prasetiyo, Budi., M, E. L.H., H, A.J., Mirqotussa’adah., R, S.H., Nurzahputra, “Data Mining Algoritma C4.5,” 2019, Accessed: Feb. 09, 2023. [Online]. Available: http://lib.unnes.ac.id/33080/6/Buku_Data_Mining.PDF
- [4] F. Ariani, Amir, N. Alam, and K. Rizal, “Klasifikasi Penetapan Status Karyawan Dengan Menggunakan Metode Naïve Bayes,” *Paradig. - J. Komput. dan Inform.*, vol. 20, no. 2, 2018.
- [5] A. Rohani, M. Taki, and M. Abdollahpour, “A novel soft computing model (Gaussian process regression with K-fold cross validation) for daily and monthly solar radiation forecasting (Part: I),” *Renew. Energy*, vol. 115, 2018, doi: 10.1016/j.renene.2017.08.061.

- [6] B. Sugara, D. Adidarma, and S. Budilaksono, "Perbandingan Akurasi Algoritma C4.5 dan Naïve Bayes untuk Deteksi Dini Gangguan Autisme pada Anak," *J. IKRA-ITH Inform.*, vol. 3, no. 1, 2019.
- [7] S. J. S. Tyas, M. Febianah, F. Solikhah, A. L. Kamil, and W. A. Arifin, "Analisis Perbandingan Algoritma Naive Bayes Dan C.45 Dalam Klasifikasi Data Mining Untuk Memprediksi Kelulusan," *J. Teknol. Inf. Dan Komun.*, vol. 8, no. 1, 2021.
- [8] 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," in *Journal of Physics: Conference Series*, 2019, vol. 1280, no. 2. doi: 10.1088/1742-6596/1280/2/022022.
- [9] F. Syahputra, H. Hartono, and R. Rosnelly, "Penerapan Algoritma C4.5 Dalam Memprediksi Ketersediaan Uang Pada Mesin ATM," *J. MEDIA Inform. BUDIDARMA*, vol. 5, no. 2, 2021, doi: 10.30865/mib.v5i2.2933.
- [10] 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, 2021, doi: 10.30865/mib.v5i2.2937.
- [11] N. B. Putri and A. W. Wijayanto, "Analisis Komparasi Algoritma Klasifikasi Data Mining Dalam Klasifikasi Website Phishing," *Komputika J. Sist. Komput.*, vol. 11, no. 1, 2022, doi: 10.34010/komputika.v11i1.4350.
- [12] I. W. Saputro and B. W. Sari, "Uji Performa Algoritma Naïve Bayes untuk Prediksi Masa Studi Mahasiswa," *Creat. Inf. Technol. J.*, vol. 6, no. 1, 2020, doi: 10.24076/citec.2019v6i1.178.