

Perbandingan Dan Analisis Metode Klasifikasi Untuk Menentukan Konsentrasi Jurusan

Indah Hidayanti¹⁾, Tri Basuki Kurniawan²⁾, Afriyudi³⁾

¹⁾²⁾³⁾Program Studi Magister Teknik Informatika, Fakultas Ilmu Komputer, Universitas Bina Darma
Jl. Jenderal Ahmad Yani No.3, 9/10 Ulu, Kota Palembang, Sumatera Selatan 30111

Email : indahhidayanti23@gmail.com¹⁾, tribasukikurniawan@binadarma.ac.id²⁾, afriyudi@binadarma.ac.id³⁾

Abstract

Higher education is the organizer of further education after the high school level. Five institutions of higher education, namely universities, institutes, high schools, academies and polytechnics. One of the factors that determine the quality of higher education is the percentage of students' ability to complete their studies on time. This is certainly obtained by choosing the right concentration of students during the lecture period and the academic ability of students. Determination of student concentration specifications is needed to determine the interests and talents of the students themselves. With the selection of the right concentration it is expected that students can graduate on time. At this time students have difficulty in determining the concentration of majors. At Bina Darma University in Palembang, especially in the Faculty of Computer Science, there are Informatics Engineering (IT). At the Faculty of Computer Science, the IT Study Program has a concentration of expertise such as Database, Software, and Network Infrastructure that are in accordance with the IT Study Program curriculum at Bina Darma University. Determination of concentration is carried out at the end of semester 4 or precisely the lecture period for semester 5. However, what happens is that students do not know their interests and abilities. At this time students have difficulty in determining the concentration of majors. The choice of concentration of the department is only based on the wishes of the students or by joining friends, so that there is a need for preference in helping students choose concentration. The methods used in data mining are C4.5 and Naïve Bayes by using the Rapid Miner application as a tool to classify student majors. In this study, it is know that C4.5 algorithm has a high accuracy of 48,06% and Naïve Bayes 42,79%

Keywords : C4.5, Naïve Bayes, Klasifikasi Jurusan, Rapid Miner

Abstrak

Perguruan tinggi merupakan penyelenggara pendidikan lanjutan setelah tingkat sekolah menengah atas. Salah satu faktor yang menentukan kualitas perguruan tinggi yaitu persentasi kemampuan mahasiswa untuk menyelesaikan studi tepat pada waktunya. Hal ini tentunya didapatkan dengan pemilihan konsentrasi mahasiswa yang tepat pada masa perkuliahan serta kemampuan akademis dari mahasiswa. Penentuan spesifikasi konsentrasi mahasiswa sangat dibutuhkan untuk menentukan minat dan bakat mahasiswa itu sendiri. Dengan pemilihan konsentrasi yang tepat diharapkan mahasiswa dapat lulus tepat pada waktunya. Pada saat ini mahasiswa kesulitan dalam menentukan konsentrasi jurusan. Di Universitas Bina Darma Palembang terutama di Fakultas Ilmu Komputer, terdapat Program Studi Teknik Informatika (TI). Pada Fakultas Ilmu Komputer ini, Program Studi TI mempunyai konsentrasi keahlian seperti Database, Software, dan Jaringan Infrastructure yang sesuai dengan kurikulum Program Studi TI di Universitas Bina Darma. Penentuan konsentrasi jurusan yaitu pada akhir semester 4 atau tepatnya masa perkuliahan untuk semester 5. Akan tetapi yang terjadi yaitu mahasiswa tidak tahu minat dan kemampuannya masing-masing. Saat ini mahasiswa kesulitan dalam menentukan pemilihan konsentrasi jurusan. Pemilihan konsentrasi jurusan hanya berdasarkan pada keinginan mahasiswa ataupun ikut-ikutan teman, untuk itu dirasa perlu adanya preferensi dalam membantu mahasiswa memilih konsentrasi. Metode yang digunakan yaitu C4.5 dan Naïve Bayes dengan menggunakan aplikasi Rapid Miner sebagai alat bantu untuk mengklasifikasikan penjurusan mahasiswa. Pada penelitian ini diketahui algoritma C4.5 memiliki tingkat akurasi 48,06 % dan naïve bayes 42,79%.

Kata kunci: C4.5, Naïve Bayes, Klasifikasi Jurusan, Rapid Miner

1. Pendahuluan

Semakin banyak dan bertambahnya jumlah perguruan tinggi maka semakin meningkat pula jumlah sumber daya manusia berkualitas yang dihasilkan perguruan tinggi tersebut. Salah satu faktor yang menentukan kualitas perguruan tinggi adalah persentasi kemampuan mahasiswa untuk menyelesaikan studi tepat waktu. Hal ini tentunya didapatkan dengan pemilihan konsentrasi mahasiswa yang tepat pada masa perkuliahan serta kemampuan akademis dari mahasiswa. Penentuan spesifikasi konsentrasi mahasiswa sangat dibutuhkan untuk menentukan minat dan bakat mahasiswa itu sendiri. Dengan pemilihan konsentrasi yang tepat diharapkan mahasiswa dapat lulus tepat waktu.

Di Universitas Bina Darma Palembang terutama di Fakultas Ilmu Komputer, terdapat Program Studi Teknik Informatika (TI). Program Studi TI mempunyai konsentrasi keahlian seperti *Database*, *Software*, dan Jaringan *Infrastructure* yang sesuai dengan kurikulum Program Studi TI di Universitas Bina Darma. Sistem pemilihan konsentrasi keahlian idealnya dipilih berdasarkan minat mahasiswa serta kemampuan akademik. Mahasiswa diwajibkan untuk memilih salah satu dari konsentrasi jurusan yang ada. Penentuan konsentrasi dilaksanakan pada akhir semester 4 atau tepatnya masa perkuliahan untuk semester 5. Akan tetapi yang terjadi mahasiswa tidak tahu minat dan kemampuannya masing-masing. Pada saat ini mahasiswa kesulitan dalam menentukan konsentrasi jurusan. Pemilihan konsentrasi jurusan hanya berdasarkan dari keinginan mahasiswa ataupun ikut-ikutan teman, untuk itu dirasa perlu adanya preferensi dalam membantu mahasiswa memilih konsentrasi. Hal ini diharapkan agar menjadi bahan rekomendasi bagi mahasiswa dalam menentukan pemilihan konsentrasi nantinya.

Pada penelitian ini, penulis akan membandingkan kinerja kedua metode klasifikasi dalam *data mining* yaitu algoritma C4.5 dan *Naive Bayes* untuk mendapatkan hasil pengujian paling akurat dalam mengelola informasi data mahasiswa sebagai dasar penentuan dalam pemilihan konsentrasi keahlian mahasiswa dengan sample data dari Program Studi TI Universitas Bina Darma Palembang. Data itu sendiri berdasarkan dari nilai mata kuliah semester 1 sampai dengan semester 4. Perbandingan kinerja algoritma C4.5 dan *Naive Bayes* bertujuan untuk mengukur tingkat akurasi terbaik masing-masing algoritma untuk diterapkan dalam kasus pemilihan konsentrasi keahlian mahasiswa.

Penelitian yang mengangkat topik tentang klasifikasi dan penerapan data mining telah banyak dilakukan sebelumnya. Beberapa penelitian tentang klasifikasi diantaranya: Penelitian yang dilakukan oleh Choirul Anam dan Harry Budi Santoso yaitu perbandingan kinerja algoritma C4.5 dan *Naive Bayes* untuk klasifikasi penerimaan beasiswa. Penelitian menggunakan data sekunder berupa daftar pemohon dan penerima beasiswa sebagai data set yang memiliki enam faktor

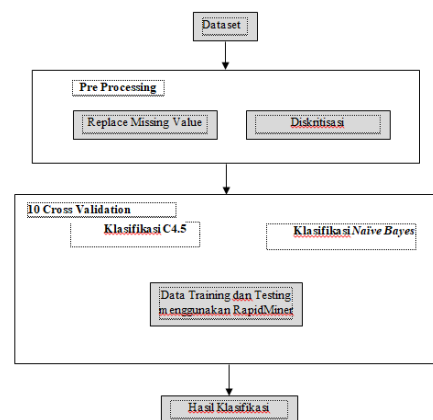
penentu, yaitu semester, IPK, prestasi ko/ekstra kurikuler, penghasilan orang tua, beban biaya listrik dan jumlah tanggungan orang tua. Pengujian dengan 10-fold cross validation sekaligus evaluasi kinerja model menggunakan *tool Rapid Miner*. Hasil penelitian menunjukkan tingkat akurasi model algoritma C4.5 sebesar 96.40% lebih baik dari tingkat akurasi model algoritma Naive Bayes sebesar 95.11%. [1]

Selanjutnya Moh Ainul Yaqin melakukan penelitian tentang komparasi metode klasifikasi dalam penentuan jurusan menggunakan 4 metode *K-NN*, *Neural Network*, C4.5 dan *Naive Bayes*. Atribut yang digunakan yaitu data mahasiswa NIM, Nama, Tanggal Lahir, Asal Sekolah, Nilai UN, Jurusan yang dipilih. Data lain yang akan diolah adalah Indeks Prestasi Kumulatif Semester 1 dan Indeks Prestasi Kumulatif Semester 2. Dari hasil penelitian di dapatkan hasil akurasi *Neural Network* adalah 15,82%, C4.5 adalah 44,28%, *Naive Bayes Classifier* adalah 73,94% dan *K-NN* adalah 54,61% [14]

2. Pembahasan

A. Tahapan Penelitian

Pada penelitian ini, data yang digunakan adalah data alumni dan mahasiswa Teknik Informatika yang masih aktif. Data nilai tersebut akan diolah menggunakan beberapa metode data mining. Berikut tahapan dalam melakukan penelitian data mining:



Gambar 1. Model Tahapan Penelitian

B. Data Penelitian

Untuk mendapatkan data penelitian, maka penulis mengumpulkan data-data berupa data alumni mahasiswa Teknik Informatika angkatan 2013 – 2014 dan data mahasiswa yang masih aktif yaitu angkatan 2015-2016. Data yang di ambil ini merupakan data yang memiliki kurikulum dan mata kuliah yang sama. Data yang digunakan seperti data Nim, IP semester 1 sampai dengan semester 4, IPK pada semester 4, dan mata kuliah. Data sendiri diambil di bagian Direktorat Sistem Teknologi Informasi Universitas Bina Darma. Data yang didapatkan sebanyak 599 data dengan *preprocessing* maka data yang dianggap tidak layak dihapus, sehingga data yang layak digunakan sebanyak 299 data.

C. Pengolahan Data Mining

Data mining sebenarnya merupakan salah satu bagian proses *Knowledge Discovery in Database* (KDD) yang bertugas untuk mengekstrak pola atau model dari data dengan menggunakan suatu algoritma yang spesifik Adapun proses KDD sebagai berikut [11] :

1. Data Selection

Data didapat dari data alumni mahasiswa Teknik Informatika angkatan 2013 – 2014 dan data mahasiswa yang masih aktif yaitu angkatan 2015-2016 pada Fakultas Ilmu Komputer, Program Studi TI Universitas Bina darma Palembang yang digunakan sebanyak 229 data. Untuk data Training dan Testing yaitu data akademik dengan konsentrasi *Database, software*, dan Jaringan *Infrastructure* angkatan 2013 sampai dengan 2016 dengan IPK di atas 3 dimana data yang di butuhkan yaitu data NIM, KHS semester 1 sampai dengan semester 4, data IPS Semester 1 Sampai dengan semester 4. Data yang dipilih merupakan data mahasiswa dengan IPK di atas 3 karena dianggap telah layak .

Uji pertama melalui data angkatan 20013-2016, atribut Nim sebagai ID sedangkan konsentrasi sebagai label. Sedangkan atribut mata kuliah seperti algoritma & pemrograman, bahasa inggris 1, Proses bisnis, praktikum struktur data, struktur data, praktikum basis data, matematika diskrit, bahasa inggris 3, praktek relation management system, dan bahasa Indonesia dihapus menurut gain terendah untuk mendapatkan tingkat akurasi yang lebih tinggi.

2. Pra Pemrosesan Data

Data yang digunakan berasal dari data alumni mahasiswa TI Universitas Bina Darma tahun ajaran 2013- 2014 dan data mahasiswa yang masih aktif 2015-2016 sebagai data set yang akan di implementasikan menggunakan aplikasi Rapid Miner dengan metode C4.5 dan *Naïve Bayes*.

Dari proses pengumpulan data didapatkan sebanyak 599 data konsentrasi dari mahasiswa yang telah lulus dan mahasiswa yang telah memilih konsentrasi dengan 42 atribut. Namun tidak semua data dan atribut dapat digunakan sebanyak 32 atribut yang dapat di uji proses klasifikasi. Adapun rincian 32 atribut dalam pra pemrosesan data yaitu sebagai berikut :

Tabel 1. Attribute Data Set Klasifikasi Jurusan

1	NIM	17	ARSITEKTUR & ORGANISASI KOMP.
2	IP1	18	PENDIDIKAN KEWARGANEGARAAN
3	IP2	19	BASIS DATA
4	IP3	20	BAHASA INGGRES 2
5	IP4	21	STATISTIK PROBABILITAS
6	IP1-4	22	SISTEM OPERASI
7	PENDIDIKAN AGAMA	23	REKAYASA PERANGKAT LUNAK
8	PENGANTAR TEKNOLOGI INFORMASI	24	TEORI GRAPH
9	PRAKTIKUM INTERNET	25	JARINGAN KOMPUTER 2
10	KONSEP BAHASA PEMROGRAMAN	26	RELATIONAL DB MANAGEMENT SYST.
11	KALKULUS	27	REKAYASA PERANG. LUNAK LANJ.
12	PRAKTIKUM PEMROGRAMAN	28	JARINGAN KOMPUTER 3
13	PENGANTAR BASIS DATA	29	SISTEM OPERASI LANJUTAN
14	JARINGAN KOMPUTER 1	30	PRAK. SISTEM OPERASI LANJUTAN
15	ETIKA PROFESI	31	INTERAKSI MANUSIA DAN KOMPUTER
16	LOGIKA INFORMATIKA	32	KONSENTRASI

Beberapa teknik pra pemrosesan data yang dilakukan yaitu [4] :

- Pembersihan data, untuk mengidentifikasi dan menghapus data yang tidak konsisten dan data yang tidak lengkap. Dalam penelitian ini record mahasiswa dengan status berhenti atau non aktif dihapus karena mengandung nilai mata kuliah yang tidak lengkap . Data yang digunakan yaitu data mahasiswa dengan IPK di atas 3 karena di anggap data yang telah layak. Selebihnya data yang tidak layak dihapus. Sehingga data yang awalnya 599 kemudian terpaksa dihapus sebanyak 300 data sehingga menjadi 229 data yang akan digunakan dalam proses data mining.
- Reduksi data, untuk memperoleh data set dengan jumlah atribut dan *record* yang lebih sedikit tetapi bersifat informatif. Dalam penelitian ini atribut Nama dan jenis kelamin tidak digunakan dalam proses mining. Data yang di butuhkan yaitu data NIM, KHS semester 1 sampai dengan semester 4, data IPS Semester 1 Sampai dengan semester 4.

3. Transformation

Pada tahapan ini yaitu merubah data kedalam bentuk yang sesuai di mining. Teknik data transformator digunakan untuk melakukan perubahan terhadap data. Teknik data transformation yaitu *descretize* yang digunakan untuk merubah data dengan tipe continuous menjadi data dengan tipe diskrit. *Descretize* adalah salah satu metode yang digunakan saat *preprocessing input*. Teknik diskritisasi data dapat digunakan untuk mengurangi jumlah nilai *attribute numeric* dengan cara membagi *range* atribut ke dalam interval. Label interval kemudian dapat digunakan untuk menggantikan nilai-nilai aktual data. Mengganti nilai dalam jumlah yang banyak dari atribut terus menerus dengan sejumlah kecil label interval dapat mengurangi dan menyederhanakan data asli. Sehingga proses yang dilakukan menjadi singkat dan mudah dalam mempresentasikan tingkat pengetahuan dari hasil proses data mining [6].

4. Data Mining

Data Mining adalah kajian yang meliputi kegiatan pengumpulan, pembersihan, pemrosesan, dan analisa sekumpulan data sehingga dengan kegiatan tersebut dapat diperoleh pemahaman yang mendalam akan data. *Data mining* adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Setelah melakukan proses *preprocessing* dan transformasi data yang sesuai untuk penggunaan teknik data mining klasifikasi dengan algoritma C4.5 dan *Naive Bayes*, maka tahapan selanjutnya adalah melakukan proses analisis data mining yang telah dikerjakan pada tahap sebelumnya. Proses analisis *data mining* ini hanya dilakukan untuk mempermudah dan mengerti penggunaan teknik *klasifikasi* dengan metode algoritma C4.5 dan *Naive Bayes* menggunakan aplikasi *Rapid Miner*.

Klasifikasi jurusan mahasiswa berdasarkan data mahasiswa menggunakan algoritma C4.5 dan *Naive Bayes* serta memanfaatkan *tools RapidMiner* sebagai *Software* untuk membantu menentukan klasifikasi

jurusan. Data yang digunakan yaitu data mahasiswa dan data KHS (kartu hasil studi) yang menampung semua data mahasiswa ditahun 2013 sampai dengan tahun 2016 yang ada di Universitas Bina Darma Palembang yang kemudian menghasilkan informasi pola klasifikasi. Data diolah menggunakan metode KDD (Knowledge Discovery in Database), dari banyaknya data mahasiswa yang di gunakan 229 record data mahasiswa yang dapat digunakan. Hasil antara pengimplementasian algoritma C4.5 sesuai dengan pemodelan yang dihasilkan oleh RapidMiner. Pemodelan yang telah di proses oleh tools RapidMiner untuk mengetahui pola klasifikasi dan keakuratan data dengan menggunakan model Performance . Dengan mengetahui jumlah data yang diklasifikasikan secara benar maka diketahui hasil akurasi yaitu dari hasil data testing serta mendapatkan pola klasifikasi yang berupa pohon keputusan dan rules dari pola yang terbentuk.

D. Algoritma C4.5

Algoritma C4.5 dan pohon keputusan merupakan dua model yang tak terpisahkan, karena untuk membangun sebuah pohon keputusan dibutuhkan algoritma C4.5. Algoritma C4.5 merupakan pengembangan dari algoritma ID3. Algoritma C4.5 dan ID3 diciptakan oleh seorang peneliti di bidang kecerdasan buatan bernama J. Rose Quinlan pada akhir tahun 1970-an[10]. Algoritma C4.5 juga menggunakan kriteria *gain* dalam menentukan fitur yang menjadi pemecahan *node* pada pohon yang induksi [9]. Untuk memilih atribut sebagai akar, didasarkan pada nilai Gain tertinggi dari atribut-atribut yang ada. Untuk menghitung Gain dapat digunkan persamaan berikut [8].

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

Keterangan:

- S : Himpunan Kasus
- A : Atribut
- n : Jumlah Partisi Atribut A
- |S_i| : Jumlah Kasus Pada Partisi Ke - i
- |S| : S Jumlah Kasus Dalam S

Untuk menghitung nilai Entropy dapat dilihat pada persamaan berikut:

$$Entropy(S) = \sum_{i=1}^n P_i * Log_2 P_i \dots \dots (2)$$

Keterangan:

- S : Himpunan Kasus
- A : Fitur
- n : Jumlah Partisi S
- P_i : Proporsi Dari S_i Terhadap S

E. Naïve Bayes Classifier

Naive Bayes Classifier adalah sebuah metoda klasifikasi yang berdasar pada teorema Bayes. Metode pengklasifikasian ini menggunakan metode probabilitas dan statistik yang pertama kali dikemukakan oleh ilmuwan Inggris bernama Thomas Bayes, yaitu suatu metode untuk memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya, sehingga metode ini dikenal sebagai Teorema Bayes. Ciri utama dari Naive Bayes Classifier ini adalah asumsi yang

sangat kuat akan independensi dari masing-masing kondisi atau kejadian [7] .

Naive Bayes sering bekerja jauh lebih baik dalam kebanyakan situasi dunia nyata yang kompleks dari pada yang diharapkan. Persamaan dari teorema *Bayes* adalah [14] :

Prediksi *Bayes* didasarkan pada teorema *Bayes* dengan formula umum sebagai berikut:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \dots \dots (3)$$

Penjelasan dari formula tersebut adalah sebagai berikut:

- X : Data dengan *class* yang belum diketahui
- H : Hipotesis data merupakan suatu *class* spesifik
- P(H|X) : Probabilitas hipotesis H berdasar kondisi X (posteriori probabilitas)
- P(H) : Probabilitas hipotesis H (prior probabilitas)
- P(X|H) : Probabilitas X berdasarkan kondisi pada hipotesis H
- P(X) : Probabilitas X

F. Rapid Miner

Rapid Miner memiliki kurang lebih 500 operator data mining, termasuk operator untuk input, output, data *preprocessing* dan visualisasi. *RapidMiner* merupakan *software* yang berdiri sendiri untuk analisis data dan sebagai mesin *data mining* yang dapat diintegrasikan pada produknya sendiri. *Rapid Miner* ditulis dengan menggunakan bahasa java sehingga dapat bekerja di semua sistem operasi [12].

G. Evaluasi Model Klasifikasi Menggunakan Confusion Matrix

Confusion matrix adalah suatu metode yang biasanya digunakan untuk melakukan perhitungan akurasi pada konsep data mining. Rumus ini melakukan perhitungan dengan 3 keluaran, yaitu: *accuracy*, *recall*, *precision* [3]. *Confusion Matrix* sangat berguna untuk menganalisa kualitas *classifier* dalam mengenali tupel-tupel dari kelas yang ada. Metode *Confusion Matrix* mempresentasikan hasil evaluasi model dengan menggunakan tabel matrix, jika data set terdiri dari dua kelas, kelas pertama dianggap positif dan kelas kedua dianggap negatif. Evaluasi menggunakan *Confusion Matrix* menghasilkan nilai *accuracy*, *precision* dan *reccal*. Contoh *Confusion Matrix* dapat dilihat pada Gambar 2:

Gambar 2. Confusion Matrix

		Ya	Tidak	Jumlah
Kelas Aktual	Ya	TP	FN	P
	Tidak	FP	TN	N
	Jumlah	P	N	P+N

Beberapa kegiatan yang dapat dilakukan dengan menggunakan data hasil klasifikasi dalam confusion matrix diantaranya [2] :

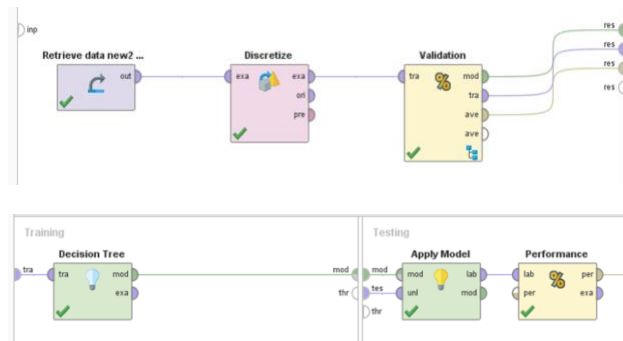
Menghitung nilai rata-rata keberhasilan klasifikasi (overall success rate) ke dalam kelas yang sesuai dengan cara membagi jumlah data yang terklasifikasi dengan benar, dengan seluruh data yang diklasifikasi.

Selain itu dilakukan pula perhitungan persentase kelas positif (true positive dan false positive) yang diperoleh dalam klasifikasi.

Recall precision berfungsi menghitung persentase false positive dan false negative untuk menemukan informasi di dalamnya

2. Pembahasan

A. Implementasi Klasifikasi Menggunakan Metode C4.5 Pada Rapid Miner



Gambar 3. Implementasi Data Set pada Rapid Miner

Gambar di atas merupakan proses pengolahan data mining. Implementasi proses pengolahan data menggunakan diskritisasi data yang digunakan untuk mengurangi jumlah nilai *atribute numeric* dengan cara membagi *range* atribut ke dalam interval. Label interval kemudian dapat digunakan untuk menggantikan nilai-nilai aktual data. X- *Validation* dilakukan untuk mendapatkan hasil akurasi. Hasil pemodelan yang telah di proses oleh *Tool RapidMiner* selain menghasilkan dalam bentuk pola pemodelan ini juga dapat mengetahui tingkat akurasi, *recall*, dan *precision*. Adapun hasil keakurat data yang digunakan dengan metode C4.5 terdapat pada Gambar 4.

accuracy: 48.06% +/- 5.39% (mikro: 48.03%)

	true Software	true IT Infrastructure	true Database	class precision
pred. Software	8	2	1	72.73%
pred. IT Infrastructure	71	93	41	45.37%
pred. Database	2	2	9	69.23%
class recall	9.88%	95.88%	17.65%	

Gambar 4. Hasil Perhitungan Nilai Akurasi dari C4.5

Dari gambar 4 di atas diketahui nilai akurasi pada metode C4.5 yaitu sebesar 48,06 %, *class recall* untuk *true software* 9,88%, *it infrastructure* 95,88%, *database* 17,65%. Sedangkan untuk nilai *class precision* pada *software* sebesar 72,73%, *it infrastructure* 45,83%, *database* 69,23%. Dari data sebanyak 229 yang diprediksi konsentrasasi *true software* yaitu 8 data, yang dipredidiksi konsentrasi *true IT infrastructure* sebanyak 93 data, dan yang di prediksi konsentrasi *true database* sebanyak 9 data.

B. Implementasi Klasifikasi Menggunakan Metode Naïve Bayes Pada Rapid Miner

Hasil permodelan yang telah di proses oleh *Tool RapidMiner* dengan metode *Naïve Bayes* dapat diketahui nilai akurasi data tersebut seperti terlihat pada gambar 5 di bawah ini.

accuracy: 42.79% +/- 10.57% (mikro: 42.79%)

	true Software	true IT Infrastructure	true Database	class precision
pred. Software	38	31	14	45.78%
pred. IT Infrastructure	25	38	15	48.72%
pred. Database	18	28	22	32.35%
class recall	46.91%	39.16%	43.14%	

Gambar 5. Hasil Perhitungan Nilai Akurasi dari Naïve Bayes

Dari gambar 5 di atas diketahui nilai akurasi pada metode C4.5 yaitu sebesar 48,06 %, *class recall* untuk *true software* 9,88%, *it infrastructure* 95,88%, *database* 17,65%. Sedangkan untuk nilai *class precision* pada *software* sebesar 72,73%, *it infrastructure* 45,83%, *database* 69,23%. Dari data sebanyak 229, yang diprediksi masuk konsentrasasi *software* yaitu 38 data, yang di prediksi konsentrasi *IT infrastructure* sebanyak 38 data, dan yang di prediksi masuk konsentrasi *database* sebanyak 22 data.

C. Perbandingan Tingkat Akurasi Metode Algoritma C4.5 dan Naive Bayes Classifier

Berdasarkan hasil implementasi yang telah dilakukan, berikut ini tabel perbandingan tingkat akurasi antara *Algoritma C4.5* dan metode *Naive Bayes Classifier* :

Tabel 2. Perbandingan Tingkat Akurasi Metode Algoritma C4.5 dan Naive Bayes Classifier

No	Metode	Akurasi
1	<i>Algoritma C4.5</i>	48,06%
2	<i>Naive Bayes Classifier</i>	42,79%

Pengujian data yang dilakukan diatas dengan mengurangi atribut data set, pengurangan dilakukan untuk mendapatkan hasil yang lebih tinggi dengan menghapus nilai gain yang kecil didapatkan hasil analisis tingkat akurasi dengan algoritma C4.5 sebesar 48.06%, sedangkan *Naïve Bayes* 42,79%.

Berdasarkan jurnal penelitian sebelumnya yang di bahas di pendahuluan , jumlah atribut yang digunakan yaitu data mahasiswa NIM, Nama, Tanggal Lahir, Asal Sekolah, Nilai UN, Jurusan yang dipilih serta nilai IP semester 1 dan 2 untuk jumlah atribute label sebanyak 2 atribut. Sedangkan yang penulis gunakan sebagai atribut sebanyak 32 atribut dan data set label sebanyak 3 atribut. Hal ini mempengaruhi tingkat performance. Semakin banyak atribut label maka semakin banyak noise yang dihasilkan. Kemudian *performance* tingkat akurasi juga bergantung pada komposisi data dan jenis data. Ketidak konsistenan data nilai mata kuliah yang ada pada data set

menjadi faktor penentu tingkat akurasi data yang kecil. Penulis beranggapan kemungkinan ini kesalahan dalam menentukan label, karena label tidak relevannya antara konsentrasi yang dipilih mahasiswa yang berdasarkan minat dengan karakteristik data mahasiswa tersebut. Misalkan mahasiswa dengan nilai karakteristik masuk ke dalam *it infrastruktur* malah memilih konsentrasi *software*.

Metode C4.5 dalam pengujian ini lebih tinggi dibandingkan *naïve bayes* dilihat dari karakteristik masing-masing metode. Pada metode C4.5 sendiri mempunyai konsep yang jelas, pohon keputusan yang dihasilkan dapat menjelaskan model klasifikasi data. C4.5 bersifat mengeksplorasi data dan berusaha menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target. Sedangkan *Naïve Bayes* merupakan teknik prediksi berbasis probabilitas sederhana dengan asumsi independensi (ketidak tergantungan) yang memandang variable pada sebuah data tidak berkaitan.

3. Kesimpulan

Pada penelitian ini dilakukan pemodelan menggunakan algoritma C4.5 dan *Naïve Bayes* dengan menggunakan data yang sudah diolah menggunakan metode KDD (*Knowledge Discovery in Database*). Dari hasil observasi terhadap dataset mahasiswa Fakultas Teknik Informatika Universitas Bina Darma melalui proses menggunakan metode klasifikasi dan algoritma C4.5 dan *Naïve Bayes* dengan tools Rapid Miner. Dari 32 atribut yang digunakan, variable mata kuliah yang paling tinggi nilai gain yaitu interaksi manusia dan komputer yang dijadikan sebagai node akar. Menurut implementasi C4.5, variable yang mempengaruhi dalam klasifikasi jurusan yaitu nilai mata kuliah interaksi manusia dan komputer, relation DB management sistem, jaringan komputer 2, jaringan komputer 1, bahasa inggris 2, jaringan komputer 3, konsep bahasa pemrograman dan IP semester 4. Berdasarkan prediksi data mining menggunakan algoritma C4.5 dan *Naïve Bayes* dapat ditarik kesimpulan bahwa masing-masing metode telah memprediksi klasifikasi konsentrasi dengan perbandingan keakurasian metode masing-masing yaitu C4.5 48,06% sedangkan *naïve bayes* 42,79%. Selisih nilai akurasi antara kedua metode tersebut adalah sebesar 3,16%.

Dilihat dari hasil akurasi masih sangat rendah, penulis beranggapan kemungkinan ada kesalahan dalam penentuan label. Dari pemilihan konsentrasi mahasiswa yang random berdasarkan minatnya saja dapat dilihat bahwa pemilihan konsentrasi itu terjadi kesalahan. Untuk penelitian selanjutnya perlu dilakukan klastering untuk menentukan label yang tepat.

Daftar Pustaka

Anam, C., & Santoso, H. B. 2018. Perbandingan Kinerja Algoritma C4.5 dan Naive Bayes Untuk Klasifikasi Penerimaan Beasiswa. *ISSN : 2088-4591*.

- Andri, Kunang, Y. N., & Murniati, S. 2013. Implementasi Teknik Data Mining Untuk Memprediksi Tingkat Kelulusan Mahasiswa Pada Universitas Bina Darma Palembang. *Seminar Nasional Informatika*, ISSN: 1979-2328.
- Aprillia, D., & dkk. 2013. *Belajar Data Mining dengan RapidMiner*. Jakarta.
- Dewi, S. 2019. Komparasi Metode Algoritma Data Mining Pada Prediksi Uji Kelayakan Credit Approval Pada Calon Nasabah Kredit Perbankan. *Jurnal Khatulistiwa Informatika, Vol VII, No. 1*, ISSN : 2579 - 633X59.
- Gumilar, D., & dkk. 2017. Prediksi Kepopuleran Lagu Berdasarkan Tangga Lagu Billboard Menggunakan Decision Tree dan K-Means. *ISBN:978-602-1180-50-1*.
- Kusmini, L., & Taufiq, E. 2009. *Algoritma Data Mining*. Yogyakarta: Andi.
- Mirqotussa'adah, & Muslim, M. A. 2017. Penerapan Dizcretization dan Teknik Bagging Untuk Meningkatkan Akurasi Klasifikasi Berbasis Ensemble pada Algoritma C4.5 dalam Mengdiagnosa Diabetes. *Lontar Komputer Vol 8 No.2*, ISSN 2541-5832.
- Mustafa, M. S., & Ramadhan, M. R. 2017. Implementasi Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma. *Citec Journal, Vol. 4, No. 2*, ISSN : 2460 - 4259.
- Novianti, B. 2016. Implementasi Data Mining Dengan Algoritma C4.5 Untuk Penjurusan Siswa (Studi Kasus SMA Negeri 1 Pontianak). *Coding, Sistem Komputer Untan ISSN : 2338-493X*, 76-77.
- Prasetyo, E. 2014. *Mengelolah Data Menjadi Informasi Menggunakan Matlab*. Yogyakarta: Andi.
- Purwanto, A., & Darmadi, E. A. 2018. Perbandingan Minat Siswa SMU Pada Metode Klasifikasi Menggunakan 5 Algoritma. *Ikraith-Informatika, Vol.2 No.1*.
- Sembiring, M. A., Larasati, M. F., & Sapta, A. 2017. Analisa Kinerja Algoritma C4.5 dalam Memprediksi Hasil Belajar. *Journal of Science and Social Research*, ISSN 2615-3262.
- Supriyanti, W., Kusri, & Amborowati, A. 2016. Perbandingan Kinerja Algoritma C4.5 Dan Naive Bayes Untuk Ketepatan Pemilihan Konsentrasi Mahasiswa . *Jurnal Informa Politeknik Indonusa Surakarta*, ISSN : 2442-7942 Vol. 1.
- Suyanto. 2017. *Data Mining Untuk Klasifikasi dan Klastering Data*. Bandung: Informatika.
- Yaqin, M. A. 2017. Komparasi Metode Klasifikasi Dalam Penentuan Penjurusan Dengan Menggunakan 4 Metode (K-NN, Neural Network, C.4.5 Dan Naïve Bayes. *JIT-Vol. 1 No 2*.