

ALGORITMA KLASIFIKASI DATA MINING UNTUK PREDIKSI STATUS MAHASISWA MENGGUNAKAN ALGORITMA DECISION TREE (C4.5)

Wachid Darmawan

STMIK Widya Pratama Pekalongan
Jl. Patriot 25 Pekalongan Telp (0285) 427816
email: wachid.dw@gmail.com

ABSTRAK

Mahasiswa yang mengalami gagal studi merupakan salah satu kerugian bagi sebuah Perguruan Tinggi (PT), salah satunya tentang akreditasi. Mahasiswa gagal studi disini maksudnya mahasiswa yang tidak melakukan pembayaran biaya kuliah untuk semester berjalan atau mahasiswa non-aktif. Jika masalah mahasiswa non-aktif bisa di prediksi lebih cepat, maka pihak manajemen dapat mencegah dan mengantisipasi lebih awal agar tidak ada mahasiswa non-aktif. Dari banyaknya jumlah mahasiswa non-aktif maka perlu dianalisis lebih awal guna menemukan informasi yang berbasis pengetahuan dan bermanfaat. Oleh sebab itu diperlukan suatu cara untuk menggambarkan sekumpulan data secara ringkas. Serangkaian proses untuk mendapatkan pola atau pengetahuan dari kumpulan data disebut Data mining. Salah satu algoritma data mining adalah klasifikasi serta banyak bidang ilmu yang menerapkan teknik klasifikasi dalam ilmu data mining untuk menyelesaikan masalah. Banyak dataset yang digunakan untuk penelitian klasifikasi, salah satu algoritma klasifikasi yang digunakan adalah algoritma Pohon Keputusan (Decision Tree). Algoritma Decision Tree yang digunakan untuk klasifikasi status mahasiswa adalah algoritma C4.5. Algoritma C4.5 salah satu algoritma klasifikasi populer dan mudah di pahami. Berdasarkan penelitian yang sudah dilakukan didapatkan hasil sebagai berikut: algoritma Naive Bayes menghasilkan akurasi sebesar 92,72%, sedangkan algoritma Decision Tree (C4.5) akurasinya sebesar 93,05%.

Kata Kunci: Status Mahasiswa, Algoritma Klasifikasi, Decision Tree, Naive Bayes, Data Mining

1. PENDAHULUAN

Perguruan Tinggi (PT) merupakan sarana pendidikan yang dapat menyelenggarakan proses Pendidikan Tinggi (Peraturan Pemerintah 2014). PT menjadikan peranan penting dalam mewujudkan sumber daya manusia yang handal dan berkualitas. Berdasarkan undang-undang nomor 12 tentang pendidikan tinggi menjelaskan bahwa PT dapat menyelenggarakan pendidikan profesi jika PT menyelenggarakan pendidikan akademik dan menyelenggarakan vokasi dalam satu rumpun ilmu pengetahuan dan sudah memenuhi syarat (Undang-Undang 2012).

Dengan demikian PT dituntut untuk melahirkan sumber daya manusia (mahasiswa) yang ahli di bidangnya. Dalam perkembangannya mahasiswa di sebuah PT yang sedang melangsungkan perkuliahan tidak semuanya lulus berjalan. Ada kalanya sebagian mahasiswa

putus ditengah jalan dalam perkuliahan atau tidak melanjutkan kuliah lagi. Mahasiswa yang menempuh pendidikan di sebuah PT dibedakan oleh status kemahasiswaan, diantaranya: mahasiswa dengan status lulus tepat waktu, mahasiswa dengan status lulus tidak tepat waktu, mahasiswa dengan status non-aktif bahkan mahasiswa dengan status drop out. Ini dapat menyebabkan mereka gagal dalam menempuh studi.

Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK) Widya Pratama merupakan perguruan tinggi di Kota Pekalongan yang berdiri sejak tahun 2002 berdasarkan SK Mendiknas Nomor: 149/D/O/2002 yang diselenggarakan oleh Yayasan Pendidikan Widya Pratama yang memiliki mahasiswa sebanyak 1778. Tabel 1 menunjukkan bahwa mahasiswa dengan status non-aktif bertambah setiap semester.

Tabel 1. Mahasiswa dengan Status Non-Aktif Tahun 2012

MHS	SMT 1	SMT 2	SMT 3	SMT 4	SMT 5	SMT 6
Aktif	722	667	650	633	616	594
Non Aktif	0	55	72	89	106	128
TOTAL	722	722	722	722	722	722

Berdasarkan penelitian Romero dan Ventura (2010) banyak masalah yang mendasari kegagalan studi setiap mahasiswa dengan berbagai faktor penyebabnya (Marquez-Vera et al. 2011). Penyebab gagal studi bagi mahasiswa salah satunya adalah mahasiswa non-aktif. Jika masalah status mahasiswa non-aktif dapat diketahui lebih dini penyebabnya, maka pihak manajemen STMIK Widya Pratama dapat mengantisipasi dan mencegah berkurangnya jumlah mahasiswa yang aktif dalam perkuliahan.

Serangkaian proses untuk mendapatkan pola atau pengetahuan dari kumpulan data disebut Data mining (Witten et al. 2011)(Kusrini & Luthfi 2009). Salah satu algoritma data mining adalah klasifikasi. Banyak dataset yang digunakan untuk penelitian klasifikasi, salah satunya adalah algoritma Pohon Keputusan (*Decision Tree*) (Witten et al. 2011)(Kusrini & Luthfi 2009). Algoritma *Decision Tree* salah satu algoritma klasifikasi populer yang mudah di pahami (Prasetyo 2012).

Algoritma *Decision Tree* dalam pengembangannya ada dua algoritma yang digunakan yaitu *ID3* dan *C4.5*. Algoritma *C4.5* dalam pemecahan node pada pohon yang diinduksi menggunakan kriteria *gain* untuk menentukan fitur (Prasetyo 2014). Algoritma *C4.5* menggunakan representasi struktur pohon (*tree*) yang mana atributnya ada disetiap internal *node*, cabangnya merupakan nilai dari atribut sedangkan daunnya merupakan label (kelas) (Kusrini & Luthfi 2009) (Maimon & Rokach 2010).

Permasalahan mahasiswa non-aktif ini banyak menarik perhatian para peneliti. Hastuti (2012) melakukan klasifikasi mahasiswa non-aktif dengan 22 atribut. Penelitian ini mengkomparasi beberapa algoritma klasifikasi diantaranya *logistic regression*, *decision tree*, *naïve bayes* dan *neural network*, dengan menggunakan 3861 dataset mahasiswa Universitas Dian Nuswantoro. Hasilnya adalah algoritma *decision tree* akurasi paling baik (Hastuti 2012). Royanti (2015) melakukan klasifikasi mahasiswa non-aktif

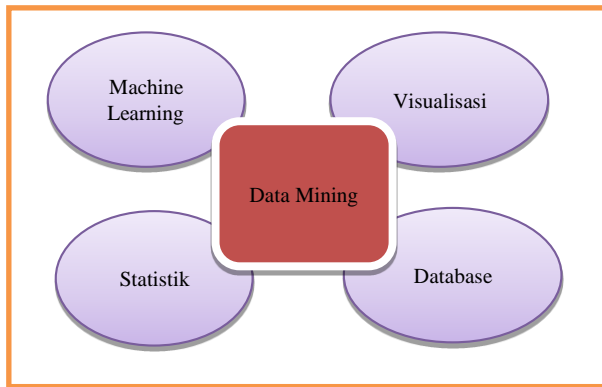
dengan 12 atribut, dengan menggunakan 632 dataset mahasiswa STMIK Widya Pratama. Penelitian ini mengkomparasi algoritma klasifikasi *naïve bayes* dengan tahap seleksi fitur dengan *gain ratio* pada tahap *preprocessing* untuk memprediksi mahasiswa non-aktif. Hasil dari penelitian ini adalah komparasi algoritma *naïve bayes* dan *gain ratio* akurasi meningkat (Royanti 2015).

Untuk mencari solusi terbaik terhadap masalah klasifikasi status mahasiswa, maka dalam penelitian ini menggunakan algoritma klasifikasi *Decision Tree* dengan algoritma *C4.5*. Dengan pemanfaatan algoritma *C4.5* diharapkan dapat diperoleh model terbaik untuk menyelesaikan masalah klasifikasi status mahasiswa. Sehingga manajemen dapat dengan mudah memprediksi status mahasiswa non-aktif di lingkungan kampus.

2. LANDASAN TEORI

2.1. Data Mining

Serangkaian proses untuk mendapatkan pola atau pengetahuan dari kumpulan data disebut Data mining (Witten et al. 2011). Teknik pengenalan pola yang digunakan data mining seperti halnya pada statistik dan matematika, sehingga data mining dapat menganalisa kasus yang sudah ada untuk menemukan pola yang ada dari suatu data (Larose 2005). Data Mining disebut juga sebagai *Knowledge Discovery in Database (KDD)* adalah bidang ilmu yang membahas tentang pola pada suatu data (Witten et al. 2011). Data mining dapat menganalisa data yang besar dengan membuat sebuah aturan, pola, ataupun model tertentu untuk mengenali data baru (Prasetyo 2012). Kegiatan untuk pengumpulan dan pemakaian data historis pada suatu data untuk menemukan pola, keteraturan dalam hubungan dalam data set disebut data mining (Witten et al. 2011). Data mining memiliki kaitan dengan berbagai bidang ilmu yang lain seperti *Machine Learning*, Statistik, Visualisasi serta database (Prasetyo 2014). Gambar 1 merupakan posisi data mining dengan berbagai disiplin ilmu lain.



Gambar 1. Posisi Data Mining dengan Berbagai Disiplin Ilmu

Beberapa peran utama data mining adalah: *Estimation, Prediction, Classification, Clustering* dan *Association*. Dari semua peranan data mining tersebut terbagi menjadi 2 berdasarkan metode pembelajarannya yaitu *Supervised Learning* dan *Unsupervised Learning* (Santosa 2007).

2.2. Klasifikasi

Klasifikasi merupakan salah satu metode yang sering digunakan di data mining. Klasifikasi termasuk kedalam *supervised learning* karena dalam proses klasifikasi terdapat proses pembelajaran dengan data yang telah lampau (Witten et al. 2011). Teknik klasifikasi banyak diterapkan dalam dunia nyata seperti halnya dalam dunia pendidikan (Kotsiantis et al. 2003)(Hastuti 2012)(Ragab et al. 2014), medis (Christobel & Sivaprakasam 2011), teknik bangunan (Ashari et al. 2013), dan dibanyak bidang lainnya.

Label dalam klasifikasi atau bisa juga disebut dengan *class* merupakan atribut yang akan dicari perhitungan algoritma data mining. Sebagai contoh dalam dunia medis jika ada pasien baru dengan gejala penyakit tertentu akan tetapi jenis penyakit yang dideritanya belum diketahui. Maka klasifikasi dapat menjadi sebuah alat untuk menentukan keputusan. Adanya data lampau atau yang nantinya disebut dengan data training akan banyak membantu dalam proses klasifikasi tersebut. Karena dengan banyaknya data training akan mempengaruhi akurasi keakuratan klasifikasi suatu algoritma data mining. Performa suatu algoritma juga dapat dipengaruhi oleh banyaknya atribut (Prasetyo 2012)(Prasetyo 2014).

Dalam melakukan suatu klasifikasi dibutuhkan data lampau yang nantinya akan diolah menjadi sebuah aturan ataupun sebuah

pengetahuan baru. Masalah klasifikasi pada dasarnya adalah sebagai berikut (Susanto & Suryadi 2010)asu:

1. Masalah Klasifikasi berangkat dari data training yang tersedia.
2. Data training akan diolah dengan menggunakan algoritma klasifikasi.
3. Masalah klasifikasi berakhir dengan dihasilkannya sebuah pengetahuan yang direpresentasikan dalam bentuk diagram, aturan atau pengetahuan.

Algoritma yang biasa dipakai dalam proses klasifikasi sangatlah banyak. Beberapa algoritma klasifikasi terbaik menurut Wu et al (2010) antara lain *Decision Tree (C4.5)*, *Support Vektor Machine (SVM)*, *K-Nearest Neighbour (k-NN)* dan *Naïve Bayes (NB)* (Wu et al. 2008).

2.3. Algoritma Decision Tree (C4.5)

Salah satu metode klasifikasi dan prediksi yang terkuat dan terkenal disebut Decision Tree (Kusrini & Luthfi 2009). *Decision Tree* berguna untuk mengeksploitasi data dan akan menemukan hubungan yang tersembunyi antara jumlah variabel input dengan variabel target (Kusrini & Luthfi 2009)(Witten et al. 2011). Struktur pada *Decision Tree* dapat digunakan untuk membagi kumpulan dataset yang besar menjadi dataset kecil, rangkain pembagiannya adalah anggota himpunan yang mirip satu dengan yang lain (Berry & Linoff 2004). *Decision Tree* lebih mengarah kepada perhitungan probabilitas dari tiap-tiap *record* terhadap kategori yang ada untuk mengklasifikasikan *record* dengan kelompok yang sama dalam satu kelas (Kusrini & Luthfi 2009).

Banyak algoritma *Decision Tree* diantaranya: *ID3*, *CART* dan *C4.5* (Larose 2005), untuk algoritma *C4.5* merupakan perbaikan/pengembangan dari algoritma *ID3* (Han & Kamber 2006)(Prasetyo 2014). Perbaikan pada algoritma *C4.5* adalah fitur dengan tipe numerik dapat ditangani dengan baik. Selain itu algoritma *C4.5* menggunakan pemecahan node pada pohon yang diinduksi menggunakan kriteria gain untuk menentukan fitur (Han & Kamber 2006)(Prasetyo 2014). Secara umum algoritma *C4.5* untuk membangun pohon keputusan, adalah (Prasetyo 2012):

1. Pilih atribut sebagai akar
2. Buat cabang untuk tiap-tiap nilai
3. Bagi kasus dalam cabang

4. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama

Untuk memilih akar disarankan untuk memilih nilai *gain* yang tinggi dari atribut yang ada. Untuk menghitung *gain* menggunakan persamaan 1, berikut (Prasetyo 2012):

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

Keterangan:

- S : himpunan kasus
- A : atribut
- n : jumlah partisi atribut A
- |S_i| : jumlah kasus pada partisi ke-i
- |S| : jumlah kasus dalam S

Sementara itu, penghitungan nilai entropi dapat dilihat pada persamaan 2, berikut:

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i \quad (2)$$

Keterangan:

- S : himpunan kasus
- A : fitur
- N : jumlah partisi S
- p_i : proporsi dari S_i terhadap S

2.4. Cross validation

Cross validation merupakan sebuah tindakan pembuktian dari sebuah metode atau performa suatu algoritma. Dalam proses pengujian data

mining yang paling banyak digunakan adalah *cross validation*. *Cross validation* merupakan pembuktian dengan membagi data sebagian sebagai data training dan sebagian yang lain sebagai data testing dengan komposisi tertentu. Pembagian paling banyak digunakan dalam penelitian klasifikasi data mining adalah membagi data secara acak menjadi 10 bagian. Satu bagian menjadi data testing dan 9 bagian dijadikan data training. Validasi yang seperti ini disebut juga dengan *10 folds cross validation* (Witten et al. 2011).

2.5. Confusion Matrix

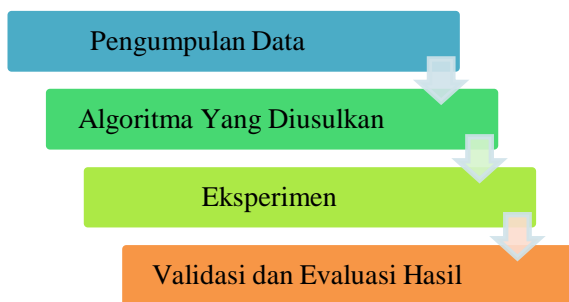
Confusion Matrix merupakan sebuah hasil evaluasi dari sebuah klasifikasi data mining yang diwujudkan dalam sebuah tabel (Gorunescu 2011). *Confusion Matrix* berisi tentang perhitungan jumlah objek data testing yang diprediksikan kedalam sebuah kelas dengan klasifikasi yang sebenarnya. Bentuk *confusion matrix* secara umum dapat dilihat pada tabel 2. Dalam *confusion matrix* terdapat total *record* yang dipakai dalam dataset baik yang diprediksikan kedalam kelas positif ataupun negatif. *Tupel/record* dengan prediksi klasifikasi positif, prediksi klasifikasi negatif, serta kesalahan dalam klasifikasi dapat terlihat dalam tabel 2.

Tabel 2. Confusion Matrix

Classification		Predicted class	
		Class: YES	Class: NO
Observed class	Class: YES	A (True Positive – TP)	D (False Negative – FN)
	Class: NO	C (False Positive – FP)	D (True Negative – TN)

3. METODE PENELITIAN

Kerangka pemikiran penelitian secara keseluruhan dapat dilihat pada gambar 2 berikut ini:



Gambar 2. Kerangka Penelitian

1. Pengumpulan Data

Pada tahap pengumpulan data, dataset status mahasiswa diambil dari P3SDI STMIK Widya Pratama. Jumlah dataset status mahasiswa yang digunakan untuk penelitian ini sebesar 632 *record* dengan 12 atribut, terlihat pada tabel 1. 11 atribut untuk Prodi, Jenis Kelamin, Umur, Saat Masuk, Domisili, Marital, Status Kerja, Asal, Biaya, Pekerjaan Ortu, Ipk, Status Skripsi dan Kelas. Sedangkan 1 atribut digunakan untuk label yaitu atribut Status Mahasiswa, terlihat pada gambar 3.

NO	PRODI	JENIS KELAMIN	UMUR SAAT MASUK	DOMISILI	MARITAL	STATUS KERJA	ASAL BIAYA	PEKERJAAN ORTU	IPK	STATUS SKRIPSI	KELAS	STATUS MHS
1	MANAJEMEN INFORMATIKA	L	21	LUAR KOTA	TIDAK MENIKAH	BELUM BEKERJA	ORANG TUA / WALI	PNS	3,03	TIDAK SKRIPSI	PAGI	AKTIF
2	MANAJEMEN INFORMATIKA	L	20	DALAM KOTA	TIDAK MENIKAH	BEKERJA	ORANG TUA / WALI	WIRAUSAHA	3,09	TIDAK SKRIPSI	MALAM	AKTIF
3	MANAJEMEN INFORMATIKA	L	22	DALAM KOTA	TIDAK MENIKAH	BEKERJA	ORANG TUA / WALI	PNS	2,85	TIDAK SKRIPSI	MALAM	AKTIF
4	MANAJEMEN INFORMATIKA	L	28	LUAR KOTA	TIDAK MENIKAH	BELUM BEKERJA	SENDIRI	WIRAUSAHA	3,70	TIDAK SKRIPSI	MALAM	TIDAK AKTIF
5	MANAJEMEN INFORMATIKA	L	29	DALAM KOTA	TIDAK MENIKAH	BEKERJA	ORANG TUA / WALI	PNS	2,28	TIDAK SKRIPSI	MALAM	AKTIF
6	KOMPUTERISASI AKUTANSI	L	24	LUAR KOTA	TIDAK MENIKAH	BELUM BEKERJA	ORANG TUA / WALI	SWASTA	2,94	TIDAK SKRIPSI	PAGI	AKTIF
7	KOMPUTERISASI AKUTANSI	P	21	LUAR KOTA	TIDAK MENIKAH	BELUM BEKERJA	ORANG TUA / WALI	SWASTA	3,71	TIDAK SKRIPSI	PAGI	AKTIF
8	KOMPUTERISASI AKUTANSI	P	31	LUAR KOTA	TIDAK MENIKAH	BEKERJA	SENDIRI	PNS	3,60	TIDAK SKRIPSI	MALAM	AKTIF
9	KOMPUTERISASI AKUTANSI	L	25	LUAR KOTA	TIDAK MENIKAH	BELUM BEKERJA	SENDIRI	WIRAUSAHA	2,78	TIDAK SKRIPSI	PAGI	AKTIF
10	KOMPUTERISASI AKUTANSI	P	19	LUAR KOTA	TIDAK MENIKAH	BELUM BEKERJA	BEASISWA	WIRAUSAHA	3,48	TIDAK SKRIPSI	PAGI	AKTIF
11	SISTEM INFORMASI	L	21	DALAM KOTA	TIDAK MENIKAH	BEKERJA	SENDIRI	PNS	2,48	TIDAK SKRIPSI	MALAM	AKTIF
12	SISTEM INFORMASI	L	19	LUAR KOTA	TIDAK MENIKAH	BEKERJA	SENDIRI	SWASTA	0,82	TIDAK SKRIPSI	MALAM	TIDAK AKTIF
13	SISTEM INFORMASI	P	19	LUAR KOTA	TIDAK MENIKAH	BEKERJA	BEASISWA	SWASTA	3,84	TIDAK SKRIPSI	MALAM	AKTIF
14	SISTEM INFORMASI	P	19	DALAM KOTA	TIDAK MENIKAH	BELUM BEKERJA	ORANG TUA / WALI	WIRAUSAHA	3,42	TIDAK SKRIPSI	PAGI	AKTIF
15	SISTEM INFORMASI	L	20	DALAM KOTA	TIDAK MENIKAH	BEKERJA	SENDIRI	PNS	3,49	TIDAK SKRIPSI	MALAM	AKTIF
16	TEKNIK INFORMATIKA	L	19	LUAR KOTA	TIDAK MENIKAH	BEKERJA	BEASISWA	WIRAUSAHA	3,68	TIDAK SKRIPSI	PAGI	AKTIF
17	TEKNIK INFORMATIKA	L	22	DALAM KOTA	TIDAK MENIKAH	BEKERJA	ORANG TUA / WALI	LAIN-LAIN	3,52	TIDAK SKRIPSI	MALAM	AKTIF
18	TEKNIK INFORMATIKA	P	24	LUAR KOTA	TIDAK MENIKAH	BEKERJA	ORANG TUA / WALI	SWASTA	3,47	TIDAK SKRIPSI	MALAM	AKTIF
19	TEKNIK INFORMATIKA	L	22	DALAM KOTA	TIDAK MENIKAH	BELUM BEKERJA	SENDIRI	WIRAUSAHA	3,54	TIDAK SKRIPSI	MALAM	AKTIF
20	TEKNIK INFORMATIKA	P	21	LUAR KOTA	TIDAK MENIKAH	BELUM BEKERJA	BEASISWA	SWASTA	0,64	TIDAK SKRIPSI	PAGI	TIDAK AKTIF

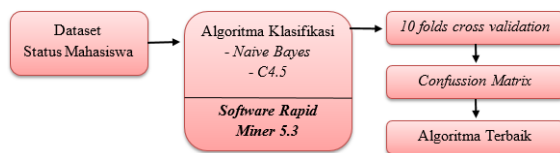
Gambar 3. Dataset Status Mahasiswa

2. Algoritma Yang diusulkan

Algoritma yang digunakan dalam penelitian tentang status mahasiswa adalah algoritma klasifikasi data mining menggunakan *decision tree* dengan algoritma *C4.5*. Algoritma *C4.5* akan dibandingkan dengan algoritma *Naive Bayes* yang sudah pernah dilakukan dalam penelitian sebelumnya.

3. Eksperimen

Untuk melakukan pengujian algoritma, dilakukan eksperimen menggunakan software Rapid Miner 5.3. Dalam tahapan eksperimen algoritma *C4.5* akan digunakan untuk mengolah dataset status mahasiswa dan akan dibandingkan dengan menggunakan algoritma *Naive Bayes*. Apakah algoritma *C4.5* akurasi lebih baik dari algoritma *Naive Bayes*.



Gambar 4. Tahapan Eksperimen

4. Validasi dan Evaluasi Hasil

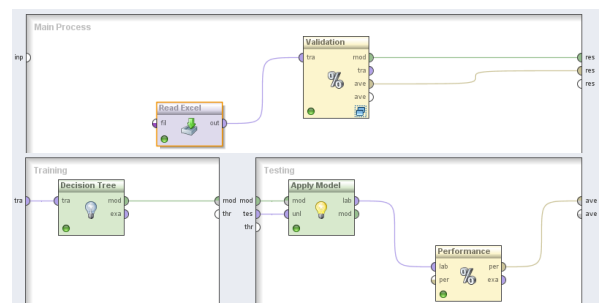
Pada proses validasi menggunakan teknik *10 folds cross validation*. Dalam melakukan evaluasi digunakan tabel *confussion matrix*, digunakan untuk membandingkan akurasi dari pengolahan data antara algoritma *C4.5* dengan algoritma *Naive Bayes*.

4. HASIL DAN PEMBAHASAN

3.1. Perhitungan Akurasi Algoritma

Dalam menghitung hasil akurasi algoritma yang diusulkan menggunakan bantuan software Rapid Miner 5.3. Dalam melakukan perhitungan akurasi algoritma menggunakan teknik *10 fold cross validation*, pada proses ini dataset akan dibagi menjadi 10 bagian secara acak. Untuk pembagiannya adalah 9 bagian digunakan untuk data training dan 1 bagian digunakan untuk data testing.

Untuk menghitung tingkat akurasi menggunakan tabel *confussion matrix*. Tabel ini digunakan untuk mengukur tingkat akurasi dari algoritma yang digunakan. Dari 10 kali percobaan yang sudah dilakukan tingkat akurasi dihitung secara rerata. Berikut perhitungan algoritma *C4.5* dengan menggunakan bantuan software Rapid Miner 5.3, seperti terlihat pada gambar 5.

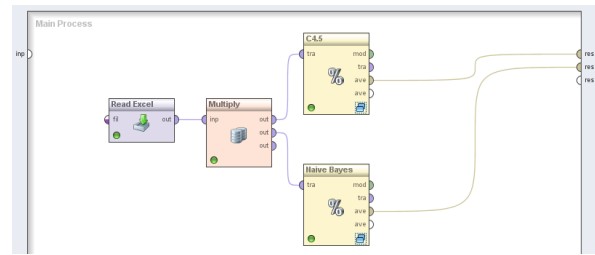


Gambar 5. Perhitungan Algoritma *Decision Tree (C4.5)*

Untuk perhitungan algoritma *Naive Bayes* dilakukan dengan tahap yang sama dengan perhitungan pada algoritma *C4.5*.

3.2. Hasil Akurasi Algoritma

Untuk menghitung komparasi algoritma menggabungkan semua algoritma kedalam satu lembar kerja untuk mengetahui algoritma apa yang memiliki tingkat akurasi terbaik. Berikut lembar kerja dalam satu dataset yang sama dari algoritma yang digunakan untuk komparasi, seperti terlihat pada gambar 5.



Gambar 5. Lembar Kerja Komparasi Algoritma

Dari hasil yang sudah dilakukan menggunakan software *Rapid Miner 5.3* didapatkan hasil akurasi dari algoritma *Decision Tree (C4.5)* dan *Naive Bayes*. Hasil akurasinya dapat dilihat pada gambar 6 dan 7.

accuracy: 93.05% +/- 3.30% (mikro: 93.04%)			
	true AKTIF	true TIDAK AKTIF	class precision
pred. AKTIF	545	28	95.11%
pred. TIDAK AKTIF	16	43	72.88%
class recall	97.15%	60.56%	

Gambar 6. Akurasi Algoritma *Decision Tree (C4.5)*

accuracy: 92.72% +/- 2.26% (mikro: 92.72%)			
	true AKTIF	true TIDAK AKTIF	class precision
pred. AKTIF	538	23	95.90%
pred. TIDAK AKTIF	23	48	67.61%
class recall	95.90%	67.61%	

Gambar 7. Akurasi Algoritma *Naive Bayes*

Setelah melakukan penelitian tingkat akurasi pada dataset status mahasiswa didapatkan hasil keseluruhan algoritma. Tahap selanjutnya adalah membandingkan tingkat akurasi dari algoritma *Decision Tree (C4.5)* dan Algoritma *Naive Bayes*, seperti terlihat di tabel 3.

Tabel 3. Hasil Komparasi Akurasi Algoritma

Algoritma	Tingkat Akurasi	Micro
<i>Decision Tree C4.5</i>	93,05 %	93,04 %
<i>Naive Bayes</i>	92,72 %	92,72 %

5. KESIMPULAN DAN SARAN

5.1. Kesimpulan

Dari hasil penelitian yang sudah dilakukan, maka dapat disimpulkan, Dalam penelitian tentang status mahasiswa menggunakan dataset *privat* yaitu dataset status mahasiswa yang diambil dari P3SDI STMIK Widya Pratama. Dataset status mahasiswa sudah pernah dilakukan menggunakan algoritma *Naive Bayes*.

Dalam penelitan status mahasiswa diketahui bahwa untuk klasifikasi status mahasiswa dengan menggunakan algoritma *Decision Tree (C4.5)* merupakan algoritma terbaik. Algoritma *Decision Tree (C4.5)* tingkat akurasi sebesar 93,05 %, sedangkan algoritma *Naive Bayes* sebesar 92,72 %.

5.2. Saran

Penelitian ini menggunakan software *Rapid Miner 5.3* untuk melakukan proses perhitungan dan pembuktian akurasi algoritma. Dalam penelitian selanjutnya diharapkan dalam melakukan penelitian bisa menggunakan software buatan sendiri untuk mendeteksi status mahasiswa. Untuk menentukan tingkat akurasi bisa juga melakukan komparasi antara algoritma klasifikasi dengan tambahan algoritma *feature selection* atau yang lainnya.

DAFTAR PUSTAKA

- Ashari, A., Paryudi, I. & Tjoa, A.M., 2013. Performance Comparison between Naïve Bayes , Decision Tree and k-Nearest Neighbor in Searching Alternative Design in an Energy Simulation Tool. , 4(11), pp.33–39.
- Berry, M.J.A. & Linoff, G.S., 2004. *Data Mining Techniques for Marketing, Sales and Customer Relationship Management* Second Edi., Simultaneously in Canada.
- Christobel, A. & Sivaprakasam, D., 2011. An Empirical Comparison of Data Mining Classification Methods. , 3(2), pp.24–28.
- Gorunescu, F., 2011. *Data Mining: Concept, Models and Techniques* vol 12., Berlin: Heidelberg: Springer Berlin Heidelberg.
- Han, J. & Kamber, M., 2006. *Data mining: Concepts and Techniques, Second Edition*, Diane Cerra. Available at: http://ccs1.hnue.edu.vn/hungtd/DM2012/DataMining_BOOK.pdf.
- Hastuti, K., 2012. Analisis Komparasi Algoritma Klasifikasi Data Mining untuk Prediksi Mahasiswa Non-aktif. , 2012(Semantik 2012), pp.241–249.
- Kotsiantis, S.B., Pierrakeas, C.J. & Pintelas, P.E., 2003. Preventing Student Dropout in Distance Learning Using Machine Learning Techniques. , pp.267–274.
- Kusrini & Luthfi, E.T., 2009. *Algoritma Data Mining* Ed. 1. T. A. Prabawati, ed., Yogyakarta: Andi Offset.
- Larose, D.T., 2005. *Discovering Knowledge in Data an Introduction to Data Mining*, Maimon, O. & Rokach, L., 2010. *Data Mining and Knowledge Discovery Handbook*, Available at: http://dx.doi.org/10.1007/0-387-25465-x_2%5Cnhttp://link.springer.com/content/pdf/10.1007/978-0-387-09823-4.pdf.
- Marquez-Vera, C., Romero, C. & Ventura, S., 2011. Predicting School Failure Using Data Mining. *Proceedings of the 4th International Conference on Educational Data Mining*, (December). Available at: http://educationaldatamining.org/EDM2011/wp-content/uploads/proc/edm2011_paper11_short_Marquez-Vera.pdf.
- Prasetyo, E., 2012. *Data Mining, Konsep dan Aplikasi menggunakan Matlab*, Yogyakarta: Andi Offset.
- Prasetyo, E., 2014. *Data Mining: Mengolah Data Menjadi Informasi Menggunakan Matlab*, Ragab, A.H.M. et al., 2014. A Comparative Analysis of Classification Algorithms for Students College Enrollment Approval Using Data Mining. *Proceedings of the 2014 Workshop on Interaction Design in Educational Environments - IDEE '14*, pp.106–113.
- Royanti, N.I., 2015. *Peningkatan Akurasi Naive Bayes Dengan Seleksi Fitur Gain Ratio Untuk Klasifikasi Status Mahasiswa*. Dian Nuswantoro.
- Santosa, B., 2007. *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*, Yogyakarta: Graha Ilmu.
- Susanto, S. & Suryadi, D., 2010. *Pengantar Data Mining Menggali Pengetahuan Dari Bongkahan Data* Ed. 1., Yogyakarta: Andi Offset.
- Undang-Undang, 2012. Undang-Undang Republik Indonesia Nomor 12 Tahun 2012 Tentang Pendidikan Tinggi. <http://www.polsri.ac.id/>, 8, pp.130–183.
- Witten, I.H., Frank, E. & Hall, M. a., 2011. *Data Mining: Practical Machine Learning Tools and Techniques* Third Edit., USA: Morgan Kaufmann Publishers.
- Wu, X. et al., 2008. *Top 10 algorithms in data mining*,

