PENERAPAN METODE *DECISION TREE* ALGORITMA C4.5 UNTUK KLASIFIKASI MAHASISWA BERPRESTASI

Nadiya Hijriana[1)] dan Riadhul Muttaqin[1)]

[1]*Program Studi Teknik Informatika Universitas Islam Kalimantan, Banjarmasin*
E-mail: nadyahijriana@yahoo.com

ABSTRACT

Informatic Engineering is a field of study from Information Technology Faculty (FTI) at Islamic University of Kalimantan (UNISKA). Informatic Engineering was established in 2007 and it has around 2441 students until now included by student who has graduated and student who still studying. Fact of so many students in Informatic Engineering can make a lot of data that it can used as strategic resource for Faculty to manage and monitoring of academic activity. This research using Decision Tree Method from Algorithm C4.5 and RapidMiner 5.3 as software for making classification of achievement students in field of study Informatic Engineering at Islamic University of Kalimantan. The result can be used as tool for monitoring of academic activity and it can help Faculty to improve the quality of their students.

Keywords: *Classification, Data, Student, Algoritma C4.5*

PENDAHULUAN

Universitas Islam Kalimantan (UNISKA) Muhammad Arsyad Al Banjary adalah PTS tertua dan terbesar di Kalimantan Selatan, didirikan oleh yayasan UNISKA berdasarkan akte Notaris Bachtiar Nomor 6 tanggal 7 Juli 1981. Saat ini UNISKA MAAB telah memiliki Program PASCA SARJANA, 8 (Delapan) Fakultas, 12 (Dua Belas) Program Studi S-1 dan 1 (Satu) Program Studi D-3. Salah satu dari kedua belas Program Studi S-1 tersebut adalah Program Studi Teknik Informatika yang merupakan bagian dari Fakultas Teknologi Informasi (FTI). Sejak berdirinya Prodi Teknik Informatika sampai sekarang, Prodi yang didirikan mulai tahun 2007 ini telah memiliki mahasiswa sebanyak 2441 orang baik yang lulus maupun belum lulus. Bagi sebuah Prodi yang baru berumur kurang lebih 9 tahun dengan jenjang perkuliahan Strata Satu atau S-1, jumlah tersebut merupakan suatu keberhasilan yang membanggakan.

Dengan membludaknya jumlah mahasiswa yang ada di FTI, maka data-data akademik yang dimiliki juga semakin bertambah banyak. Apabila tidak diolah dengan baik dan benar, hanya akan menjadi tumpukan data yang tidak berguna dan tidak bermanfaat. Padahal data tersebut bisa dijadikan sebagai sumber data strategis yang sangat bermanfaat bagi Fakultas Teknologi Informasi.

Berdasarkan permasalahan tersebut, maka peneliti akan mengklasifikasikan mahasiswa Prodi Teknik Informatika yang berprestasi dengan menggunakan metode *Decision Tree* Algoritma C4.5. Tujuan dari penelitian ini adalah untuk mengklasifikasi mahasiswa berprestasi yang ada di Prodi Teknik Informatika FTI UNISKA dimana hasil penelitian tersebut dapat digunakan sebagai alat bantu bagi Fakultas untuk pengembangan akademis dalam hal pembimbingan dan pengarahan yang lebih intensif bagi mahasiswa IPK-nya kurang memuaskan.

METODE PENELITIAN

Perancangan Penelitian

Pengertian penelitian dalam akademik yaitu digunakan untuk mengacu pada aktivitas yang rajin dan penyelidikan sistematis atau investigasi di suatu daerah, dengan tujuan menemukan atau merevisi fakta, teori, aplikasi dan tujuannya adalah untuk menemukan dan menyebarkan pengetahuan baru (Berndtssom et. al. 2008).

Metode penelitian yang dilakukan adalah metode penelitian eksperimen, dengan tahapan sebagai berikut:

1. Pengumpulan Data
   Pada tahap ini ditentukan data yang akan diproses. Mencari data yang tersedia, memperoleh data tambahan yang dibutuhkan, mengintegrasikan semua datake dalam data set, termasuk variabel yang diperlukan dalam proses.
2. Pengolahan Data Awal (*Data Pre-processing)*
   Di tahap ini dilakukan penyeleksian data, data dibersihkan dan ditransformasikan ke bentuk yang diinginkan sehingga dapat dilakukan persiapan dalam pembuatan model.
3. Model/Metode yang Diusulkan (*Proposed Model/Method*)
   Pada tahap ini data dianalisis, dikelompokkan variabel mana yang berhubungan dengan satu sama lainnya. Setelah data dianalisis lalu diterapkan model-model yang sesuai dengan jenis data. Pembagian data ke dalam data latihan (*training data*) dan data uji (*testing data*) juga diperlukan untuk pembuatan model.
4. Eksperimen dan Pengujian Metode (*Method Test and Experiment*)
   Pada tahap ini model yang diusulkan akan diuji untuk melihat hasil berupa rule yang akan dimanfaatkan dalam pengambilan keputusan.
5. Evaluasi dan Validasi Hasil (*Result Evaluation and Validation*)
   Pada tahap ini dilakukan evaluasi terhadap model yang ditetapkan untuk mengetahui tingkat keakurasian model.

Metode Pengumpulan Data

Data sekunder adalah data yang diperoleh secara tidak langsung bersumber dari dokumentasi, literatur, buku, jurnal dan informasi lainnya yang ada hubungannya dengan masalah yang diteliti. Data sekunder pada penelitian ini adalah : buku-buku, jurnal tentang metode *Decision Tree* Algoritma C4.5 dan data mining serta data mahasiswa Prodi Teknik Informatika FTI UNISKA Banjarmasin tahun akademik 2014/2015. Sedangkan Data primer adalah data yang diperoleh dari hasil penelitian. Data primer dalam penelitian ini adalah data hasil uji dengan menggunakan metode *Decision Tree* Algoritma C4.5.

Data yang dibutuhkan dalam penelitian ini adalah data mahasiswa Program Studi Teknik Informatika Angkatan 2012, dimana atribut yang nantinya akan dipakai antara lain NPM, Status Mahasiswa, Indeks Prestasi Semester 5 dan Indeks Prestasi Semester 6.

Pengolahan Data Awal

Data awal yang didapatkan dari BAK UNISKA MAB Banjarmasin yaitu data mahasiswa seluruh angkatan dengan atribut NIM, Nama, Program Studi, Status Mahasiswa, Indeks Prestasi Semester (IPS), SKS per Semester, Indeks Prestasi Kumulatif, SKS Total. Dari data awal tersebut dilakukan pengolahan data lagi sehingga didapatkan data mahasiswa Angkatan 2012 dan data Indeks Prestasi untuk semester 5 dan semester 6. Selanjutnya dilakukan pengolahan data untuk menggabungkan kedua semester dan memastikan status mahasiswa akan diuji berada pada status "Aktif" baik untuk semester 5 maupun untuk semester 6 sehingga hasil akhir berupa atribut yang akan digunakan dalam pengujian adalah NPM, Status Mahasiswa, Indeks Prestasi Semester 5 dan Indeks Prestasi Semester 6 dimana jumlah mahasiswa sebanyak 444 orang.

Metode yang diusulkan

Metode yang diusulkan untuk menentukan klasifikasi mahasiswa yang berprestasi adalah metode *Decision Tree* Algoritma C4.5 dengan menggunakan *software* RapidMiner 5.3 untuk pembuatan analisa dan pengujian model.

Tahapan Algoritma Decision Tree C4.5:

1) Menyiapkan data training
2) Menentukan akar dari pohon.
3) Hitung nilai Gain:

$$Entropy(S) = \sum_{i=1}^{n} - pi * \log_2 pi \qquad (1)$$

4) Ulangi langkah ke-2 hingga semua tupel terpartisi

$$Gain(S, A) = S - \sum_{i=1}^{n} \frac{|S_i|}{|S|} * S_i \qquad (2)$$

5) Proses partisi pohon keputusan akan berhenti saat semua tupel dalam node N mendapat kelas yang sama dan atau tidak ada atribut di dalam tupel yang dipartisi lagi dan atau tidak ada tupel di dalam cabang yang kosong.

Eksperimen dan Pengujian Metode

Metode yang telah dikembangkan dalam penelitian ini akan diterapkan pada data mahasiswa Prodi Teknik Informatika UNISKA Banjarmasin

Angkatan 2012 dengan sampel data pada tahun akademik 2014/2015 melalui suatu simulasi menggunakan software RapidMiner 5.3. Data Sampel terdiri dari atribut NPM, Status Mahasiswa (Aktif), IPK Semester 5, IPK Semester 6 dimana jumlah mahasiswa sebanyak 444 orang. Sebanyak 90 % data akan digunakan untuk membangun struktur pohon keputusan melalui metode *Decision Tree* Algoritma C4.5. Sedangkan 10 % lainnya akan digunakan sebagai data uji.

Evaluasi dan Validasi Hasil

Validasi hasil penelitian dilakukan dengan mengambil sampel data mahasiswa dari Angkatan 2012 yang mempunyai status Aktif. Hasil pengolahan data awal didapatkan sebanyak 444 mahasiswa aktif. Data diuji dengan menggunakan metode *Decision Tree* Algoritma C4.5 sebanyak 9 kali, dimana field data yang akan diuji adalah NPM, Status Mahasiswa (Aktif), IPK Semester 5 dan IPK Semester 6. Selain itu, untuk membentuk pohon keputusan maka atribut IPK Semester 5 dan IPK Semester 6 perlu diklasifikasi menjadi:

Tabel 1. Klasifikasi Nilai

No	IPK Semester	Klasifikasi
1	>= 3,51	Dengan Pujian
2	>= 2,76	Sangat Memuaskan
3	>= 2,0	Memuaskan
4	< 2,0	Kurang

Evaluasi dilakukan dengan menganalisa hasil klasifikasi dari penggunaan metode algoritma *Decision Tree* C4.5, yaitu membandingkan nilai Akurasi, nilai *Precision*, dan nilai *Recall*. Perhitungan akurasi dengan tabel *confusion matrix* adalah sebagai berikut (sumber):

$$Akurasi = \frac{F\,11+F\,00}{F\,11+F\,10+F\,01+F00} \qquad (3)$$

Penjelasan tentang pengukuran *Precision* dan *Recall* dapat dilihat pada Tabel 2 dan perhitungan berikut:

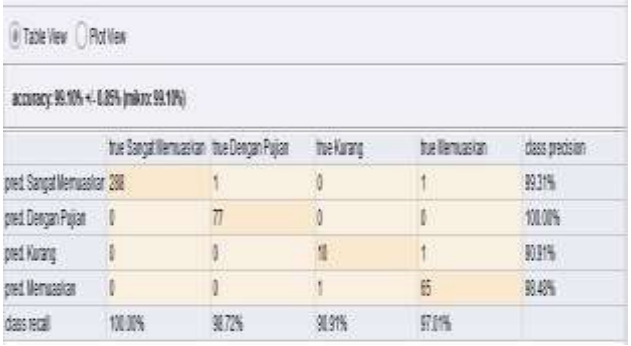
Tabel 2. Perhitungan Precision dan Recall

	Relevant	Not Relevant
Retrieved	A	B
Not Retrieved	C	D

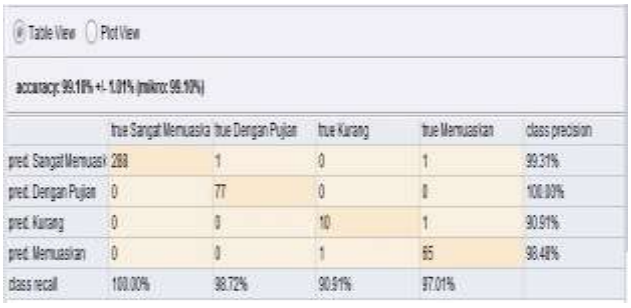
HASIL DAN PEMBAHASAN

Hasil Eksperimen dan Pengujian Model/Metode

Nilai akurasi dalam penelitian ini didapatkan dengan melakukan pengujian terhadap sampel data mahasiswa aktif Angkatan 2012 sebanyak 444 orang. Atribut yang digunakan antara lain NPM, Status Aktif, IPK Semester 5 dan IPK Semester 6. Pengujian data dilakukan dengan menggunakan *software* RapidMiner 5.3 dimana uji data dilakukan sebanyak sembilan kali. Dari kesembilan pengujian tersebut, dapat diketahui bahwa pengujian Validasi K-6 dan Validasi K-9 memiliki bobot nilai akurasi paling tinggi yaitu 99.10% dengan +/- 1.01% dan Mikro 99.10%. Hasil pengujian dapat dilihat pada gambar berikut:



Gambar 1. Hasil Perhitungan Akurasi pada Validasi K-6



Gambar 2. Hasil Perhitungan Akurasi pada Validasi K-9

Evaluasi dan Validasi Hasil

Setelah melakukan sembilan kali pengujian data menggunakan RapidMiner 5.3, selanjutnya dilakukan analisa data untuk mencari nilai akurasi yang paling tinggi. Hasilnya adalah bahwa nilai akurasi tertinggi diperoleh melalui pengujian Validasi K-6 dan Validasi K-9, yakni sebesar 99.10%. Dengan demikian maka klasifikasi

keakuratan tes diagnostiknya termasuk dalam kategori sangat baik.

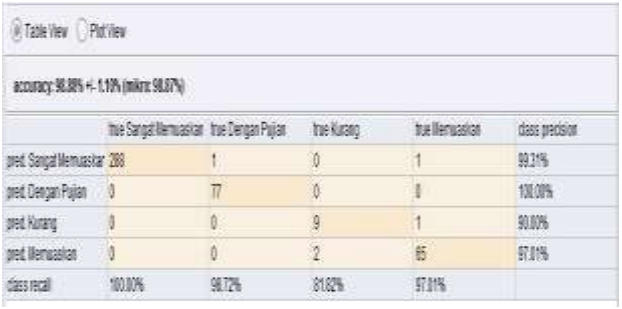
Berikut hasil evaluasi pengukuran data *Performance Vector* dengan membandingkan nilai *Precision* dan nilai *Recall* seperti pada tabel dibawah ini:

Tabel 3. Performance Vector C4.5 dari Validasi K-6

Klasifikasi (IPK)	Jumlah Mahasiswa (orang)	*Class Recall*	*Class Precision (%)*
Dengan Pujian	78	98,72	100,00
Sangat memuaskan	288	100,00	99,31
Memuaskan	67	97,01	98,48
Kurang	11	90,91	90,91

Analisis Evaluasi dan Validasi Model

Dari hasil pengujian yang sudah dilakukan, dapat dilihat bahwa pengujian Validasi K-6 dan Validasi K-9 memiliki nilai akurasi paling tinggi yakni sebesar 99.10% dengan +/- 1.01% dan Mikro 99.10% seperti yang tampak pada gambar berikut:



Gambar 3. Hasil Perhitungan Akurasi pada Validasi K-6 beserta nilai *Class Recall* dan nilai *Class Precision*

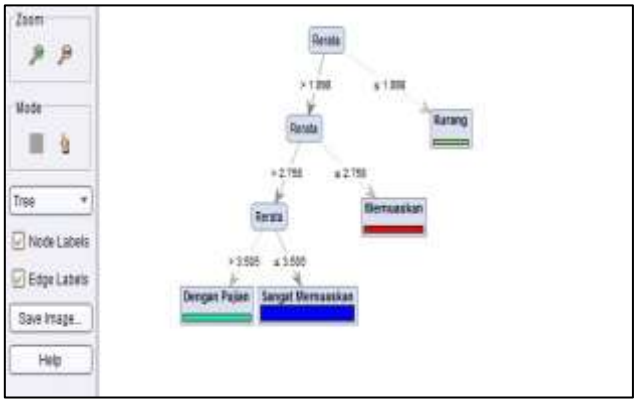
Dari gambar 3 dapat dilihat bahwa IPK mahasiswa dengan predikat "Sangat Memuaskan" merupakan jumlah paling banyak yakni 288 orang, diikuti oleh jumlah mahasiswa yang IPK-nya mencapai predikat "Dengan Pujian" sebanyak 78 orang. Urutan ketiga adalah mahasiswa yang IPK-nya mencapai predikat "Memuaskan" dengan jumlah mahasiswa adalah 67 orang. Urutan terakhir adalah 11 mahasiswa mendapat IPK dengan predikat "Kurang".

Berdasarkan hasil pengujian yang telah dilakukan dengan menerapkan metode *Decision Tree* Algoritma C4.5 dan menggunakan software RapidMiner 5.3, dapat diketahui bahwa dari sembilan kali proses pengujian tersebut, didapatlkan hasil nilai akurasi tertinggi melalui pengujian Validasi K-6 dan Validasi K-9, yakni sebesar 99.10%. Dari nilai akurasi tertinggi tersebut didapatkan tabel klasifikasi Indeks Prestasi Kumulatif (IPK) mahasiswa sebagai berikut:

Tabel 4. Hasil Klasifikasi IPK Mahasiswa Berdasarkan Urutan Jumlah Terbanyak

No	Range IPK	Klasifikasi Nilai (IPK)	Jumlah Mahasiswa
1	≥ 2,76	Sangat Memuaskan	288 orang
2	≥ 3,51	Dengan Pujian	78 orang
3	≥ 2,0	Memuaskan	67 orang
4	< 2,0	Kurang	11 orang

Selain itu validasi dari tingginya hasil nilai akurasi diatas dibuktikan pula menggunakan klasifikasi pohon keputusan (*Decision Tree*), dimana hasilnya adalah sebagai berikut:



Gambar 4. Hasil pengujian pohon keputusan ((*Decision Tree*)

Dari gambar 4 terlihat klasifikasi kelompok nilai Indeks Prestasi Kumulatif (IPK) mahasiswa, dimana kelompok IPK dengan predikat "Sangat Memuaskan" menempati urutan pertama atau kelompok yang terbesar, diikuti oleh kelompok IPK dengan predikat "Dengan Pujian". Urutan ketiga dan keempat adalah kelompok IPK dengan predikat "Memuaskan" dan kelompok IPK dengan predikat "Kurang".

Dengan demikian, dari hasil pengujian terlihat bahwa klasifikasi keakuratan tes diagnostiknya termasuk dalam kategori sangat baik dimana hasil akurasi pengujian Validasi K-6 dan Validasi K-9 juga telah sesuai dengan pengujian pohon keputusan

(*Decision Tree*). Dan metode *Decision Tree* Algoritma C4.5 dapat diterapkan untuk memberikan solusi dalam permasalahan mengolah data untuk klasifikasi, dalam hal ini mengolah data akademik untuk mengklasifikasi mahasiswa berprestasi yang ada di program studi Teknik Informatika FTI UNISKA MAB Banjarmasin.

KESIMPULAN

Hasil klasifikasi dengan software Rapid Miner menggunakan metode *Decision Tree* Algoritma C4.5 menunjukkan bahwa data akademik dapat diolah berdasarkan atribut-atribut yang dibutuhkan menjadi data yang bermanfaat bagi Fakultas, dibuktikan dengan hasil klasifikasi yang terdapat pada penelitian ini. Hasil klasifikasi mahasiswa berprestasi pada Angkatan 2012 yang diperoleh tidak hanya dapat dijadikan sebagai acuan untuk pengolahan data lebih lanjut misalnya untuk pengolahan data penerima beasiswa, tetapi juga dapat dijadikan sebagai alat bantu bagi Fakultas untuk monitoring kegiatan akademik, salah satunya untuk memonitor mahasiswa yang masuk ke dalam klasifikasi mahasiswa yang berpotensi lulus dengan IPK berpredikat "Kurang" yang nantinya bisa menurunkan mutu lulusan sebuah Perguruan Tinggi.

DAFTAR PUSTAKA

Alimancon Sijabat. 2015. Penerapan Data Mining Untuk Pengolahan Data Siswa Dengan Metode Decision Tree (Studi Kasus Yayasan Perguruan Kristen Andreas). Majalah Ilmiah Informasi dan Teknologi Ilmiah (INTI) Vol.5 No.3.

Berndtssom, M., Hansson, J., Olsson, B., and Lundell, B. 2008. A Guide For Students In Computer Science And Information Systems. London. Springer.

Dwi Untari. 2014. Data Mining Untuk Menganalisa Prediksi Mahasiswa Berpotensi Non-Aktif Menggunakan Metode Decision Tree C4.5. Semarang. Universitas Dian Nuswantoro.

Fajar Astuti Hermawan. 2013. Data Mining. Yogyakarta. Penerbit Andi.

Indri Rahmayuni. 2014. Perbandingan Performansi Algoritma C4.5 dan Cart dalam Klasifikasi Data Nilai Mahasiswa Prodi Teknik Komputer Politeknik Negeri Padang. Jurnal TEKNOIF Vol.2 No.1.

Kusrini dan Emma Taufiq Luthfi. 2009. Algoritma Data Mining. Yogyakarta. Penerbit Andi

Kusrini, Sri Hartati, Retantyo Wardoyo dan Agus Harjoko. 2009. Perbandingan Metode Nearest Neighbor dan Algoritma C4.5 Untuk Menganalisis Kemungkinan Pengunduran Diri Calon Mahasiswa di STMIK AMIKOM Yogyakarta. Jurnal DASI Vol.10 No.1

Liliana Swastina. 2013. Penerapan Algoritma C4.5 Untuk Penentuan Jurusan Mahasiswa. Jurnal GEMA AKTUALITA Vol.2 No.1.

Lillyan Hadjaratie. 2014. Prediksi dan Pemetaan Data Mahasiswa Fakultas Teknik Menggunakan Pendekatan Data Mining. Gorontalo. Universitas Negeri Gorontalo.

Setyawan dan Yusuf Sulistyo, ST., M.Eng. 2014. Klasifikasi Prestasi Akademik Mahasiswa FKI UMS Menggunakan Metode Decision Tree. Solo. Universitas Muhammadiyah Surakarta.