

# Pemanfaatan Data Mining Untuk Prediksi Mahasiswa Peserta “Capping Day” di Universitas Klabat Menggunakan Metode C4.5

## Data Mining for Student Prediction Participant of "Capping Day" at Universitas Klabat Using the C4.5 Algorithm

**Green Ferry Mandias**

Program Studi Sistem Informasi, Universitas Klabat

e-mail: green@unklab.ac.id

### **Abstrak**

*Fakultas Keperawatan Universitas Klabat melalui hikmat yang ada menginginkan mahasiswa perawat yang berdedikasi. Fakultas keperawatan dalam falsafanya juga menginginkan pelayanan dengan menghasilkan perawat-perawat profesional yang kompeten baik knowledge, skill dan attitude. Tujuan penelitian ini yaitu memanfaatkan proses data mining untuk memprediksi peserta capping mahasiswa fakultas ilmu keperawatan di Universitas Klabat. Metode yang digunakan dalam penelitian ini yaitu model klasifikasi dengan algoritma C4.5. Hasil yang diperoleh yaitu berdasarkan model pohon keputusan algoritma C4.5 didapati bahwa 3 atribut yang menjadi penentu mahasiswa itu mengikuti capping tepat waktu ataupun tidak. Yang pertama adalah berdasarkan nilai gain yang paling tinggi yaitu nilai Keperawatan Dasar, bila nilai Keperawatan Dasar Fail, Excellent dan Low maka prediksi untuk mahasiswa itu “No”, tidak akan mengikuti capping tepat waktu. Bila nilai High maka mengikuti capping tepat waktu. Bila nilai Average maka dilihat pada nilai Psikologi Keperawatan bila nilainya Excellent High dan Fail maka prediksinya yaitu “Yes”. Bila nilainya Low maka prediksi adalah “No”. Bila nilai adalah Average dilihat pada nilai Anatomi dan Fisiologi. Bila nilai Anatomi dan Fisiologi Fail, Average, Low dan Excellent maka prediksinya adalah “No” dan bila nilai Average maka prediksinya adalah “Yes”.*

**Keywords : Data Mining, Klasifikasi, Algoritma C4.5, Capping Day, Universitas Klabat**

### **Abstract**

*The Faculty of Nursing of Universitas Klabat through the existing wisdom wants dedicated nursing students. Faculty of nursing in philosophy also want service by menghasilkan professional nurses' competent good knowledge, skill and attitude. The purpose of this research is to utilize data mining process to predict capping participants of nursing faculty students at Klabat University. The method used in this research is the classification model with C4.5 algorithm. The results obtained are based on the model decision tree algorithm C4.5 found that the 3 attributes that became the determinant of the students to follow the capping on time or not. The first is based on the highest gain value of the basic nursing value, if the value of Basic Fail Nursing, Excellent and Low then the predictions for the student "No", will not follow capping on time. When the value of High then follows the capping on time. If the value is Average then seen on the value of Psychology of Nursing if the value of Excellent High and Fail then the prediction is "Yes". If the value is Low then the prediction is "No". When the value is Average seen on the value of Anatomy and Physiology. If the value of Anatomy and Physiology Fail,*

*Average, Low and Excellent then the prediction is "No" and if the Average value then the prediction is "Yes".*

**Keywords :** *Data Mining, Klasifikasi, C4.5 Algorithm, Capping Day, Universitas Klabat*

## 1. PENDAHULUAN

Sejarah keperawatan yang panjang telah membawa profesi ini mengembangkan tradisinya sendiri. Dalam dunia pendidikan keperawatan di Indonesia, hampir semua sekolah keperawatan baik akademi maupun sekolah tinggi/universitas melakukan *Capping Day* setiap tahun. *Capping Day* adalah suatu seremonial yang dilakukan oleh mahasiswa baru yang telah menjalani perkuliahan selama kurang lebih satu semester, ketika acara berlangsung mereka akan mengucapkan janji mahasiswa, mendengarkan riwayat hidup *Florence Nightingale*, serta diberi pin dan kap (khusus perempuan). [1]

Universitas Klabat yang dibawah oleh Uni Konferens Indonesia Kawasan Timur, pada bulan Agustus 2006 telah menambah satu Fakultas lagi yaitu Fakultas Ilmu Keperawatan dengan maksud menyediakan tenaga perawat profesional yang kompeten dan berkualitas tinggi yang sanggup menyajikan pelayanan kesehatan, yang mencakup biopsikososio spiritual.

Fakultas Keperawatan Universitas Klabat melalui hikmat yang ada menginginkan mahasiswa perawat yang berdedikasi. Fakultas keperawatan dalam falsafahnya juga menginginkan pelayanan dengan menghasilkan perawat-perawat profesional yang kompeten baik *knowledge, skill* dan *attitude*. Visi dari fakultas ini yaitu menjadi program studi keperawatan swasta yang menghasilkan perawat profesional yang terbaik di Sulawesi Utara dalam penekanan khusus di bidang medikal-bedah dan komunitas. Dalam misinya fakultas ini ingin mendidik dan menyiapkan perawat profesional yang kompeten yang mengacu pada standar nasional dengan keunggulan dalam bidang keperawatan medikal-bedah dan komunitas.

Sistem pendidikan telah dibuat sedemikian agar mahasiswa dapat menyelesaikan program studi yang telah ditentukan dan memberikan mahasiswa kesempatan agar dapat menentukan mata ajar yang sesuai. Sistem pendidikan yang telah dibuat bertujuan agar fakultas dapat mengevaluasi kemajuan belajar mahasiswa dengan sebaik-baiknya. Walau demikian ada ada banyak mahasiswa yang merupakan mahasiswa yang seharusnya mengikuti *capping day*, tidak dapat mengikuti kegiatan tersebut tepat pada waktunya. Untuk itu pada penelitian ini akan dilakukan pemanfaatan data mining dalam memprediksi mahasiswa yang akan mengikuti *capping day* di fakultas Keperawatan Universitas Klabat menggunakan metode C4.5. Secara sederhana data mining adalah penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar [2]. Data mining juga disebut sebagai serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari suatu kumpulan data [3]. Data mining, sering juga disebut sebagai *knowledge discovery in database (KDD)*. KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar [4]. Adapun tujuan penelitian ini yaitu memanfaatkan proses data mining untuk memprediksi peserta *capping day* mahasiswa Fakultas Ilmu Keperawatan di Universitas Klabat menggunakan metode C4.5. C4.5 merupakan sebuah algoritma *decision tree*. *Decision tree* berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target [5]

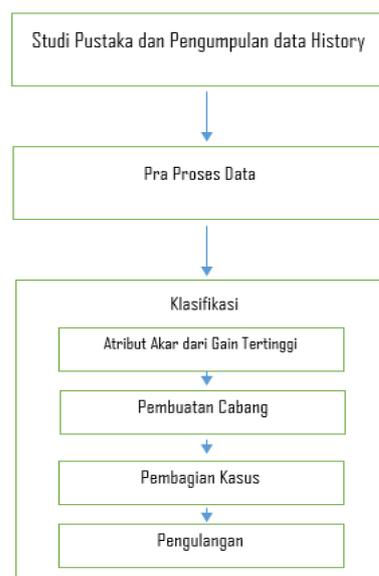
*Decision tree* tidak menggunakan vector jarak untuk mengklasifikasikan obyek. Seringkali data observasi mempunyai atribut-atribut yang bernilai nominal. Misalkan obyeknya adalah sekumpulan buah-buahan yang bisa dibedakan berdasarkan atribut bentuk, warna, ukuran dan rasa. Bentuk, warna, ukuran dan rasa adalah besaran nominal, yaitu bersifat kategoris dan tiap nilai tidak bisa dijumlahkan atau dikurangkan. Dalam atribut warna ada beberapa nilai yang mungkin yaitu hijau, kuning, merah. Dalam atribut ukuran ada nilai besar, sedang dan kecil.

Dengan nilai-nilai atribut ini, kemudian dibuat decision tree untuk menentukan suatu obyek termasuk jenis buah apa jika nilai tiap-tiap atribut diberikan.

Ada beberapa macam algoritma decision tree diantaranya CART dan C4.5. Beberapa isu utama dalam *decision tree* yang menjadi perhatian yaitu seberapa detail dalam mengembangkan *decision tree*, bagaimana mengatasi atribut yang bernilai continues, memilih ukuran yang cocok untuk penentuan atribut, menangani data *training* yang mempunyai data yang atributnya tidak mempunyai nilai, memperbaiki efisiensi perhitungan. *Decision tree* sesuai digunakan untuk kasus-kasus yang keluarannya bernilai diskrit. Walaupun banyak variasi model *decision tree* dengan tingkat kemampuan dan syarat yang berbeda, pada umumnya beberapa ciri yang cocok untuk diterapkannya *decision tree* adalah pertama data dinyatakan dengan pasangan atribut dan nilainya, label/keluaran data biasanya bernilai diskrit dan data mempunyai missing value (nilai dari suatu atribut tidak diketahui). Dengan cara ini akan mudah mengelompokkan obyek ke dalam beberapa kelompok. Untuk membuat *decision tree* perlu memperhatikan hal-hal yaitu atribut mana yang akan dipilih untuk pemisahan obyek, urutan atribut mana yang akan dipilih terlebih dahulu, struktur tree, kriteria pemberhentian dan *pruning*.

Manfaat dari penelitian ini yaitu bagi Fakultas Ilmu Keperawatan diharapkan dengan adanya prediksi menggunakan data mining dapat membantu menyajikan informasi dan pengetahuan tentang kinerja mahasiswa sebelum menjadi peserta *capping day* sehingga dapat menjadi timbal balik yang dapat dijadikan input bagi proses peningkatan mutu pendidikan.

## 2. METODE PENELITIAN



Gambar 1 Kerangka Konseptual Penelitian

Kegiatan yang dilakukan dalam pengembangan implementasi data mining ini adalah:

### 1. Persiapan Awal

Tahap ini dilakukan survei terhadap proses penilaian mahasiswa di Fakultas Ilmu Keperawatan Universitas Klabat sebelum mengikuti acara *capping day* kemudian mengumpulkan data *history* mahasiswa fakultas keperawatan semester I sampai IV untuk mengetahui atribut dan kinerja mahasiswa. Tahap ini juga dilakukan studi pustaka.

### 2. Praproses data.

Praproses data dilakukan yaitu seleksi data, cleaning dan transformasi.

### 3. Klasifikasi Data.

Klasifikasi data menggunakan metode C.45. Langkah yang dibuat yaitu memilih atribut sebagai akar berdasar nilai Gain tertinggi. Kemudian dibuat cabang untuk tiap-tiap nilai didalam akar tersebut. Langkah berikutnya yaitu membagi kasus dalam cabang. Kemudian ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama. Data input akan diproses untuk membentuk tabel probabilitas sebagai dasar proses klasifikasi dan output dari sistem ini berupa klasifikasi kinerja akademik mahasiswa yang diprediksi kelulusannya. Gambar 1 adalah kerangka konseptual penelitian.

### 3. HASIL DAN PEMBAHASAN

Data yang diperlukan dalam penelitian ini yaitu data dari mahasiswa baru yang masuk tahun 2014 dan tahun 2015 di semester ganjil. Data yang diperoleh adalah data yang dikumpulkan dari Ketua Program Studi Keperawatan menunjukkan sebagian mahasiswa yang telah mengikuti *capping day* pada tahun 2016 dan tahun 2017. Sesuai kurikulum yang berlaku di fakultas keperawatan maka mahasiswa dijadwalkan untuk dapat mengikuti *capping day* di tahun 2016 semester ganjil bila mahasiswa itu masuk atau menjadi mahasiswa baru di tahun 2014 semester ganjil.

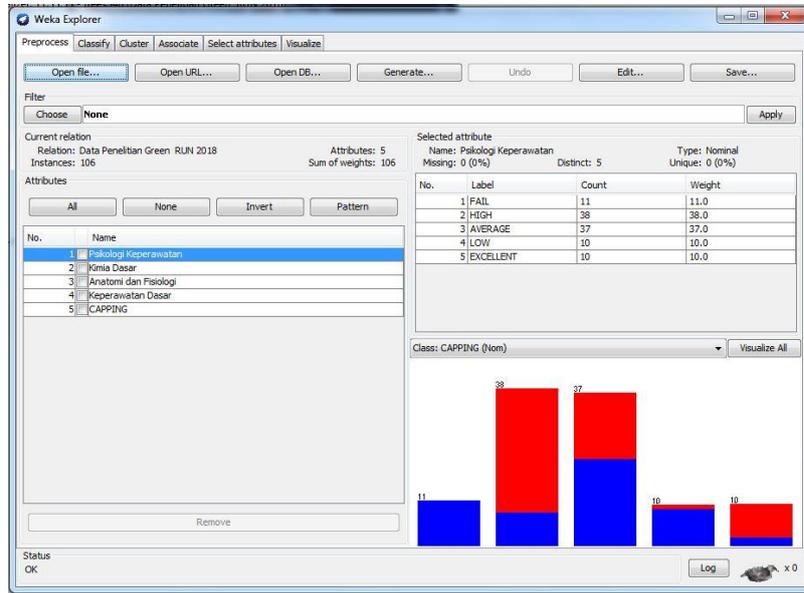
Setelah diperoleh data mahasiswa baru, peneliti mencari dan mengumpulkan data mahasiswa yang sudah mengikuti *capping day* pada tahun 2016 dan tahun 2017. Beberapa mahasiswa yang masuk tahun 2014 mengikuti *capping day* dan beberapa tidak sesuai jadwal. Begitu pula dengan mahasiswa yang masuk di tahun 2015 beberapa mahasiswa mengikuti *capping day* di tahun 2017 dan yang lain tidak mengikuti *capping day* sesuai jadwal.

Model prediksi dibangun dari data yang sudah dikumpulkan dengan pemanfaatan data mining menggunakan algoritma C4.5. Adapun beberapa atribut yang digunakan adalah nilai Psikologi Keperawatan, Kimia Dasar, Anatomi dan Fisiologi serta Keperawatan Dasar. Alasan nilai ini yang diambil karena nilai ini adalah nilai yang merupakan awal serta syarat dari beberapa matakuliah lanjutan dan merupakan syarat untuk mengikuti Capping. Setiap nilai yang diperoleh diubah dengan ketentuan bila nilai itu diantara 0-66 Fail, 67-74 Low, 75-84 Average, 85-94 High dan 95-100 Excellent. Terdapat 106 record yang siap diolah. Sebelum data diolah di Weka, maka data diubah dalam format CSV seperti terlihat pada gambar 2.

	A	B	C	D	E
1	Psikologi Keperawatan	Kimia Dasar	Anatomi dan Fisiologi	Keperawatan Dasar	CAPPING
2	FAIL	FAIL	FAIL	FAIL	No
3	HIGH	AVERAGE	AVERAGE	HIGH	No
4	AVERAGE	AVERAGE	HIGH	EXCELLENT	No
5	AVERAGE	AVERAGE	LOW	AVERAGE	No
6	AVERAGE	LOW	AVERAGE	LOW	No
7	HIGH	LOW	HIGH	AVERAGE	Yes
8	HIGH	HIGH	HIGH	HIGH	Yes
9	HIGH	HIGH	AVERAGE	AVERAGE	Yes
10	HIGH	AVERAGE	AVERAGE	AVERAGE	No
11	AVERAGE	AVERAGE	LOW	AVERAGE	Yes
12	AVERAGE	AVERAGE	LOW	LOW	No
13	AVERAGE	LOW	FAIL	FAIL	No
14	AVERAGE	AVERAGE	AVERAGE	AVERAGE	No
15	FAIL	FAIL	FAIL	FAIL	No
16	HIGH	AVERAGE	AVERAGE	AVERAGE	Yes
17	LOW	LOW	LOW	LOW	No
18	EXCELLENT	HIGH	HIGH	EXCELLENT	No
19	LOW	FAIL	LOW	LOW	No
20	AVERAGE	AVERAGE	HIGH	AVERAGE	No
21	HIGH	HIGH	HIGH	HIGH	Yes
22	AVERAGE	LOW	LOW	HIGH	Yes
23	FAIL	FAIL	LOW	LOW	No
24	HIGH	AVERAGE	AVERAGE	AVERAGE	Yes
25	HIGH	AVERAGE	AVERAGE	HIGH	Yes
26	AVERAGE	AVERAGE	FAIL	LOW	No
27	FAIL	FAIL	FAIL	FAIL	No
28	HIGH	AVERAGE	HIGH	HIGH	Yes

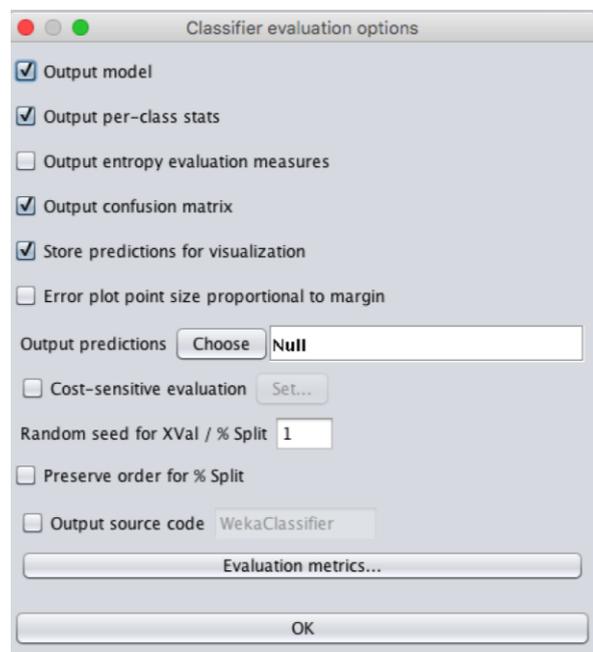
Gambar 2 Data Penelitian

Data yang sudah siap dalam format CSV sekarang dimasukkan dalam program Weka. Program Weka akan menerima data yang berformat .artff namun format CSV juga bisa diterima. Data yang sudah dimasukkan didalam Weka berhasil diterima seperti yang terlihat pada gambar 3.



Gambar 3 Preproses data di Weka

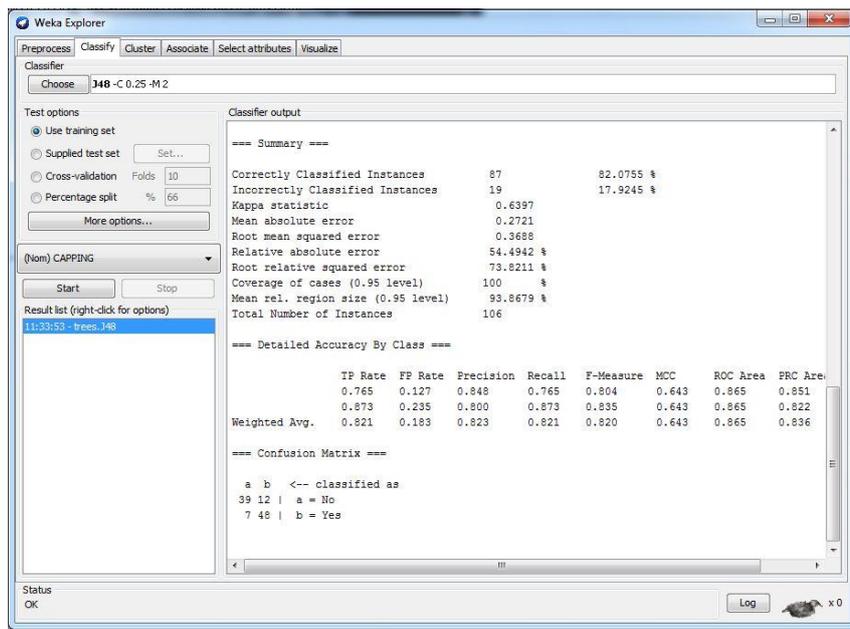
Terdapat 5 atribut yang terlihat dengan sum of weight sebanyak 106. Atribut yang menjadi target adalah atribut *capping day* dimana memiliki instance Yes dan No yang artinya mengikuti *capping day* tepat waktu atau No yang artinya tidak mengikuti *capping day* tepat waktu.



Gambar 4 Classifier evaluation Options

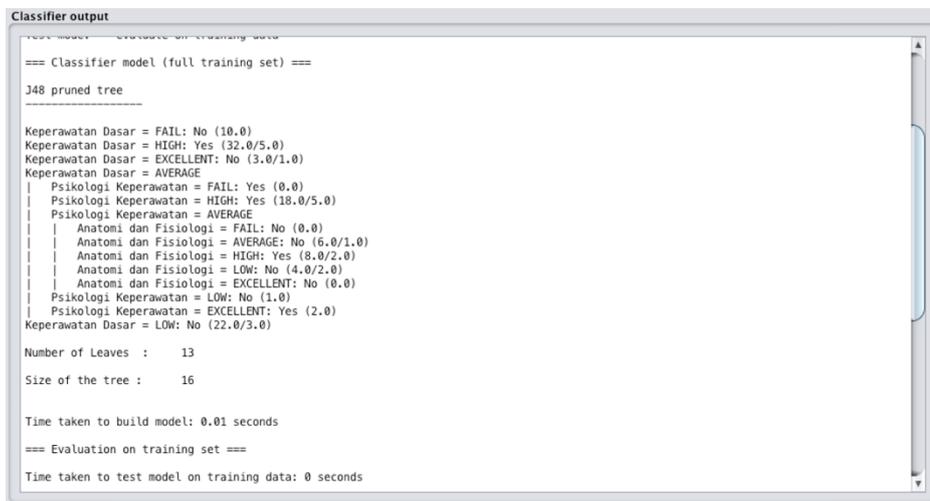
Setelah data dimasukkan ke Weka maka yang dilakukan selanjutnya adalah membuat aturan model pohon keputusan dengan memilih pada menu *classify* dan pilihan J48 yang adalah prediksi C4.5 dengan pengaturan seperti terlihat pada gambar 4.

Gambar 5 menunjukkan data yang dibuatkan pohon keputusan dengan algoritma C4.5 setelah dijalankan menghasilkan *Correctly Classified Instance* menunjukkan angka sebesar 87 atau 82.075% dan *Incorrectly Classified Instance* menunjukkan angka 19 atau sebesar 17.924%. *Kappa statistic* menunjukkan angka 0.6397, *mean absolute error* menunjukkan angka 0.2721, *root mean squared error* menunjukkan angka sebesar 0.3688 dan *relative absolute error* sebesar 54.4942 dan *root relative squared error* menunjukkan angka 73.8211. Bila dilihat dari angka-angka yang diperoleh maka pohon keputusan yang dihasilkan bisa memprediksikan sebesar 82% sesuai atau 18% yang misklasifikasi



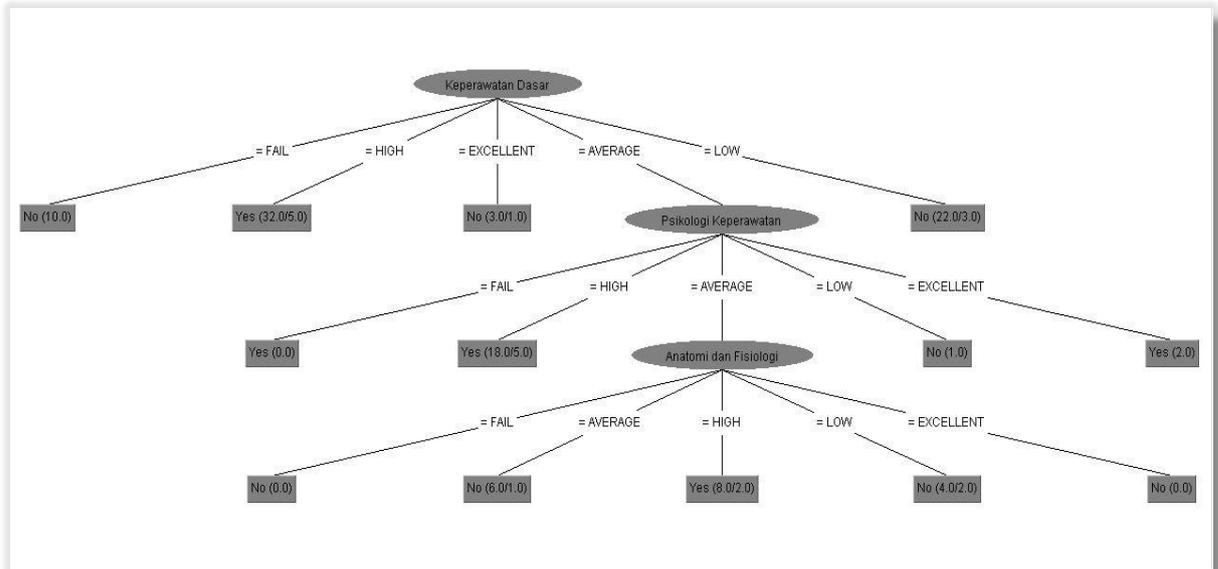
Gambar 5 Classify dengan Algoritma C4.5

Gambar 6 menunjukkan jumlah cabang sebanyak 13 dan jumlah ukuran dari ukuran sebanyak 16. Waktu yang dibutuhkan untuk membangun model selama 0.01 detik.



Gambar 6 Classifier Model

Gambar 7 menunjukkan gambar pohon keputusan untuk memprediksi mahasiswa peserta *capping day*.



Gambar 7 Model Prediksi dengan Algoritma C4.5

Berdasarkan model pohon keputusan algoritma C4.5 didapati bahwa 3 atribut yang menjadi penentu mahasiswa itu mengikuti *capping day* tepat waktu ataupun tidak. Yang pertama adalah berdasarkan nilai Gain yang paling tinggi yaitu nilai Keperawatan Dasar, bila nilai Keperawatan Dasar Fail, Excellent dan Low maka prediksi untuk mahasiswa itu tidak akan mengikuti *capping day* tepat waktu. Bila nilai High maka mengikuti Capping tepat waktu. Bila nilai Average maka dilihat pada nilai Psikologi Keperawatan bila nilainya Excellent High dan Fail maka prediksinya yaitu Yes. Bila Nilainya Low maka prediksi adalah No. Bila nilai adalah Average dilihat pada nilai Anatomi dan Fisiologi. Bila nilai Fail, Average, Low dan Excellent maka prediksinya adalah No dan bila nilai maka prediksinya adalah Yes.

#### 4. KESIMPULAN

Setelah didapatkan hasil dari model prediksi yang sudah dibangun maka ada yang tepat da nada yang misklasifikasi. Pohon keputusan yang terbentuk untuk nilai Keperawatan Excellent yang mempunyai prediksi No dan nilai Psikologi Keperawatan Low yang mempunyai prediksi Yes serta nilai Anatomi dan Fisiologi Excellent yang mempunyai Prediksi No adalah hasil diluar prediksi. Namun sesuai data yang sudah dikumpulkan hasil yang diperoleh demikian sehingga untuk fakultas keperawatan yang akan menggunakan model ini untuk prediksi presentasi yang akan didapat yaitu sebesar 82% yang tepat dan ada sebesar 18% yang misklasifikasi.

#### 5. SARAN

Saran yang dapat diberikan untuk fakultas keperawatan di Universitas Klabat adalah supaya dapat memperhatikan para mahasiswa yang berada di Tingat I ketika mengambil mata kuliah yang menjadi syarat mengikuti *capping day* agar mempunyai extra perhatian kepada mahasiswa dan untuk penelitian selanjutnya agar menambahkan beberapa atribut seperti nilai tes masuk juga demografi dari mahasiswa.

## DAFTAR PUSTAKA

- [1] Ellis and Hartler. *Nursing in Today's World*. Lippincot Williams & Wilkins. Philadelphia. 2012
- [2] Davies and P. Beynon, *Database Systems Third Edition*. New York: Palgrave Macmillan, 2004.
- [3] I Pramudiono. (2007) Ilmu Komputer. [Online]. <http://www.ilmukomputer.org/wp-content/uploads/2006/08/iko-datamining.zip>
- [4] B. Santosa, *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu, 2007.
- [5] M. Berry and G. Linoff, *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*, Ney York: John Wiley & Sons, 2004.