

# Implementasi Model Pohon Keputusan Untuk Mengklasifikasi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 [Universitas Diponegoro]

**Ahmad Fatoni**

Mahasiswa Teknik Informatika, Universitas Dian Nuswantoro  
Jl. Nakula No. 9-11 Semarang  
[ahmadfatoni2104@gmail.com](mailto:ahmadfatoni2104@gmail.com)

## ABSTRACT

*The rapid growth of data accumulation has created the conditions for data-rich but still difficult to analyze solutions problems for decision-making. Trees in the analysis of the decision-making problem solving is the mapping of the problem-solving alternatives that can be drawn from these problems. By utilizing the student master data and data graduation students, are expected to predict about the graduation rates of students with student master data through data mining techniques. Categories graduation rate is measured from the time of study and GPA. The algorithm used is a C4.5 algorithm, because it can be used to perform classification / segmentation or prediction. Besides, the information that is displayed in the form of support and confidence value of each category of graduation rates.*

*This final report will describe the attributes that are used in classifying students study period and the resulting decision tree model. Design data mining applications include student input and display data in 2009/2010, setting the variables used, the case data is entered, the results of the decision tree, the establishment of the rule of decision tree algorithm C4.5 made, do the testing rules and testing of student data for determine the outcome of the decision. What are the things that have been done and what has not been done on the development of this data mining will be reviewed at the end of this report.*

*Keyword: Decision Tree Models, Algorithms C.45, Students, Data graduation, Bachelor (S1)*

## 1. Pendahuluan

Perkembangan teknologi informasi telah menyebabkan banyak orang dapat memperoleh data dengan mudah bahkan cenderung berlebihan. Data tersebut semakin lama semakin banyak dan terakumulasi, akibatnya pemanfaatan data yang terakumulasi tersebut menjadi tidak optimal. Banyaknya data yang dimiliki oleh sebuah organisasi bisa menyebabkan kesulitan dalam pengklasifikasian data tersebut untuk kepentingan organisasi. Kegiatan pengklasifikasian yang dilakukan oleh manusia masih memiliki keterbatasan, terutama pada kemampuan manusia dalam menampung jumlah data yang ingin diklasifikasikan. Selain itu bisa juga terjadi kesalahan dalam pengklasifikasian yang dilakukan. Salah satu cara mengatasi masalah ini adalah dengan menggunakan *Data Mining* (DM) dengan teknik klasifikasi. Data mining dapat membantu sebuah organisasi yang memiliki data melimpah untuk memberikan informasi yang dapat mendukung pengambilan keputusan [1].

Dalam dunia pendidikan, data yang berlimpah dan berkesinambungan mengenai siswa yang dibina dan alumni terus dihasilkan. Data yang berlimpah membuka peluang

diterapkannya data mining untuk pengelolaan pendidikan yang lebih baik dan data mining dalam pelaksanaan pembelajaran berbantuan komputer yang lebih efektif [2].

Dalam dunia pendidikan, data yang berlimpah dan berkesinambungan mengenai siswa yang dibina dan alumni terus dihasilkan. Data yang berlimpah membuka peluang diterapkannya data mining untuk pengelolaan pendidikan yang lebih baik dan data mining dalam pelaksanaan pembelajaran berbantuan komputer yang lebih efektif [2].

Universitas Diponegoro atau yang biasa disingkat dengan Undip adalah sebuah perguruan tinggi negeri yang terdapat di Semarang-Jawa Tengah. Undip senantiasa berusaha menjadi universitas nasional yang dikenali dan terakreditasi secara internasional sebagai universitas riset. Undip memiliki 11 fakultas dan program pascasarjana yang memiliki jumlah mahasiswa baru dan lulusan tahun 2009-2014 adalah sebagai berikut :

Tabel 1. Perkembangan jumlah mahasiswa baru, terdaftar dan lulusan per Tahun Akademik Undip jenjang S1 [3]

No.	Tahun Akademik	Jumlah Mahasiswa Baru	Lulusan
1.	2009/2010	6085	5313
2.	2010/2011	7861	4819
3.	2011/2012	7135	5173
4.	2012/2013	7844	5838
5.	2013/2014	7837	5202

Berdasarkan buku wisuda angkatan ke-131, yang lulus 1581 mahasiswa sekitar 25,98% yang lulus tepat waktu dari jumlah mahasiswa terdaftar tahun akademik 2009/2010 adalah 6085 mahasiswa, sehingga masih 75% mahasiswa yang tidak lulus tepat waktu atau lebih dari 8 semester, data di peroleh dari Biro Akademik Undip.

Sesuai dengan peraturan akademik Universitas Diponegoro Tahun 2010 bahwa program sarjana (S1) adalah program pendidikan akademik setelah pendidikan menengah, yang memiliki beban studi sekurang-kurangnya 144 (seratus empat puluh empat) sks dan sebanyak-banyaknya 160 (seratus enam puluh) sks yang dijadwalkan untuk 8 (delapan) semester dan dapat ditempuh dalam waktu kurang dari 8 (delapan) semester dan paling lama 14 (empat belas) semester. [4]

Sehingga dilihat dari segi rata-rata lama lulusan Undip, masih terdapat mahasiswa yang melebihi masa studi 8 (delapan) semester. Ini yang akan menyebabkan penumpukan data mahasiswa yang mangkir atau belum lulus dan menjadi beban kinerja perguruan tinggi kurang baik. Disamping itu juga mengurangi penilaian sebuah perguruan tinggi dari sebuah lembaga yang menilai dan mengevaluasi sebuah perguruan tinggi.

Melihat permasalahan yang terjadi pada Universitas Diponegoro dalam mengambil keputusan untuk mengklasifikasiki masa studi mahasiswa Sarjana (S1), maka diperlukan pembuatan model pohon keputusan untuk mengklasifikasi masa studi mahasiswa jenjang Sarjana (S1).

Oleh karena itu penulis membuat judul **"Implementasi Model Pohon Keputusan untuk Mengklasifikasi Masa Studi Mahasiswa menggunakan Algoritma C4.5 [Universitas Diponegoro]"**.

Tujuan penelitian ini dilakukan adalah

- a. Untuk menentukan jumlah mahasiswa yang lulus tepat waktu atau tidak di lingkungan Undip

- b. Untuk menentukan hasil pohon keputusan dalam mengklasifikasi masa studi mahasiswa.

## 2. Data Mining

*Data Mining* adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual. *Data mining* juga dapat dijabarkan sebagai analisis otomatis dari data yang berjumlah besar atau kompleks dengan tujuan untuk menemukan pola atau kecenderungan yang penting yang biasanya tidak disadari keberadaannya[5].

*Data Mining* bukanlah suatu bidang yang sama sekali baru. Salah satu kesulitan untuk mendefenisikan *data mining* adalah kenyataan bahwa data mining mewarisi banyak aspek dan teknik dari bidang-bidang ilmu yang sudah mapan terlebih dahulu. *Data mining* memiliki akar yang panjang dari bidang ilmuperti kecerdasan buatan(*artificial intelegent*) *machine learning*, statistic, database, dan juga information retrieval[5].

*Data Mining* dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan[6], yaitu.

### 1. Deskripsi

Terkadang peneliti dan analisis sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data. Sebagai contoh petugas pengumpulan suara mungkin tidak dapat menemukan keterangan atau fakta bahwa siapa yang tidak cukup professional akan sedikit didukung dalam pemilihan presiden. Deskripsi dari pola dan kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan.

### 2. Estimasi

Estimasi hamper sama dengan klasifikasi, kecuali variable target estimasi lebih ke arah numeric daripada ke arah kategori. Model dibangun menggunakan record lengkap yang menyediakan nilai dari variable target sebagai nilai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variable target dibuat berdasarkan nilai variable prediksi. Sebagai contoh akan dilakukan estimasi tekanan darah sistolik pada pasien rumah sakit berdasarkan umur pasien, jenis kelamin, indeks berat badan, dan level sodium darah. Hubungan antara tekanan darah sistolik dan nilai variable prediksi dalam proses pembelajaran akan menghasilkan model estimasi. Model estimasi yang dihasilkan dapat digunakan untuk kasus baru lainnya.

### 3. Prediksi

Prediksi hamper sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada di masa mendatang.

Contoh prediksi data bisnis dan penelitian adalah:

- Prediksi harga beras dalam tiga bulan yang akan datang.
- Prediksi presentase kenaikan kecelakaan lalu lintas tahun depan jika batas bawah kecepatan dinaikkan.

Beberapa metode dan teknik yang digunakan dalam klasifikasi dan estimasi dapat pula digunakan (untuk keadaan yang tepat) untuk prediksi.

#### 4. Klasifikasi

Dalam klasifikasi, terdapat target variable kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.

Contoh lain klasifikasi dalam bisnis dan penelitian adalah :

- Menentukan apakah suatu transaksi kartu kredit merupakan transaksi yang curang atau bukan.
- Memperkirakan apakah suatu pengajuan hipotek oleh nasabah merupakan suatu kredit yang baik atau buruk.
- Mendiagnosis penyakit seorang pasien untuk mendapatkan termasuk kategori penyakit apa.

#### 5. Pengklusteran

Pengklusteran merupakan pengelompokan record, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. Kluster adalah kumpulan record yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan dengan record-record dalam kluster lain.

Pengklusteran berbeda dengan klasifikasi yaitu tidak adanya variable target dalam pengklusteran. Pengklusteran tidak mencoba untuk melakukan klasifikasi, mengestimasi, atau memprediksi nilai dari variable target. Akan tetapi, algoritma pengklusteran mencoba melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan, yang mana kemiripan dalam satu kelompok akan bernilai maksimal, sedangkan kemiripan dengan record dalam kelompok yang lain akan bernilai minimal.

Contoh pengklusteran dalam bisnis dan penelitian adalah:

- Mendapatkan kelompok-kelompok konsumen untuk target pemasaran dari suatu produk bagi perusahaan yang tidak memiliki dana pemasaran yang besar.
- Untuk tujuan audit akutansi, yaitu melakukan pemisahan terhadap perilaku financial dalam baik dan mencurigakan.
- Melakukan pengklusteran dalam ekspresi dari gen, untuk mendapatkan kemiripan dari perilakudari gen dalam jumlah besar.

#### 6. Asosiasi

Tugas asosiasi dalam *data mining* adalah menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja.

Contoh asosiasi dalam bisnis dan penelitian adalah:

- Meneliti jumlah pelanggan dari perusahaan telekomunikasi seluler yang diharapkan untuk memberikan respon positif terhadap penawaran upgrade layanan yang diberikan.
- Menemukan barang dalam supermarket yang dibeli secara bersamaan dan barang yang tidak pernah dibeli secara bersamaan.

### 3. Model Pohon Keputusan

#### • Sejarah Pohon Keputusan

Di dalam kehidupan manusia sehari-hari, manusia selalu dihadapkan oleh berbagai macam masalah dari berbagai macam bidang. Masalah-masalah yang dihadapi oleh manusia memiliki tingkat kesulitan dan kompleksitas yang sangat bervariasi, mulai dari masalah yang teramat sederhana dengan sedikit faktor-faktor yang terkait, sampai dengan masalah yang sangat rumit dengan banyak sekali faktor-faktor yang terkait dan perlu untuk diperhitungkan. Untuk menghadapi masalah-masalah ini, manusia mulai mengembangkan sebuah sistem yang dapat membantu manusia agar dapat dengan mudah mampu untuk menyelesaikan masalah-masalah tersebut.

Adapun pohon keputusan ini adalah sebuah jawaban akan sebuah sistem yang manusia kembangkan untuk membantu mencari dan membuat keputusan untuk masalah-masalah tersebut dan dengan memperhitungkan berbagai macam faktor yang ada di dalam lingkup masalah tersebut. Dengan pohon keputusan, manusia dapat dengan mudah mengidentifikasi dan melihat hubungan antara faktor-faktor yang mempengaruhi suatu masalah dan dapat mencari penyelesaian terbaik dengan memperhitungkan faktor-faktor tersebut.

Pohon keputusan ini juga dapat menganalisa nilai resiko dan nilai suatu informasi yang terdapat dalam suatu alternatif pemecahan masalah. Peranan pohon keputusan sebagai alat bantu dalam mengambil keputusan (*decision support tool*) telah dikembangkan oleh manusia sejak perkembangan teori pohon yang dilandaskan pada teori graf. Kegunaan pohon keputusan yang sangat banyak ini membuatnya telah dimanfaatkan oleh manusia dalam berbagai macam sistem pengambilan keputusan.

#### • Pengertian Pohon Keputusan

Pohon dalam analisis pemecahan masalah pengambilan keputusan adalah pemetaan mengenai alternatif-alternatif pemecahan masalah yang dapat diambil dari masalah tersebut. Pohon tersebut juga memperlihatkan faktor-faktor kemungkinan/probabilitas yang akan mempengaruhi alternatif-alternatif keputusan tersebut, disertai dengan estimasi hasil akhir yang akan didapat bila kita mengambil alternatif keputusan tersebut.

Pohon keputusan merupakan metode klasifikasi yang paling populer digunakan. Selain karena pembangunannya relatif cepat, hasil dari model yang dibangun mudah untuk dipahami. Pada decision tree terdapat 3 jenis node, yaitu:

- a. Root Node, merupakan node paling atas, pada node ini tidak ada input dan bisa tidak mempunyai output atau mempunyai output lebih dari satu.
- b. Internal Node, merupakan node percabangan, pada node ini hanya terdapat satu input dan mempunyai output minimal dua.
- c. Leaf node atau terminal node, merupakan node akhir, pada node ini hanya terdapat satu input dan tidak mempunyai output.

- **Manfaat Pohon Keputusan**

Pohon keputusan adalah salah satu metode klasifikasi yang paling populer karena mudah untuk diinterpretasi oleh manusia. Pohon keputusan adalah model prediksi menggunakan struktur pohon atau struktur berhirarki. Konsep dari pohon keputusan adalah mengubah data menjadi pohon keputusan dan aturan-aturan keputusan. Manfaat utama dari penggunaan pohon keputusan adalah kemampuannya untuk mem-*break down* proses pengambilan keputusan yang kompleks menjadi lebih simpel sehingga pengambil keputusan akan lebih menginterpretasikan solusi dari permasalahan.

Pohon Keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target. Pohon keputusan memadukan antara eksplorasi data dan pemodelan, sehingga sangat bagus sebagai langkah awal dalam proses pemodelan bahkan ketika dijadikan sebagai model akhir dari beberapa teknik lain. Sering terjadi tawar menawar antara keakuratan model dengan transparansi model. Dalam beberapa aplikasi, akurasi dari sebuah klasifikasi atau prediksi adalah satu-satunya hal yang ditonjolkan, misalnya sebuah perusahaan *direct mail* membuat sebuah model yang akurat untuk memprediksi anggota mana yang berpotensi untuk merespon permintaan, tanpa memperhatikan bagaimana atau mengapa model tersebut bekerja.

- **Kelebihan Pohon Keputusan**

Kelebihan dari metode pohon keputusan adalah:

- a) Daerah pengambilan keputusan yang sebelumnya kompleks dan sangat global, dapat diubah menjadi lebih simpel dan spesifik.
- b) Eliminasi perhitungan-perhitungan yang tidak diperlukan, karena ketika menggunakan metode pohon keputusan maka sample diuji hanya berdasarkan kriteria atau kelas tertentu.
- c) Fleksibel untuk memilih fitur dari internal node yang berbeda, fitur yang terpilih akan membedakan suatu kriteria dibandingkan kriteria yang lain dalam node yang sama.
- d) Kefleksibelan metode pohon keputusan ini meningkatkan kualitas keputusan yang dihasilkan jika dibandingkan ketika menggunakan metode penghitungan satu tahap yang lebih konvensional.
- e) Dalam analisis multivariat, dengan kriteria dan kelas yang jumlahnya sangat banyak, seorang penguji biasanya perlu untuk mengestimasi baik itu distribusi dimensi tinggi ataupun parameter tertentu dari distribusi kelas tersebut. Metode pohon keputusan dapat
- f) Menghindari munculnya permasalahan ini dengan menggunakan kriteria yang jumlahnya lebih sedikit pada setiap node internal tanpa banyak mengurangi kualitas keputusan yang dihasilkan.

- **Kekurangan Pohon Keputusan**

- a) Terjadi overlap terutama ketika kelas-kelas dan kriteria yang digunakan jumlahnya sangat banyak. Hal tersebut juga dapat menyebabkan meningkatnya waktu pengambilan keputusan dan jumlah memori yang diperlukan.
- b) Pengakumulasi jumlah eror dari setiap tingkat dalam sebuah pohon keputusan yang besar.
- c) Kesulitan dalam mendesain pohon keputusan yang optimal.
- d) Hasil kualitas keputusan yang didapatkan dari metode pohon keputusan sangat tergantung pada bagaimana pohon tersebut didesain.

#### 4. Algoritma C4.5

Algoritma adalah kumpulan perintah yang tertulis secara sistematis guna menyelesaikan permasalahan logika dari matematika.

Pengertian Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan. Untuk membangun pohon keputusan dalam algoritma C4.5, hal pertama yang dilakukan yaitu memilih atribut sebagai akar, 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. Untuk memilih atribut dengan akar, didasarkan pada nilai *gain* tertinggi dari atribut-tribut yang ada.

Algoritma C4.5 merupakan kelompok algoritma decision tree. Algoritma ini mempunyai input berupa training samples dan samples. Training samples berupa data contoh yang akan digunakan untuk membangun sebuah tree yang telah diuji kebenarannya. Sedangkan samples merupakan field-field data yang nantinya akan kita gunakan sebagai parameter dalam melakukan klasifikasi data[7].

Algoritma C4.5 adalah algoritma hasil pengembangan dari algoritma ID3. Perbaikan algoritma C4.5 dari algoritma ID3 dilakukan dalam hal(Santosa,ID3):

1. Bisa mengatasi missing value
2. Bisa mengatasi data kontinyu
3. Pruning
4. Adanya aturan

Secara garis besar langkah-langkah yang dilakukan oleh algoritma C4.5 dalam membentuk pohon keputusan adalah sebagai berikut:

1. Pada saat awal pembentukan pohon akan dimulai dengan membuat suatu node yang melambangkan training sample.
2. Jika sampel-sampel tersebut memiliki kelas yang sama, maka node tersebut dijadikan leaf node dengan label kelas tersebut.
3. Jika sampel-sampel tersebut tidak memiliki kelas yang sama maka algoritma ini akan mencari gain ratio tertinggi dari atribut yang tersedia, sebagai cara untuk memilih atribut yang paling berpengaruh paada training sample yang tersedia. Nantinya atribut ini akan dijadikan atribut “penguji” atau “keputusan” pada node tersebut. Hal yang perlu diperhatikan adalah ketika atribut tersebut bernilai kontinu, maka atribut tersebut harus di diskritkan terlebih dahulu.
4. Cabang untuk setiap node akan dibentuk berdasarkan nilai-nilai yang diketahui dari atribut pengujian.
5. Algoritma ini akan terus melakukan proses yang sama rekursif untuk membentuk suatu pohon keputusan untuk setiap sample di setiap bagiannya.
6. Proses rekursif ini akan berhenti, ketika salah satu dari kondisi di bawah telah terpenuhi. Kondisi-kondisinya adalah:

- a. Semua sample yang diberikan pada node adalah berasal dari satu kelas yang sama.

- b. Tidak ada atribut lainnya yang dapat digunakan untuk mempartisi sample lebih lanjut.

Tidak ada sample yang memenuhi test-attribute=  $a_i$ . Dalam hal ini, sebuah daun dibuat dan dilabeli dengan kelas yang memiliki sample terbanyak (majority voting). Pemilihan atribut sebagai simpul, baik akar (*root*) atau simpul internal didasarkan pada nilai *Gain* tertinggi dari atribut-atribut yang ada. Perhitungan nilai *Gain* digunakan rumus seperti dalam Persamaan 1.

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

di mana:

- $S$  : Himpunan kasus
- $A$  : Atribut
- $n$  : Jumlah partisi atribut  $A$
- $|S_i|$  : Jumlah kasus pada partisi ke- $i$
- $|S|$  : Jumlah kasus dalam  $S$

Untuk menghitung nilai *Entropy* dapat dilihat pada Persamaan 2.

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

di mana:

- $S$  : Himpunan kasus
- $n$  : Jumlah partisi  $S$
- $p_i$  : Proporsi dari  $S_i$  terhadap  $S$

## 5. Metode Penelitian

Dalam penelitian ini menggunakan metode deskriptif. Dimana penelitian ini bertujuan untuk memecahkan masalah yang terjadi di Universitas Diponegoro. Pada penelitian ini adalah prediksi masa studi mahasiswa. Metode deskriptif mempunyai ciri sebagai berikut :

- a. Berpusat pada penyelesaian masalah yang terjadi saat ini dan masalah yang bersifat aktual.
- b. Data yang dikumpulkan sebelumnya harus di susun, dijelaskan dan dianalisis karena metode ini sering disebut metode analitik.

## 6. Pengujian Data

Pada tahap pengujian, data yang sudah ditransformasikan ke dalam bentuk kelompok data atribut yang lebih sederhana. Data yang digunakan adalah data mahasiswa yang lulus tepat waktu dan tidak tepat waktu pada Universitas Diponegoro yang diambil secara

acak/random sebanyak 30 data sampel. Data akan diolah dengan menerapkan *data mining* algoritma C4.5 dengan memilih atribut yang dipilih kemudian dibuat model pohon keputusan untuk menentukan hasil mengklasifikasi masa studi mahasiswa pada Universitas Diponegoro. Beberapa atribut atau komponen variabel yang digunakan untuk penentuan gain tertinggi yaitu terlihat pada (Lampiran 1).

Kemudian hitung *entropy* dengan rumus sebagai berikut:

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

Maka hasil perhitungannya sebagai berikut  
Tabel 2. Hasil Perhitungan *entropy*

Hasil	Frekuensi	Proporsi	Log2 Proporsi	-Proporsi Log
Yes	13	0.43333	-1.20645	0.52280
No	17	0.56667	-0.81943	0.46434
Total	30		Entropi	0.98714

Gambar 1 menunjukkan deskripsi lengkap dari pohon keputusan (*decision tree*) yang terbentuk dari 30 data sampel mahasiswa untuk mengklasifikasi masa studi mahasiswa.

1	Jika ipk_semester 1 = SEDANG Dan jalur_masuk = SBMPTN Dan jurusan_sita = IPA Maka No
2	Jika ipk_semester 1 = SEDANG Dan jalur_masuk = SBMPTN Dan jurusan_sita = IPS Maka Yes
3	Jika ipk_semester 1 = SEDANG Dan jalur_masuk = SNMPTN Dan pilihan_masuk = KEDUA Maka Yes
4	Jika ipk_semester 1 = SEDANG Dan jalur_masuk = SNMPTN Dan pilihan_masuk = KETIGA Maka No
5	Jika ipk_semester 1 = SEDANG Dan jalur_masuk = MANDIRI Maka No
6	Jika ipk_semester 1 = TINGGI Dan jalur_masuk = SBMPTN Maka Yes
7	Jika ipk_semester 1 = TINGGI Dan jalur_masuk = SNMPTN Maka Yes

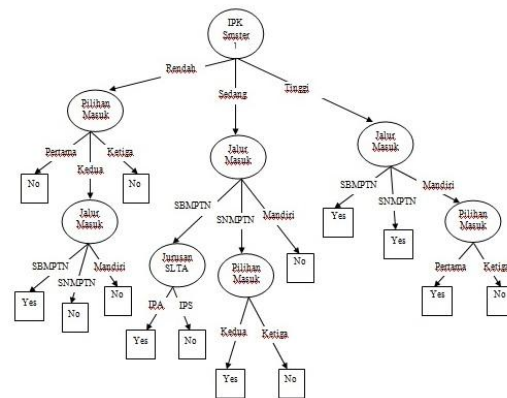
Gambar 1. Deskripsi *Decision Tree*

Gambar 1. menunjukkan hasil deskripsi secara lengkap dari pohon keputusan (*decision tree*) yang telah terbentuk dengan menggunakan algoritma C4.5. Dari hasil dekripsi pada Gambar 5 juga menunjukkan bahwa penggunaan *data mining* algoritma C4.5 baik digunakan dalam proses menggali data (*data mining process*) untuk menarik beberapa kesimpulan yang divisualisasikan dengan pohon keputusan (*decision tree*).

## 8. Hasil Penelitian

Hasil penelitian secara lengkap dari pohon keputusan (*decision tree*) yang telah terbentuk dengan menunjukkan bahwa penggunaan data mining algoritma C4.5 baik digunakan dalam proses menggali data (*data mining process*) untuk menarik beberapa kesimpulan dengan pohon keputusan () untuk mengklasifikasi jumlah masa studi mahasiswa yang tepat waktu dan yang tidak tepat waktu.

Berikut ini adalah hasil dari bentuk pohon keputusannya :



Gambar 2. Hasil pohon keputusan

## 7. Kesimpulan dan Saran

Berdasarkan dari penelitian yang telah dilakukan maka dapat diambil beberapa kesimpulan sebagai berikut :

1. Lama masa studi atau dalam hal ini ketepatan masa studi setiap mahasiswa dapat diklasifikasi berdasarkan faktor-faktor yang berkaitan dengan latar belakang sekolah sebelumnya dan data akademik serta pribadi saat berada di perguruan tinggi.
2. Fungsi prediksi dengan memanfaatkan teknik data mining menggunakan algoritma C4.5 telah dapat dibuat dan digunakan untuk memklasifikasi (menentukan kelas) dari masa studi atau ketepatan masa studi dari mahasiswa dengan data training dan data testing yang telah diperoleh.
3. Tingkat kesalahan dari fungsi klasifikasi yang digunakan untuk mengklasifikasi masih dimungkinkan dapat dipengaruhi oleh jumlah data training maupun testing yang digunakan serta tingkat konsisten data yang digunakan.

Adapun saran yang dapat diberikan adalah :

1. Sebaiknya jumlah data yang digunakan training maupun testing ditambah hingga dapat diperoleh hasil akurasi fungsi algoritma yang lebih baik.

2. Untuk pengembangan ke depannya dimungkinkan dilakukan uji coba menggunakan algoritma lain yang hasilnya bisa dilakukan perbandingan dan analisa.
3. Cakupan faktor atau variabel untuk mengklasifikasi dapat ditambah lebih banyak dan variasi nilai data yang juga bisa lebih banyak serta konsistensi data juga diperhatikan.

#### **Daftar Pustaka:**

- [1] Kiron, D., Shockley, R., Kruschwitz, N., Finch, G., & Haydock, M. 2012. *Analytics: The Widening Divide*. MIT Sloan Management Review, 53(2), 1-22.
- [2] Ayub, Mewati, 2007. Proses Data Mining dalam Sistem Pembelajaran Berbantuan Komputer. *Jurnal Sistem Informasi* Vol. 2 No. 1 Maret 2007 : 21-30
- [3] Undip. (2013), Direktori Kemahasiswaan, Universitas Diponegoro, Semarang : BAPSI
- [4] Undip. (2010), Buku Peraturan akademik, Universitas Diponegoro (Undip) Semarang tahun akademik 2010/2011
- [5] Budi Sutedjo Dharma Oetomo, 2002, *Perencanaan dan Pembangunan Sistem Informasi*, Andi Offset, Yogyakarta.
- [6] Bunafit Nugroho, 2005, *Database Relational dengan MySql*, Graha Ilmu, Yogyakarta.
- [7] Jananto, Arif. (2013). “Algoritma Naive Bayes untuk Mencari Perkiraan Waktu Studi Mahasiswa”, *Jurnal Teknologi Informasi DINAMIK*, Volume 18, No.1, Januari 2013 : 09-16
- [8] Pramudiono, I, 2006. Apa itu Data Mining? Dalam <http://datamining.japati.net/cgi-bin/indodm.cgi?bacaarsip&1155527614&artikel>.
- [9] Kusriani, Luthfy, E.T. 2009. *Algoritma Data Mining*. Yogyakarta: Andi.
- [10] <http://sistemdata.undip.ac.id/lulusan/>
- [11] <http://www.unisbank.ac.id/ojs/index.php/fti1/article/view/1669>
- [12] Kadir, K. (2008). “Belajar Basis Data Dengan MySql”, Andi, Yogyakarta.
- [13] [http://www.internetworkingindonesia.org/Issues/Vol-1-No2-Fall2009/ijj\\_vol1\\_no2\\_meinanda.pdf](http://www.internetworkingindonesia.org/Issues/Vol-1-No2-Fall2009/ijj_vol1_no2_meinanda.pdf)

## Lampiran 1

Tabel Penentuan gain tertinggi

	Cabang	Jmlh	Yes	No	P(Yes)	P(No)	$-P(\text{Yes}) \times \text{Log}_2 P(\text{Yes})$	$-P(\text{No}) \times \text{Log}_2 P(\text{No})$	Total	$P(\text{Lulus}=\dots)$	Total X $P(\text{Lulus}=\dots)$	$E(\text{Total} \times P(\text{Lulus}=\dots))$	Gain	
1	Jalur Masuk	SBMPTN	11	8	3	0.73	0.27	0.33	0.51	0.85	0.37	0.31	0.82	0.17184594
		SNMPTN	8	3	5	0.38	0.63	0.53	0.42	0.95	0.27	0.25		
		MANDIRI	11	2	9	0.18	0.82	0.45	0.24	0.68	0.37	0.25		
2	Pilihan Masuk	PERTAMA	8	5	3	0.63	0.38	0.42	0.53	0.95	0.27	0.25	0.87	0.11733020
		KEDUA	11	6	5	0.55	0.45	0.48	0.52	0.99	0.37	0.36		
		KETIGA	11	2	9	0.18	0.82	0.45	0.24	0.68	0.37	0.25		
3	Jurusan SLTA	IPA	13	4	9	0.31	0.69	0.52	0.37	0.89	0.43	0.39	0.95	0.03644283
		IPS	13	7	6	0.54	0.46	0.48	0.51	1.00	0.43	0.43		
		Lain	4	2	2	0.50	0.50	0.50	0.50	1.00	0.13	0.13		
4	IPK Semester 1	RENDAH	10	2	8	0.20	0.80	0.46	0.26	0.72	0.33	0.24	0.70	0.28555557
		SEDANG	11	3	8	0.27	0.73	0.51	0.33	0.85	0.37	0.31		
		TINGGI	9	8	1	0.89	0.11	0.15	0.35	0.50	0.30	0.15		