

MENENTUKAN KEMUNGKINAN MASUKNYA CALON MAHASISWA BARU PADA SEBUAH PERGURUAN TINGGI SWASTA MANGGUNAKAN TEKNIK KLASIFIKASI POHON KEPUTUSAN DENGAN APLIKASI RAPIDMINER 5.1

Petrus Dwi Ananto Pamungkas

Program Studi Teknik Informatika, STMIK-Akademi Bina Insani-Bekasi

Jl. Siliwangi No. 6, Bekasi, Jawa Barat - Indonesia 17113

Telp. (021) 88958130, (021) 82400924

e-mail : petrusdwi@binainsani.ac.id

ABSTRACT

Classification techniques with the formation of the decision tree is a data processing technique enormous able to present the data into information that is easily understood by almost everyone didandingkan in table form. Usually, the results of data collection is in table form. But not everyone is able to read the tables quickly and accurately as you wish manufacturer, especially if the data is very large. Data in this study hundreds of data presented in tabular form. All prospective students information collected in one table with a lot of items, ranging from personal data, parents to home school. The resulting table will be difficult to understand so that the absorption of information becomes old and inaccurate. Presentation of information in the form of a decision tree into a proper alternative for the data very much. By using an application RapidMiner 5.1, the making of a decision tree data from prospective students to more easily and quickly so that the university management can see and know what are the factors that make potential students to continue the registration process to become a student.

Keywords: data, information, decision tree, RapidMiner 5.1 application.

ABSTRAK

Teknik Klasifikasi dengan pembentukan pohon keputusan merupakan salah satu teknik pengolahan data yang sangat besar yang mampu mempresentasikan data menjadi informasi yang mudah untuk dipahami oleh hampir semua orang didandingkan dalam bentuk tabel. Biasanya, hasil pengumpulan data adalah dalam bentuk tabel. Tapi tidak semua orang mampu membaca tabel dengan cepat dan tepat sesuai keinginan pembuatnya, apalagi jika datanya sangat besar. Data dalam penelitian ini ada ratusan data yang disajikan dalam bentuk tabel. Semua informasi calon mahasiswa dikumpulkan dalam satu tabel dengan banyak item, mulai dari data pribadi, orang tua hingga asal sekolah. Tabel yang dihasilkan tersebut akan sulit dipahami sehingga penyerapan informasinya menjadi lama dan tidak akurat. Penyajian informasi dalam bentuk pohon keputusan menjadi alternatif tepat untuk data yang sangat banyak tersebut. Dengan menggunakan aplikasi RapidMiner 5.1 maka pembuatan pohon keputusan dari data calon mahasiswa menjadi lebih mudah dan cepat sehingga pihak manajemen perguruan tinggi dapat melihat dan mengetahui faktor-faktor apa saja yang membuat calon mahasiswa untuk tetap melanjutkan proses pendaftaran hingga menjadi mahasiswa.

Kata kunci: data, informasi, pohon keputusan, aplikasi RapidMiner 5.1

1. PENDAHULUAN

Berdasarkan data Dikti tahun 2014 bahwa terdapat lebih dari 13.000 perguruan tinggi swasta di Indonesia yang aktif melakukan kegiatan belajar-mengajar. Sebagai perguruan tinggi yang melakukan pengelolaan dana sendiri maka perguruan tinggi swasta harus aktif mencari orang-orang yang berminat menjadi mahasiswa. Semakin banyak calon mahasiswa yang menjadi mahasiswa maka semakin besar sumber dana yang dimiliki oleh perguruan tinggi swasta tersebut sehingga kegiatan belajar-mengajar dapat dilakukan secara optimal. Untuk menarik minat calon mahasiswa maka dibutuhkan biaya, waktu dan tenaga yang lumayan besar. Jika memang sebuah perguruan tinggi swasta mampu menganalisis faktor-faktor apa saja yang membuat orang melakukan pendaftaran hingga menjadi mahasiswa maka dapat dihemat biaya, waktu dan tenaga. Kegiatan pendaftaran mahasiswa baru merupakan langkah awal dari kegiatan belajar-mengajar yang akan dilakukan oleh mahasiswa pada sebuah perguruan tinggi. Bagi sebuah perguruan tinggi swasta, kegiatan penerimaan mahasiswa mempunyai peranan yang sangat penting untuk keberlangsungan kegiatan, mulai dari penentuan jumlah kebutuhan dosen hingga jumlah ruangan yang akan

digunakan untuk kegiatan belajar-mengajar. Semuanya butuh biaya yang harus diperhitungkan dengan sebaik-baiknya untuk keberlangsungan hidup perguruan tinggi tersebut.

Berdasarkan latar belakang masalah sebelumnya, maka penelitian difokuskan pada masalah bagaimana menghasilkan informasi dengan mudah dan cepat mengenai faktor-faktor yang menyebabkan calon mahasiswa mendaftar hingga menjadi mahasiswa pada sebuah perguruan tinggi swasta di Bekasi?

Adapun tujuan penelitian yang dilakukan adalah untuk menghasilkan informasi dengan mudah dan cepat mengenai faktor-faktor yang mungkin menjadi penentu para calon mahasiswa mendaftar hingga menjadi mahasiswa pada sebuah perguruan tinggi swasta di Bekasi melalui penggunaan aplikasi RapidMiner 5.1.

Menurut Puspita dan Wahyudi (2015) dalam penelitian “*Algoritma C4.5 Berbasis Decision Tree untuk Prediksi Kelahiran Bayi Prematur*” menjelaskan bahwa dari hasil prediksi model algoritma C4.5 berbasis *Decision Tree* memberikan nilai akurasi yaitu 93.60%. Dengan nilai akurasi yang sudah mencapai *Excellent classification* hal ini menunjukkan bahwa *Decision Tree* merupakan model yang cukup baik.

Menurut Pamungkas (2014) dalam penelitian “Algoritma Data Mining C4.5 untuk Mengetahui Faktor Penentu Pengunduran Diri Calon Mahasiswa pada Suatu Perguruan Tinggi Swasta (Studi Kasus pada STMIK-Akademi Bina Insani, Bekasi)” dijelaskan bahwa melalui Algoritma C4.5 dapat diketahui faktor utama mundurnya calon mahasiswa dalam proses pendaftaran antara lain informasi yang diperoleh dari diri sendiri, orang tua, dan teman. Calon mahasiswa jaman sekarang dapat memperoleh informasi dengan mudah dan cepat tentang tempat-tempat kuliah melalui media internet ditambah lagi dukungan dari orang tua dan teman.

Sedangkan Syahril (2011) dalam penelitian “Konversi Data Training Tentang Penyakit Hipertensi Menjadi Bentuk Pohon Keputusan dengan Teknik Klasifikasi Menggunakan Tools RapidMiner 4.1” menjelaskan bahwa data yang ditampilkan dalam bentuk pohon keputusan menjadi lebih mudah dan cepat untuk dipahami dibandingkan bila data disajikan dalam bentuk tabulasi. RapidMiner 4.1 berperan cukup baik dalam mengkonversi data training yang ada sehingga walaupun data yang dikonversikan dalam bentuk sebuah pohon keputusan berasal dari data yang sangat besar namun pada dasarnya nilai

kandungan data yang ditampilkan memiliki kecenderungan pola yang dapat menghasilkan data yang cukup informatif.

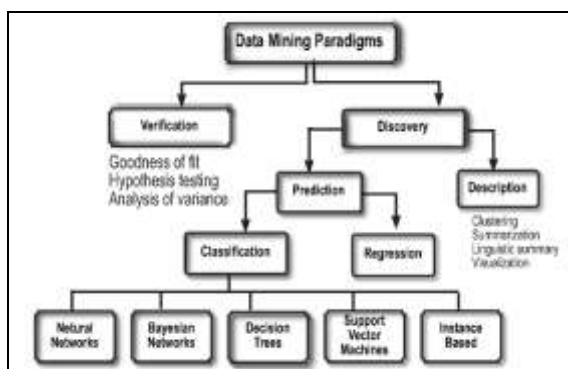
Menurut Gartner Group dalam Larose (2005) bahwa “*Data mining is the process of discovering meaningful new correlations, patterns and trends by sifting through large amounts of data stored in repositories, using pattern recognition technologies as well as statistical and mathematical techniques*”. Sedangkan Hand et al dalam Larose (2005) mengatakan “*Data mining is the analysis of (often large) observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner*”. Hal ini juga diperkuat oleh Cabena et al dalam Larose (2005) bahwa “*Data mining is an interdisciplinary field bringing together techniques from machine learning, pattern recognition, statistics, databases, and visualization to address the issue of information extraction from large data bases*”.

Data mining sangat perlu dilakukan terutama dalam mengelola data yang sangat besar untuk memudahkan *activity recording* suatu transaksi dan untuk proses *data warehousing* agar dapat memberikan informasi yang akurat bagi penggunaanya.

Alasan utama mengapa data mining sangat menarik perhatian industri

informasi dalam beberapa tahun belakangan ini adalah karena tersedianya data dalam jumlah yang besar dan semakin besarnya kebutuhan untuk mengubah data tersebut menjadi informasi dan pengetahuan yang berguna, karena sesuai fokus bidang ilmu ini yaitu melakukan kegiatan mengekstraksi atau menambang pengetahuan dari data yang berukuran/berjumlah besar, informasi inilah yang nantinya sangat berguna untuk pengembangan.

Ada banyak metode data mining yang digunakan untuk tujuan yang berbeda. Menurut Maimon & Rokach (2010) bahwa ada dua kelompok besar dalam metode data mining, yaitu *verification* dan *discovery* (Gambar 1).



Gambar 1. Taksonomi dari Metode Data Mining Menurut Maimon & Rokach

Metode *verification* umumnya meliputi teknik-teknik statistik seperti *goodness of fit*, *hypothesis testing* (Uji-T), dan analisis variansi. Sedangkan metode *discovery* dapat dibagi atas model prediktif

dan model deskriptif. Model prediktif melakukan prediksi terhadap data dengan menggunakan hasil-hasil yang telah diketahui dari data yang berbeda. Model ini dapat dibuat berdasarkan penggunaan data historis lain melalui teknik klasifikasi dan regresi. Untuk teknik klasifikasi dapat digunakan *neural networks*, *bayesian networks*, *decision trees*, *support vector machines*, dan *instance based*. Model deskriptif digunakan untuk mengidentifikasi pola-pola atau hubungan dalam data dan memberikan cara untuk mengeksplorasi sifat-sifat data yang diselidiki. Teknik-teknik yang digunakan dalam metode ini antara lain *clustering*, *summarization*, *linguistic summary*, dan *visualization*.

Pohon keputusan merupakan metode klasifikasi yang banyak digunakan untuk mengubah fakta yang sangat besar menjadi pohon keputusan yang mempresentasikan aturan yang mudah dipahami dengan bahasa alami dan juga dapat diekspresikan dalam bentuk bahasa database, seperti *Structured Query Language (SQL)*, untuk mencari record pada kategori tertentu. Pohon keputusan juga dapat digunakan untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan variabel target.

Menurut Berry & Linoff (2004) bahwa sebuah pohon keputusan adalah sebuah struktur yang dapat digunakan untuk membagi kumpulan data yang besar menjadi himpunan-himpunan *record* yang lebih kecil dengan menerapkan serangkaian aturan keputusan. Dengan masing-masing rangkaian pembagian tersebut, anggota himpunan hasil menjadi mirip satu dengan yang lainnya. Sebuah pohon keputusan mungkin dibangun dengan cara seksama secara manual atau dapat tumbuh secara otomatis dengan menerapkan salah satu atau beberapa algoritma pohon keputusan untuk memodelkan himpunan data yang terklasifikasi.

Data dalam pohon keputusan biasanya dinyatakan dalam bentuk tabel dengan atribut dan record. Atribut menyatakan suatu parameter yang dibuat sebagai kriteria dalam pembentukan pohon. Proses pada pohon keputusan adalah mengubah bentuk data dalam tabel menjadi model pohon dan mengubah bentuk pohon menjadi aturan dan menyederhanakan aturan.

Menurut Larose (2005) bahwa ada beberapa algoritma yang dapat digunakan dalam pembentukan pohon keputusan, antara lain CART, ID3, dan C4.5. Algoritma C4.5 merupakan pengembangan dari algoritma ID3. Secara umum,

algoritma C4.5 untuk membangun sebuah pohon keputusan adalah sebagai berikut:

1. pilih atribut sebagai akar,
2. buat cabang untuk tiap-tiap nilai,
3. bagi kasus dalam cabang, dan
4. ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Untuk memilih atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada. Untuk menghitung gain digunakan rumus berikut:

$Gain(S,A)=$

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

dimana:

- S = Himpunan kasus
- i = Atribut
- n = Jumlah partisi atribut A
- $|S_i|$ = Proporsi S_i terhadap S
- $|S|$ = Jumlah kasus dalam S

Sedangkan penghitungan nilai entropy dapat dilihat dalam persamaan berikut:

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

dimana:

- S = Himpunan kasus
- i = Fitur
- n = Jumlah partisi S
- p_i = Proporsi S_i terhadap S

2. METODE PENELITIAN

Penelitian dilakukan terhadap data calon mahasiswa yang masuk ke STMIK-Akademi Bina Insani, Bekasi. Rentang data yang dimaksud adalah selama tahun 2009, 2010, dan 2011. Metode klasifikasi digunakan untuk mengambil keputusan dengan memperkirakan suatu kasus sehingga perlu dibuatkan variabel target, yaitu lulus dan gagal. Data calon mahasiswa yang mendaftar dibandingkan dengan data mahasiswa yang sudah diterima menjadi mahasiswa STMIK-Akademi Bina Insani, Bekasi. Adapun variabel input yang digunakan antara lain gelombang, pekerjaan orangtua, asal sekolah, jurusan sekolah, pilihan kuliah, dan asal informasi.

3. HASIL DAN PEMBAHASAN

3.1. Penggunaan Aplikasi RapidMiner

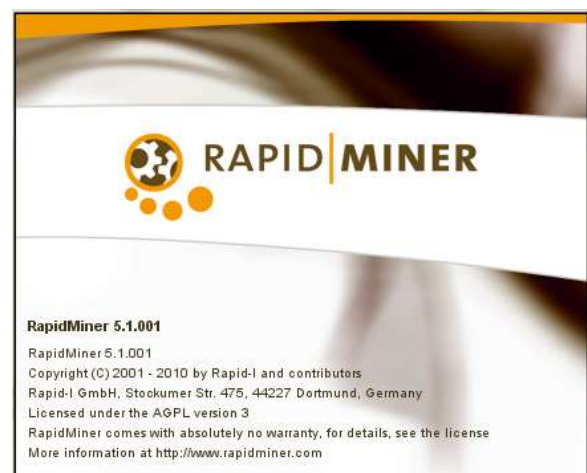
5.1

Sebelum menggunakan aplikasi RapidMiner 5.1, pastikan terlebih dahulu data yang akan diolah. Data yang dipilih adalah data calon mahasiswa yang mendaftar di STMIK-Akademi Bina Insani, Bekasi pada tahun 2009, 2010, dan 2011 sebanyak 744 orang, 905 orang, dan 883 orang. Dari data yang sudah didapat ternyata masih banyak terdapat duplikasi data, inkonsistensi data berupa field yang kosong dan kesalahan cetak (tipografi)

sehingga perlu dilakukan *data cleaning* berupa penghapusan data yang terduplikasi dan inkonsistensi. Dari hasil *data cleaning* ini tersisa data sebanyak 513 orang untuk data tahun 2009, 730 orang untuk data tahun 2010, dan .sebanyak 730 orang untuk data tahun 2011. Tahap selanjutnya adalah dilakukan penyesuaian field dengan membatasi kategori yang ada seperti asal_info hanya berasal dari teman, orangtua, kakak, kerabat, dan diri sendiri dengan datang langsung ke STMIK-Akademi Bina Insani, Bekasi. Untuk asal_jurusan hanya diberikan pilihan SMA, SMK, Madrasah, dan D1. Sedangkan untuk jurusan_sekolah hanya IPA, IPS, Bahasa, dan lainnya.

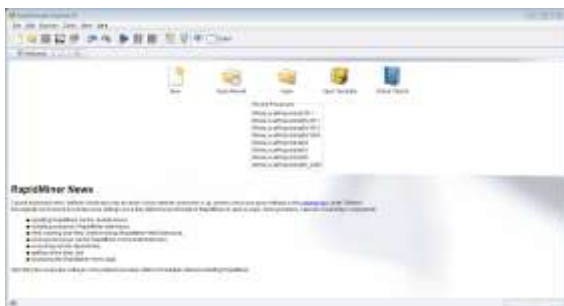
Saat semua data sudah *valid* maka mulai dilakukan pengolahan data dengan menggunakan aplikasi RapidMiner 5.1.

a. Tampilan Awal Saat Menjalankan Aplikasi RapidMiner 5.1



Gambar 2. Tampilan Awal Aplikasi RapidMiner 5.1

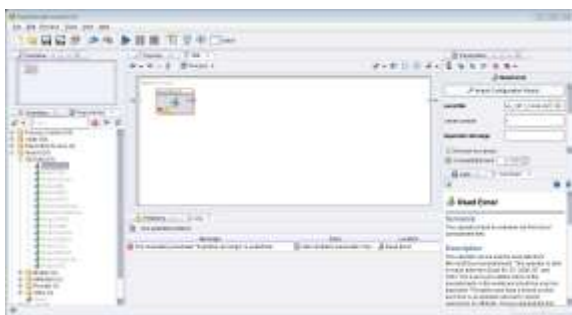
Setelah muncul tampilan logo dan informasi tentang aplikasi RapidMiner 5.1 (Gambar 2) maka muncul tampilan pemilihan tempat penyimpanan file yang nantinya akan diolah dengan menggunakan RapidMiner 5.1, bisa membuat folder baru atau membuka dari file yang ada (lihat Gambar 3).



Gambar 3. Tampilan Pemilihan Tempat Penyimpanan File

b. Tampilan Pengambilan Data Excel

Ketika sukses menentukan tempat penyimpanan file, selanjutnya mengambil data yang akan diolah dalam bentuk file excel (*import data excel*) (Gambar 4).

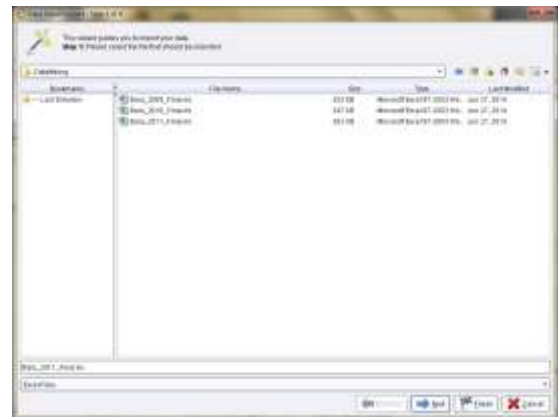


Gambar 4. Tampilan Import Data Excel

c. Penyesuaian Data Excel

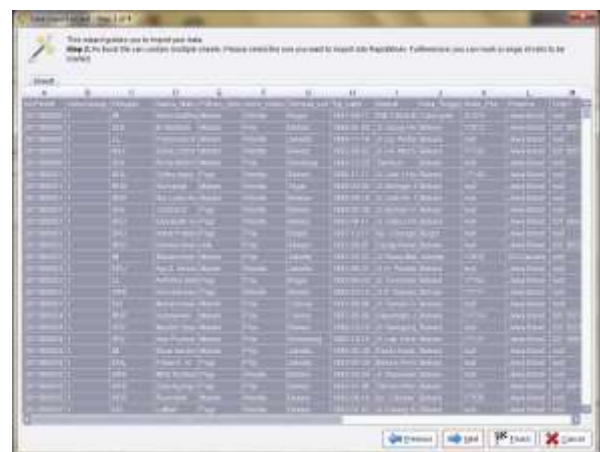
Penyesuaian data excel yang akan diolah dapat dilakukan dengan mengikuti

perintah yang terdapat dalam *Import Configuration Wizard*. Terdapat empat tahap yang harus dilakukan dalam penyesuaian data excel. Tahap pertama merupakan tahap pengambilan data excel (Gambar 5).

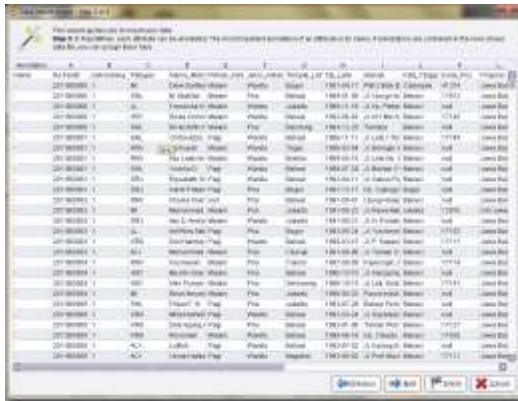


Gambar 5. Tampilan Tahap Pertama Import Configuration Wizard

Setelah tahap pertama selesai dilakukan maka tahap berikutnya adalah pengambilan item data excel yang akan diolah (Gambar 6). Seleksi item-item data apa saja yang akan diolah. Kemudian akan tampil item data yang telah diseleksi (Gambar 7).

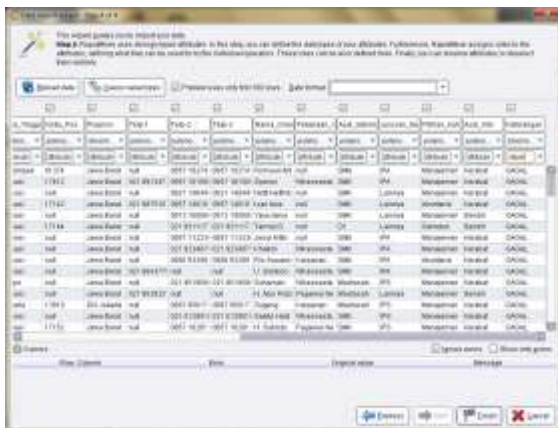


Gambar 6. Tampilan Tahap Kedua Import Configuration Wizard



Gambar 7. Tampilan Tahap Ketiga Import Configuration Wizard

Untuk tampilan terakhir dari *Import Configuration Wizard* adalah menentukan tipe data untuk item KETERANGAN menjadi bentuk LABEL (Gambar 8).



Gambar 8. Tampilan Tahap Keempat Import Configuration Wizard

d. Pemilihan Model Pengolahan Data

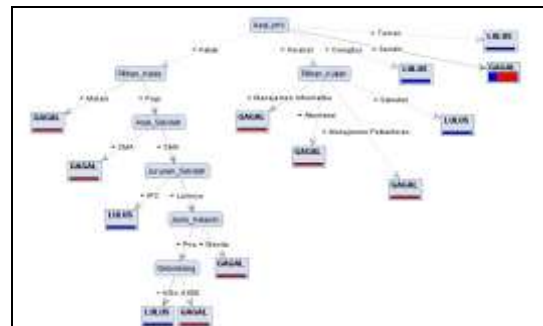
Setelah data yang akan diolah tersedia dalam aplikasi maka tahap berikutnya adalah memilih model pengolahan data. Dalam penelitian ini digunakan model *Decision Tree* (Pohon Keputusan) dan menghubungkan data yang telah disesuaikan dengan model pohon keputusan (Gambar 9).



Gambar 10. Konektivitas antara Data dan Model Pohon Keputusan

3.2. Interpretasi Data

Pola informasi yang dihasilkan dari proses klasifikasi ini ditampilkan dalam bentuk pohon keputusan (*decision tree*) sehingga mudah dimengerti oleh pihak yang berkepentingan.



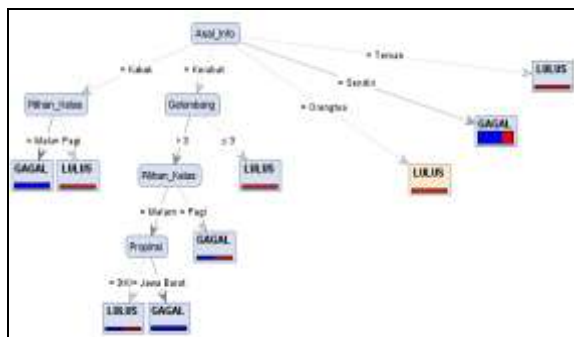
Gambar 10. Pohon Keputusan Tahun 2009

Dari gambar 10 di atas dapat dilihat bahwa ada beberapa faktor yang membuat calon mahasiswa tetap melanjutkan proses pendaftaran hingga menjadi mahasiswa antara lain:

1. Calon mahasiswa yang mendapatkan informasi yang berasal dari orangtua dan teman.
2. Calon mahasiswa yang mendapatkan informasi berasal dari kakak (saudara kandung) dan

memilih untuk kuliah pada pagi hari dan berasal dari SMK dengan jurusan IPS.

3. Calon mahasiswa yang mendapatkan informasi berasal dari kakak (saudara kandung) dan memilih untuk kuliah pada pagi hari serta berasal dari SMK dengan jurusan selain IPS.
4. Calon mahasiswa yang mendapatkan informasi berasal dari kerabat (om, tante, sepupu, maupun tetangga) dan memilih jurusan kuliah sekretari.



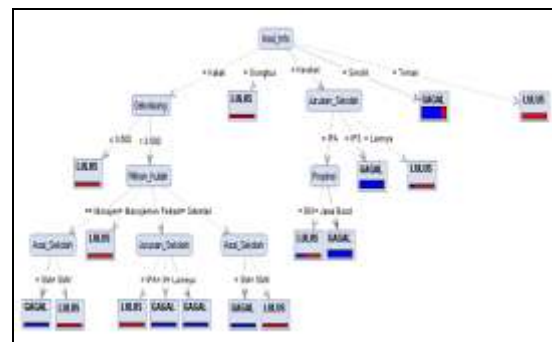
Gambar 11. Pohon Keputusan Tahun 2010

Berdasarkan gambar 11 di atas dapat dilihat bahwa ada beberapa faktor yang membuat calon mahasiswa tetap melanjutkan proses pendaftaran antara lain:

1. Calon mahasiswa yang mendapatkan informasi yang berasal dari orangtua dan teman.
2. Calon mahasiswa yang mendapatkan informasi berasal dari

kakak (saudara kandung) dan memilih untuk kuliah pagi hari.

3. Calon mahasiswa yang mendapatkan informasi yang berasal dari kerabat (om, tante, sepupu, maupun tetangga) dan mendaftar sebelum gelombang ketiga.
4. Calon mahasiswa yang mendapatkan informasi yang berasal dari kerabat (om, tante, sepupu, maupun tetangga) dan mendaftar setelah gelombang ketiga, memilih kuliah di malam hari, serta tinggal di wilayah DKI Jakarta.



Gambar 12. Pohon Keputusan Tahun 2011

Berdasarkan gambar 12 di atas dapat dilihat bahwa ada beberapa faktor yang membuat calon mahasiswa tetap melanjutkan proses pendaftaran antara lain:

1. Calon mahasiswa yang mendapatkan informasi yang berasal dari orangtua dan teman.

2. Calon mahasiswa yang mendapatkan informasi yang berasal dari kerabat (om, tante, sepupu, maupun tetangga) dan berasal dari sekolah dengan jurusan selain IPA dan IPS.
3. Calon mahasiswa yang mendapatkan informasi yang berasal dari kerabat (om, tante, sepupu, maupun tetangga) dan berasal dari sekolah dengan jurusan selain IPA, serta tinggal di DKI Jakarta.
4. Calon mahasiswa yang mendapatkan informasi berasal dari kakak (saudara kandung) dan sudah mendaftar sebelum gelombang ketiga dan keempat.
5. Calon mahasiswa yang mendapatkan informasi berasal dari kakak (saudara kandung) dan mendaftar setelah gelombang ketiga dan keempat serta memilih kuliah program studi Manajemen Informatika.
6. Calon mahasiswa yang mendapatkan informasi berasal dari kakak (saudara kandung), mendaftar setelah gelombang ketiga dan keempat serta memilih kuliah program studi Akuntansi maupun Sekretari, serta berasal dari SMK.
7. Calon mahasiswa yang mendapatkan informasi berasal dari kakak (saudara kandung), mendaftar setelah gelombang ketiga dan keempat serta memilih kuliah program studi Manajemen Administrasi, serta berasal dari sekolah manapun dengan jurusan IPA.

4. SIMPULAN

Sumber informasi yang diperoleh calon mahasiswa menjadi faktor utama lanjutnya proses pendaftaran hingga sampai menjadi mahasiswa. Dengan kemajuan teknologi internet maka semua informasi tentang perguruan tinggi dan kegiatan-kegiatannya dapat diketahui dengan cepat dan mudah. Ditambah dukungan dari orang tua dan teman maka semakin menguatkan niat calon mahasiswa untuk terus menjalani proses pendaftaran hingga menjadi mahasiswa. Testimoni kakak (saudara kandung) dan tetangga yang pernah atau sedang kuliah juga menjadi faktor yang harus diperhatikan oleh pihak manajemen sebuah perguruan tinggi.

DAFTAR PUSTAKA

- [1] Berry, Michael J. A. dan Gordon S. Linoff. 2005. *Introduction to Data Mining and Knowledge Discovery*.

- Third Edition*. Potomac: Two Crows Corporation.
- [2] Bramer, Max. 2007. *Principles of Data Mining*. London: Springer-Verlag.
- [3] Kusriani dan Emha Taufiq Luthfi. 2009. *Algoritma Data Mining*. Yogyakarta: Andi Offset.
- [4] Larose, Daniel T. 2005. *Discovering Knowledge in Data: An Introduction to Data Mining*. New Jersey: John Wiley & Sons, Inc.
- [5] Maimon, Oded dan Lior Rokach. 2010. *Data Mining and Knowledge Discovery Handbook. Second Edition*. New York: Springer Science+Business Media.
- [6] Pamungkas, Petrus Dwi Ananto. 2014. *Algoritma Data Mining C4.5 untuk Mengetahui Faktor Penentu Pengunduran Diri Calon Mahasiswa pada Suatu Perguruan Tinggi Swasta (Studi Kasus pada STMIK-Akademi Bina Insani, Bekasi)*. Bina Insani ICT Jurnal. Vol. 1; No. 2; Agustus 2014; Hal. 113-127. ISSN 2355-3421. STMIK Bina Insani: Bekasi.
- [7] Puspita, Ari, dan Mochamad Wahyudi. 2015. *Algoritma C4.5 Berbasis Decision Tree untuk Prediksi Kelahiran Bayi Prematur*. Prosiding Konferensi Nasional Ilmu Pengetahuan dan Teknologi (KNIT) 2015. Hal. 107-112. ISBN 978-602-72850-0-2. 8 Agustus 2015. Bina Sarana Informatika: Bekasi.
- [8] Rapid-IvGmbH. 2010. *RapidMiner 5: User Manual*. www.rapid-i.com
- [9] Syahril, Muhammad. 2011. *Konversi Data Training Tentang Penyakit Hipertensi Menjadi Bentuk Pohon Keputusan dengan Teknik Klasifikasi Menggunakan Tools RapidMiner 4.1*. Jurnal SAINTIKOM. Vol. 10; No. 2; Mei 2011; Hal. 103-109.