

IMPLEMENTASI RAPIDMINER DALAM MENGANALISA DATA MAHASISWA DROP OUT

¹Sri Wahyuni, ²Kana Saputra S, ³Mochammad Iswan Perangin-Angin

^{1,2,3}Program Studi Teknik Komputer, Universitas Pembangunan Pancabudi Medan
email: ¹sriwahyuni@dosen.pancabudi.ac.id, ²kanasaputras@dosen.pancabudi.ac.id,
³mochammadiswan@gmail.com

Abstract

Data Mining adalah proses penggalian data dari tumpukan database yang berukuran besar yang digunakan untuk menemukan knowledge berupa informasi penting dan bermanfaat. Klasifikasi merupakan salah satu teknik yang ada pada data mining. Algoritma yang digunakan adalah algoritma C4.5 dengan metode Decision Tree (Pohon Keputusan). Decision Tree adalah metode yang merubah fakta dalam bentuk data berguna untuk mengeksplorasi data menjadi sebuah pohon keputusan yang mempresentasikan aturan-aturan yang mudah dipahami untuk menemukan hubungan yang tersembunyi dari variabel input dan target. Salah satu Software Pengolahan Data Mining Decision Tree adalah RapidMiner. Salah satu Tujuan penelitian ini untuk mengklasifikasikan data mahasiswa di Universitas Pembangunan Panca Budi untuk mengetahui faktor mahasiswa yang mengalami dropout. Atribut yang digunakan terdiri dari Asal Sekolah, Agama, Umur Mahasiswa, Pekerjaan Orangtua, Pendapatan Orangtua, dan IPK. Atribut yang paling berpengaruh terhadap mahasiswa yang mengalami dropout adalah Agama.

Keywords: Data Mining, Decision Tree, Klasifikasi, RapidMiner

1. PENDAHULUAN

Data Mining yang juga dikenal sebagai *knowledge*, adalah salah satu bidang yang berkembang pesat karena besarnya kebutuhan akan nilai tambah dari tumpukan *database* skala besar yang terakumulasi sejalan dengan pertumbuhan teknologi informasi yang sangat pesat. *Data mining* dapat diartikan sebagai ekstraksi informasi atau pola yang penting atau menarik dari data yang ada pada *database* yang besar. Menurut Larose dalam Kusrini dan Emha Taufiq (2009). *Data mining* dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu Deskripsi, Estimasi, Prediksi, Klasifikasi, Pengklasteran, Asosiasi. Salah satu teknik yang umum digunakan adalah teknik Klasifikasi. Metode yang digunakan yaitu metode Klasifikasi dengan algoritma C4.5. Pada penelitian sebelumnya, Surjeet dan saurabh (2012) menggunakan data mining untuk membantu memprediksi peningkatan kinerja mahasiswa dengan teknik Klasifikasi, hasil penelitian ini memberikan langkah-langkah untuk meningkatkan kinerja siswa yang diperkirakan akan gagal atau dipromosikan., Angga Raditya (2008) menggunakan data mining untuk mencari pola prediksi hujan dengan algoritma C4.5, akurasi pola prediksi didapat mencapai 79%. Angga dan Riani (2012) menggunakan *data mining* untuk memprediksi kriteria nasabah kredit menggunakan metode klasifikasi yang digunakan decision tree diperoleh sebuah aplikasi yang memudahkan bagian dana bank memperoleh target pemasaran kredit. Muhammad Syahril (2011) menggunakan data mining untuk mengkonversi data training penyakit hipertensi. Pada penelitian ini penulis menggunakan algoritma *Decision Tree C4.5* yang masuk dalam fungsi *Classification*. Metode *decision tree* mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan. Data mahasiswa yang diperoleh dari Sistem Informasi UNPAB lima tahun terakhir digambarkan dalam tabel dibawah ini :

Tabel 1. Data Mahasiswa 2010 sampai 2015

TAHUN	JUMLAH MAHASISWA MASUK	JUMLAH MAHASISWA LULUS
2010	1776 Orang	199 Orang
2011	1979 Orang	630 Orang
2012	2155 Orang	83 Orang
2013	1914 Orang	166 Orang
2014	3286 Orang	81 Orang
2015	3774 Orang	-

2. KAJIAN LITERATUR DAN PEGEMBANGAN HIPOTESIS

Data mining merupakan pencarian *trend* atau pola yang akan dicari dalam *database* yang besar untuk pengambilan keputusan diwaktu yang akan datang (Fajar Astuti Hermawati, 2013). Ciri penting dari *data mining* adalah bahwa *volume* data yang sangat besar meskipun ide-ide dari area studi yang berhubungan

tadi dapat diaplikasikan pada masalah-masalah *data mining*, *scalability* yang berkaitan dengan ukuran data menjadi suatu kriteria baru yang penting.

Pengelompokan *data mining* terdiri dari deskripsi, estimasi, prediksi, asosiasi, pengklusteran, dan klasifikasi. Dalam klasifikasi, terdapat target variabel kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang dan pendapatan rendah. Kemudian untuk menemukan pendapatan *seorang pegawai*, dipakai cara klasifikasi dalam *data mining*. Secara umum proses klasifikasi dapat dilakukan dalam dua tahap, yaitu proses belajar dari data pelatihan dan klasifikasi kasus. Pada proses belajar, algoritma klasifikasi mengolah data training untuk menghasilkan sebuah model. Setelah model diuji dan dapat diterima, pada tahap klasifikasi, model tersebut digunakan untuk memprediksi kelas dari kasus baru untuk membantu proses pengambilan keputusan. Decision tree adalah salah satu metode klasifikasi yang paling populer karena mudah untuk diinterpretasi oleh manusia. Konsep dasar algoritma *Decision Tree* adalah mengubah data menjadi pohon keputusan dan aturan-aturan keputusan (*rules*) (Defiyanti dan Pardede 2010).

Konsep data dalam pohon keputusan adalah sebagai berikut (Muhammad Syahril, 2011):

1. Data dinyatakan dalam bentuk tabel dengan atribut dan record.
2. Atribut menyatakan suatu parameter yang dibuat sebagai kriteria dalam pembentukan tree. Salah satu atribut merupakan atribut yang menyatakan data solusi per-item data yang disebut dengan target atribut.
3. Atribut memiliki nilai-nilai yang dinamakan dengan *instance*. Misalkan atribut berat badan mempunyai *instance* berupa *overweight*, *average*, dan *underweight*.

Algoritma *Decision Tree C4.5* mempunyai input berupa *trainingsamples* dan *samples*. *Training samples* berupa data contoh yang akan digunakan untuk membangun sebuah *tree* yang telah diuji kebenarannya. Sedangkan *samples* merupakan *field* data yang nantinya akan kita gunakan sebagai parameter dalam melakukan klasifikasi data.

Secara umum algoritma *Decision Tree C4.5* untuk membangun pohon keputusan adalah sebagai berikut (Muhammad Syahril, 2011):

1. Pilih atribut sebagai akar.
2. Buat cabang untuk tiap-tiap nilai.
3. Bagi kasus dalam cabang.
4. Ulangi proses untuk tiap cabang memiliki kelas yang sama.

3. METODE PENELITIAN

Secara umum terdapat beberapa tahapan dalam penelitian ini dengan mengikuti pola umum penelitian ilmiah seperti yang terlihat pada Gambar 2.



Gambar 1. Kerangka Kerja Penelitian

4. HASIL DAN PEMBAHASAN

Didalam Teknik *Classification* keluaran dari setiap data yang dijadikan target atau *class* harus berupa bilangan bulat atau diskrit.

Data mahasiswa pada Universitas Pembangunan Pancabudi Medan dalam bentuk *Microsoft excel*. Data Mahasiswa yang telah dikumpulkan kemudian ditentukan atribut, dilakukan *filtering* atribut lalu ditentukan *variabel input* dan *target*, selanjutnya membuat pohon keputusan dengan *software*

RapidMiner. Data Mahasiswa yang diambil dari tahun 2010 sampai tahun 2015 dengan status aktif status aktif dan *dropout*. Universitas Pembangunan Panca Budi tidak mengenal istilah *dropout*, tetapi *student losses*

Variabel yang dimiliki yaitu sebagai berikut :

1. NPM
2. Jenis Kelamin
3. Agama
4. Usia
5. Tempat Tinggal
6. Status

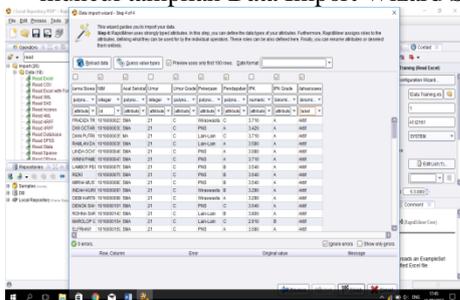
Langkah-langkah menggunakan *software* aplikasi RapidMiner 5.2 sebagai berikut :

1. Data Mahasiswa yang berisi variabel dari kondisi dan atribut keputusan atau target untuk mengetahui faktor mahasiswa *drop out* yang disimpan dengan *microsoft excel* yang berisi kasus untuk menghasilkan informasi berupa rule.

No	NPM	Nama Siswa	ILM	Asal Sekolah	Pekerjaan	Pendidikan	PM	Status Mahasiswa
1	501000001	SAHARA SULTA	SMA	Widayawati		1000000	1,00	Aktif
2	501000002	SIMAMBAKAL	SMA	PMG		2000000	1,00	Aktif
3	501000003	SINAMPURITA	SMA	Lain-Lain		2400000	0,90	Aktif
4	501000004	NI NIA NARA NARA	SMA	Widayawati		3000000	1,00	Aktif
5	501000005	SINAMPURITA	SMA	Widayawati		3000000	1,00	Aktif
6	501000006	SINAMPURITA	SMA	Widayawati		2000000	1,00	Aktif
7	501000007	SINAMPURITA	SMA	Widayawati		3000000	1,00	Aktif
8	501000008	SINAMPURITA	SMA	Widayawati		3000000	1,00	Aktif
9	501000009	SINAMPURITA	SMA	PMG		2000000	1,00	Aktif
10	501000010	SINAMPURITA	SMA	Lain-Lain		2000000	1,00	Aktif
11	501000011	SINAMPURITA	SMA	PMG		4000000	1,00	Aktif
12	501000012	SINAMPURITA	SMA	PMG		3000000	1,00	Aktif
13	501000013	SINAMPURITA	SMA	Widayawati		3000000	1,00	Aktif
14	501000014	SINAMPURITA	SMA	PMG		2000000	1,00	Aktif
15	501000015	SINAMPURITA	SMA	Lain-Lain		2000000	1,00	Aktif
16	501000016	SINAMPURITA	SMA	PMG		4000000	1,00	Aktif
17	501000017	SINAMPURITA	SMA	PMG		3000000	1,00	Aktif
18	501000018	SINAMPURITA	SMA	PMG		4000000	1,00	Aktif
19	501000019	SINAMPURITA	SMA	PMG		2000000	1,00	Aktif
20	501000020	SINAMPURITA	SMA	PMG		2000000	1,00	Aktif
21	501000021	SINAMPURITA	SMA	Widayawati		2000000	1,00	Aktif
22	501000022	SINAMPURITA	SMA	Widayawati		2000000	1,00	Aktif
23	501000023	SINAMPURITA	SMA	PMG		3000000	1,00	Aktif

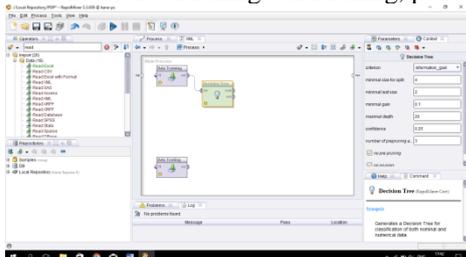
Gambar 2. Contoh data mahasiswa Universitas Pembangunan Panca Budi

2. Data mahasiswa yang telah disimpan dalam *Microsoft excel* dapat dipakai langsung pada aplikasi *software RapidMiner5.2*.
3. Buka *software RapidMiner 5.2*, klik *New*.
4. Kemudian klik *Repositories*, pilih *Import Excel Sheet*.
5. Pilih data tes yang kita simpan dalam *microsoft excel*, klik *Next*.
6. Muncul tampilan *Data Import Wizard Step* klik *Next*, pada *Annotation* pilih *Name*, klik *Next* sehingga muncul tampilan *Data Import Wizard Step 4*.



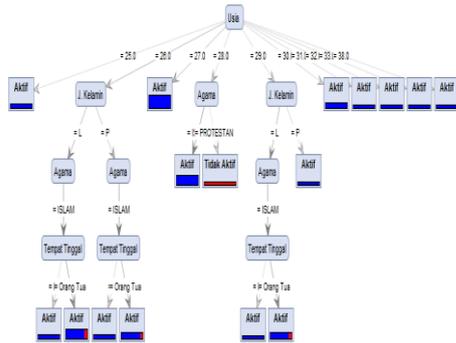
Gambar 3. Import Wizard Step 4.

7. Pada *Data Import Wizard Step 4*.
8. Klik *Next*. Pilih *NewLocalRepository* klik data import. Kemudian klik *Finish*.
9. Drag data pada layar *process*, kemudian klik *Operator*, *Modeling*, *Classification and Regression*, *Tree Induction*, *Decision Tree*. Kemudian drag ke layar *Process*.
10. Klik *No Pre Pruning* dan *Pruning*, pada *criterion* pilih *Information_Gain*.



Gambar 4. Menambahkan Operator *Decision Tree*

11. Klik ikon Run maka akan tampil hasil pohon keputusan seperti gambar 7.



Gambar 5. Hasil Pohon Keputusan Dengan RapidMiner

5. KESIMPULAN

Data mahasiswa yang diperoleh dari BPAA. Dengan menggunakan metode Klasifikasi C.45 dalam proses menganalisa mahasiswa *drop out* berdasarkan menghasilkan informasi baru yang akurat. Pemilihan variabel data Mahasiswa yaitu atribut kondisi dan atribut keputusan yang akan digunakan sangat mempengaruhi *rule* atau *knowledge* yang dihasilkan, jika data yang diproses semakin banyak maka semakin banyak *rule* atau pengetahuan yang diperoleh. *Decision tree* Algoritma C4.5 dapat memberikan informasi atau pengetahuan berupa *rule* yang mudah dipahami oleh masyarakat luas karena digambarkan dengan pohon keputusan. Atribut yang digunakan terdiri dari Jenis kelamin, Agama, Usia, Tempat Tinggal, Status. Atribut yang paling berpengaruh terhadap mahasiswa yang mengalami *dropout* adalah dari variabel Agama. Informasi yang dihasilkan dapat membantu manajemen kampus dalam melakukan evaluasi terhadap mahasiswa *drop out*.

6. REFERENSI

- Andriani Anik. 2012. Penerapan Algoritma C4.5 Pada Program Klasifikasi Mahasiswa Dropout. *Seminar Nasional Matematika*.
- As'ad Bahrawi. 2016. Prediksi Keputusan Menggunakan Metode Klasifikasi *Naïve Bayes*, *One-R*, dan *Decision Tree*. *Jurnal Penelitian Komunikasi dan Opini Publik*. 20(1): 1-10.
- Defiyanti S, Pardede DLC. 2010. Perbandingan Kinerja Algoritma ID3 dan C4.5 dalam Klasifikasi Spam-Mail. Gunadarma.
- Dhika H, Destiawati F, Fitriansyah A. 2016. Implementasi Algoritma C4.5 Terhadap Kepuasan Pelanggan. *Prosiding SNaPP2016 Sains dan Teknologi*. 6(1): 16-22.
- Fajar Astuti Hermawati. 2013. *Data Mining*. Penerbit: CV. ANDI OFFSET.
- Jantan H, Hamdan AR, Othman ZA. 2010. *Human Talent Prediction in HRM Using C4.5 Classification Algorithm*. *International Journal on Computer Science and Engineering*. 2(8): 2526-2534.
- Kusrini, Emha Taufiq Luthfi. 2009. *Algoritma Data Mining*. Penerbit: CV. ANDI OFFSET.
- Mabrur AG, Lubis R. 2012. Penerapan Data Mining Untuk Memprediksi Kriteria Nasabah Kredit. *Jurnal Komputer dan Informatika*. 1(1): 53-57.
- Muhammad Syahril. 2011. Konversi Data Training Tentang Penyakit Hipertensi Menjadi Bentuk Pohon Keputusan Dengan Teknik Klasifikasi Menggunakan Tools Rapid Miner 4.1. *Jurnal SAINTIKOM*.
- Panjaitan Febriyanti. 2017. Penerapan *Cloud Computing* Pada SMKN 1 Indralaya Selatan. *Prosiding SNaPP2017 Sains dan Teknologi*. 7(1):
- Swatina L. 2013. Penerapan Algoritma C4.5 Untuk Penentuan Jurusan Mahasiswa. *Jurnal GEMA AKTUALITA*. 2(1): 93-98.