

Association Rule Dengan Algoritma Apriori Untuk Menentukan Karakteristik Profil Calon Mahasiswa Sistem Informasi di Universitas Dian Nuswantoro

Association Rule Using Algorithm Apriori to Determine Characteristics Profiles Of Prospective System Information Student in Dian Nuswantoro University

Isma Atika Farda¹, Fajrian Nur Adnan²

^{1,2}Jurusan Sistem Informasi, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro Semarang
e-mail: ismafarda@gmail.com, fajrian@dsn.dinus.ac.id

Abstrak

Pendidikan merupakan aspek yang penting terutama di perguruan tinggi. Perguruan tinggi dituntut untuk memiliki keunggulan dengan memanfaatkan sumber dayanya, termasuk sumber daya manusia yaitu mahasiswa. Lama studi mahasiswa merupakan faktor yang penting dalam meningkatkan daya saing di dunia pendidikan. Sering terjadi ketidakseimbangan antara jumlah mahasiswa yang diterima dengan jumlah mahasiswa yang lulus tepat waktu. Hal ini disebabkan kurangnya analisa terhadap profil mahasiswa yang lulus tepat waktu sehingga tidak dapat menyaring calon mahasiswa yang berpotensi lulus tepat waktu dan dapat menurunkan daya saing perguruan tinggi dalam dunia pendidikan. Data mining adalah proses penemuan pola yang terdapat dalam suatu data dengan jumlah yang besar. Dengan memanfaatkan data profil mahasiswa dapat menghasilkan karakteristik profil calon mahasiswa yang lulus tepat waktu ketika kuliah. Algoritma yang digunakan adalah apriori, informasi yang dihasilkan berupa nilai support dan confidence dari karakteristik yang terbentuk sebanyak 4 – Itemset yang nantinya dapat digunakan sebagai acuan dalam penerimaan mahasiswa baru di UDINUS dan dapat meningkatkan daya saing dengan perguruan tinggi yang lain.

Kata kunci—Data Mining, Apriori, Karakteristik Profil Calon Mahasiswa

Abstract

Education is an important aspect, especially in college. Universities and colleges are required to have the advantage by exploiting its resources, including human resources such as students. Long study of college students is an important factor in improving competitiveness in the world of education. Often an imbalance between the number of students accepted by the number of students who graduate on time. This is due to a lack of analysis of the profile of students graduate on time so that it can not filter potentially prospective students graduate on time and may reduce the competitiveness of higher education in the world of education. Data mining is the process of discovery of patterns that are present in a large amount of data. By utilizing the student profile data is expected to produce characteristic profiles of students who graduate on time when college. The algorithm used is a priori, the information generated in the form of support and confidence values of the characteristics that form as much as 4 - itemsets that can later be used as a reference in the new admissions in UDINUS and can improve competitiveness with other universities.

Keywords—Data Mining, Apriori, Characteristic Profiles Of Prospective Student

1. PENDAHULUAN

Universitas Dian Nuswantoro (UDINUS) merupakan perguruan tinggi swasta yang cukup terkenal di Jawa Tengah tepatnya di Kota Semarang yang berdiri pada tahun 1990. Jumlah mahasiswa UDINUS saat ini mencapai lebih dari 10.000 mahasiswa dari beragam kota di Indonesia. Kemudian berdasarkan nilai akreditasi Program Studi Teknik Informatika S-1, Sistem Informasi S-1, dan Desain Komunikasi Visual S-1 Fakultas Ilmu Komputer pada BAN-PT tahun 2015 mencapai nilai A. Hal ini tentu saja menjadi kebanggaan bagi UDINUS. Namun kondisi saat ini jumlah mahasiswa yang diterima tidak seimbang dengan jumlah mahasiswa yang lulus. Jika hal ini dibiarkan maka akan mengakibatkan banyak masalah seperti kekalahan dalam persaingan perguruan tinggi yang semakin ketat di Jawa Tengah bahkan Indonesia, turunnya akreditasi program studi, rendahnya kualitas setiap lulusan mahasiswa, serta kurangnya kinerja dari program studi yang bersangkutan. Oleh sebab itu diperlukan suatu teknik yang dapat menganalisis karakteristik profil calon mahasiswa yang berpotensi lulus tepat waktu sehingga masalah tersebut dapat diminimalkan.

Teknik yang tepat untuk menjawab kebutuhan diatas adalah data mining. Salah satu metode data mining adalah *association rules* atau disebut juga *frequent itemset mining*[1]. Algoritma yang termasuk ke dalam *association rule* adalah *Apriori* yang digunakan oleh beberapa peneliti misalnya Harton Rohul Meisa Tambun dan Anofrizen. Penelitian ini menggunakan data profil mahasiswa dan data kelulusan mahasiswa dengan harapan dapat mengetahui informasi mengenai tingkat kelulusan mahasiswa dengan teknik data mining[2]. Kemudian Dewi Setianingsih dan RB Fajria Hakim juga pernah melakukan penelitian menggunakan *apriori* dengan mengolah data bencana tanah longsor di Indonesia untuk mengetahui atribut apa saja yang sering muncul bersamaan dalam kejadian bencana tanah longsor. Hasil yang didapatkan adalah pola hubungan asosiatif yang terjadi dari data tanah longsor di Indonesia pada akhir bulan desember tahun 2014 dan awal bulan januari tahun 2015 dengan nilai support tertinggi yaitu jika terjadi tanah longsor ketika tanah labil maka hujan deras dengan nilai *support* sebesar 0.5227273 dan nilai *confident* sebesar 0.8214286[3].

2. METODE PENELITIAN

Data diperoleh dari UPT Data dan Informasi (PSI) yaitu data mahasiswa program studi Teknik Informatika S-1, Sistem Informasi S-1, dan Desain Komunikasi Visual S-1 yang telah lulus dari tahun 2009 sampai tahun 2012. Atribut yang digunakan dalam penelitian ini antara lain Id mahasiswa, indeks prestasi kumulatif (ipk), angkatan, tahun yudisium, jurusan SMA, asal kota, nilai raport kelas 1 dan 2.

Dari data yang diperoleh diatas akan di proses menggunakan algoritma *apriori* sehingga menghasilkan pola atau *rule* yang dapat digunakan oleh BIAK dalam penyeleksian calon mahasiswa baru di UDINUS dan oleh program studi dalam menentukan strategi marketing yang tepat sasaran. Urutan langkah yang harus dilakukan adalah :

- a. Menentukan batasan *minimum support* dan *minimum confidence* untuk menghasilkan aturan atau *rule* yang diinginkan.
- b. Mendefinisikan data menjadi kombinasi 2 - *Itemset*.
- c. Mencari kombinasi item yang memenuhi syarat *minimum support* yang telah ditentukan.
- d. Menghilangkan *Itemset* yang tidak memenuhi nilai support.
- e. Menghitung nilai *confidence* dari seluruh itemset yang memenuhi nilai support.
- f. Membuat tabel kombinasi 3 - *Itemset* berdasarkan data awal.
- g. Ulangi tahap c sampai e seterusnya sampai 4 - *Itemset*[4].

Validasi untuk data mining *association rules* algoritma *apriori* dapat dilakukan dengan menggunakan *lift / improvement ratio*. Metode tersebut menghasilkan output berupa pengujian dari *rule* yang terbentuk[5].

3. HASIL DAN PEMBAHASAN

3.1 Pengolahan Data

Sebelum melakukan proses data mining menggunakan perhitungan apriori. Perlu dilakukan pengolahan data untuk mendapatkan data yang diinginkan. Seperti yang sudah dijelaskan sebelumnya bahwa data diperoleh dari PSI UDINUS berupa data profil mahasiswa yang telah lulus pada tahun 2009 sampai 2012. Data tersebut didapatkan dalam bentuk excel sehingga tidak dapat langsung digunakan karena data yang dibutuhkan terpisah dari beberapa sheets dan kolom sehingga penggabungan data dalam satu database yang baru perlu dilakukan. Kemudian data tersebut dimasukkan ke dalam database dan diolah sehingga didapatkan hasil yang valid sesuai dengan kebutuhan penelitian. Berikut adalah tahapan yang dilakukan.

3.2 Data Selection

Data selection adalah proses pemilihan data dari sekumpulan data operasional sebelum tahap penggalian informasi dalam KDD dimulai. Dari data yang diperoleh terdapat beberapa atribut yang tidak dibutuhkan dalam proses perhitungan data mining sehingga dilakukan proses pemilihan atau penyeleksian data seperti nama SMA, jenis kelamin, nilai raport kelas 3, pendapatan orang tua, dan pekerjaan orang tua. Atribut yang digunakan dalam penelitian ini adalah program studi ketika kuliah, angkatan, ipk, tahun yudisium, kota asal, jurusan SMA, nilai raport kelas 1 dan 2.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	
id	progd	angkatan	ipk	sex	tgl_yudisium	kota_asal	kota_sita	nm_sita	jurusan	rata_nilai_1_1	rata_nilai_1_2	rata_nilai_2_1	rata_nilai_2_2	rata_nilai_3_1	rata_nilai_3_2	hasil_ortu	job_ortu	
14896	A14	2011	3.09	2	2016-07-29	KOTA SEMARANG	KOTA SEMARANG	SMA Teuku Umar	\N	\N	\N	\N	\N	\N	\N	\N	\N	SWASTA
24833	A14	2011	3.19	1	2015-08-11	KOTA SEMARANG	KOTA SEMARANG	SMA Kristen YSKI	IPA	69.25	73.86	73.61	75.53	78.61	75.53	\N	\N	SWASTA
24836	A14	2011	3.3	1	2016-02-26	KOTA SEMARANG	KOTA SEMARANG	SMA Negeri 10	\N	\N	\N	\N	\N	\N	\N	\N	\N	SWASTA
24843	A14	2011	3.24	1	2015-11-05	KABUPATEN PEMALANG	KABUPATEN PEMALANG	SMA NEGERI 1 PETARUKAN	IPS	79.25	75.87	75.85	75.55	77.38	0.00	\N	\N	SWASTA
24847	A14	2011	3.09	2	2015-08-11	KOTA SEMARANG	KOTA SEMARANG	SMA Kesatrian 1	IPA	64.46	64.93	70.53	69.92	75.92	0.00	\N	\N	B U M N
24858	A14	2011	3.27	1	2016-02-26	KABUPATEN GROBOGAN	KABUPATEN GROBOGAN	SMA NEGERI 1 GODONG	IPA	71.37	72.18	75.64	76.45	79.23	0.00	\N	\N	P N S
24875	A14	2011	3.19	2	2015-08-11	KABUPATEN BANJARNEGARA	KABUPATEN BANJARNEGARA	SMA NEGERI 1 BANJARNEGARA	IPA	72.06	73.41	72.97	74.71	81.07	0.00	\N	\N	WIARAUSAHA
24902	A14	2011	3	1	2016-08-08	KABUPATEN KENDAL	KABUPATEN KENDAL	SMK (STM) N 3 KENDAL	TKJ	74.20	74.58	74.00	75.15	77.55	79.52	< 3 juta	\N	P N S
24903	A14	2011	3.13	1	2016-08-08	KABUPATEN WONOSOBO	KABUPATEN WONOSOBO	SMK Grafika Bhatki	\N	\N	\N	\N	\N	\N	\N	3 juta - 5 juta	\N	P N S
24949	A14	2011	3.29	2	2015-08-11	KOTA SEMARANG	KOTA SEMARANG	SMK ST. FRANCISKUS	\N	\N	\N	\N	\N	\N	\N	\N	\N	SWASTA
24991	A14	2011	3.58	2	2015-11-05	KABUPATEN SEMARANG	KABUPATEN SEMARANG	SMA SEDES SAPIENTIAE	IPA	79.69	80.50	80.38	81.23	0.00	0.00	\N	\N	P N S
24999	A14	2011	3.08	1	2015-08-11	KOTA SEMARANG	KOTA SEMARANG	SMA Kristen YSKI	IPS	64.81	67.83	69.23	72.00	0.00	0.00	\N	\N	SWASTA
25017	A14	2011	3.31	1	2015-11-05	KOTA SEMARANG	KOTA SEMARANG	SMA NEGERI 4	IPA	71.43	72.50	76.25	77.00	0.00	0.00	\N	\N	B U M N
25018	A14	2011	3.2	2	2015-08-11	KABUPATEN SEMARANG	KABUPATEN SEMARANG	SMA SINT LOUIS	IPA	70.47	72.37	70.92	72.92	0.00	0.00	\N	\N	SWASTA
25033	A14	2011	3.4	2	2015-11-05	KABUPATEN SEMARANG	KABUPATEN SEMARANG	SMK NEGERI 11 SEMARANG	MULTIMEDIA	78.00	76.13	78.05	79.23	81.13	0.00	\N	\N	SWASTA
25049	A14	2011	3.03	1	2016-11-30	KABUPATEN DEMAK	KABUPATEN DEMAK	\N	\N	\N	\N	\N	\N	\N	\N	\N	\N	P N S
25050	A14	2011	3.11	2	2015-08-11	KABUPATEN PEMALANG	KABUPATEN PEMALANG	SMA NEGERI 1 PETARUKAN	IPS	74.06	75.38	73.77	75.00	76.08	0.00	\N	\N	SWASTA
25054	A14	2011	3.11	2	2016-02-26	KABUPATEN WONOGIRI	KABUPATEN WONOGIRI	SMA NEGERI 1 PETARUKAN	IPS	73.38	74.38	75.85	75.46	76.92	0.00	\N	\N	P N S
25067	A14	2011	3.46	2	2015-11-05	KABUPATEN SEMARANG	KOTA SEMARANG	SMA NEGERI 1 PETARUKAN	\N	\N	\N	\N	\N	\N	\N	3 juta - 5 juta	\N	SWASTA
25074	A14	2011	3.12	1	2016-02-26	KABUPATEN BLORA	KOTA SEMARANG	SMA NEGERI 1 UNGARAN	\N	\N	\N	\N	\N	\N	\N	\N	\N	SWASTA
25102	A14	2011	3.2	2	2016-02-26	KOTA SEMARANG	KOTA SEMARANG	SMA NEGERI 6	IPS	71.87	75.25	74.00	75.61	0.00	0.00	\N	\N	WIARAUSAHA
25112	A14	2011	3.24	1	2015-11-05	KABUPATEN SEMARANG	KABUPATEN SEMARANG	SMA NEGERI 1 UNGARAN	IPS	70.44	72.12	77.54	76.40	80.15	0.00	\N	\N	SWASTA

Gambar 1 Penghapusan Atribut Yang Tidak Digunakan

Gambar 1 diatas merupakan penghapusan beberapa atribut dari data utama yang tidak digunakan dalam proses data mining dikarenakan memiliki beberapa alasan yang berkaitan dengan proses perhitungan algoritma apriori.

A	B	C	D	E	F	G	H	I	J	K
id	progd	angkatan	ipk	tgl_yudisium	kota_asal	jurusan	rata_nilai_1_1	rata_nilai_1_2	rata_nilai_2_1	rata_nilai_2_2
14896	A14	2011	3.09	2016-07-29	KOTA SEMARANG	\N	\N	\N	\N	\N
24833	A14	2011	3.19	2015-08-11	KOTA SEMARANG	IPA	69.25	73.86	73.61	75.53
24836	A14	2011	3.3	2016-02-26	KOTA SEMARANG	\N	\N	\N	\N	\N
24843	A14	2011	3.24	2015-11-05	KABUPATEN PEMALANG	IPS	79.25	75.87	75.85	75.55
24847	A14	2011	3.09	2015-08-11	KOTA SEMARANG	IPA	64.46	64.93	70.53	69.92
24858	A14	2011	3.27	2016-02-26	KABUPATEN GROBOGAN	IPA	71.37	72.18	75.64	76.45
24875	A14	2011	3.19	2015-08-11	KABUPATEN BANJARNEGARA	IPA	72.06	73.41	72.97	74.71
24902	A14	2011	3	2016-08-08	KABUPATEN KENDAL	TKJ	74.20	74.58	74.00	75.15
24903	A14	2011	3.13	2016-08-08	KABUPATEN WONOSOBO	\N	\N	\N	\N	\N
24949	A14	2011	3.29	2015-08-11	KOTA SEMARANG	\N	\N	\N	\N	\N
24991	A14	2011	3.58	2015-11-05	KABUPATEN SEMARANG	IPS	79.69	80.50	80.38	81.23
24999	A14	2011	3.08	2015-08-11	KOTA SEMARANG	IPA	64.81	67.83	69.23	72.00
25017	A14	2011	3.31	2015-11-05	KOTA SEMARANG	IPA	71.43	72.50	76.25	77.00
25018	A14	2011	3.2	2015-08-11	KABUPATEN SEMARANG	IPA	70.47	72.37	70.92	72.92
25033	A14	2011	3.4	2015-11-05	KABUPATEN SEMARANG	MULTIMEDIA	78.00	76.13	78.05	79.23
25049	A14	2011	3.03	2016-11-30	KABUPATEN DEMAK	\N	\N	\N	\N	\N
25050	A14	2011	3.11	2015-08-11	KABUPATEN PEMALANG	IPS	74.06	75.38	73.77	75.00
25054	A14	2011	3.11	2016-02-26	KABUPATEN WONOGIRI	IPS	73.38	74.38	75.85	75.46
25067	A14	2011	3.46	2015-11-05	KABUPATEN SEMARANG	\N	\N	\N	\N	\N
25074	A14	2011	3.12	2016-02-26	KABUPATEN BLORA	\N	\N	\N	\N	\N
25102	A14	2011	3.2	2016-02-26	KOTA SEMARANG	IPS	71.87	75.25	74.00	75.61
25112	A14	2011	3.24	2015-11-05	KABUPATEN SEMARANG	IPS	70.44	72.12	77.54	76.40

Gambar 2 Hasil Proses Data Selection

Pada gambar 2 menunjukkan hasil proses *attribute selection* yang disesuaikan dengan kebutuhan proses penelitian. *Attribute selection* adalah proses penghapusan atribut yang sekiranya tidak diperlukan dalam proses penelitian agar memudahkan dalam pengolahan data dan mengurangi resiko tertumpuknya data karena data yang terlalu banyak sehingga dapat menghemat ruang penyimpanan data.

3.3 Data Cleaning

Setelah melakukan *data selection* tahap selanjutnya adalah *data cleaning*. Pada tahap ini dilakukan proses penghapusan duplikasi data, memeriksa *inkonsistensi* data dan memperbaiki kesalahan pada data. Selain itu juga dilakukan proses *enrichment* yaitu proses memperkaya data dengan data atau informasi yang relevan dan diperlukan untuk KDD[6].

Dalam data profil mahasiswa yang didapat ditemukan sejumlah *record* yang *inkonsisten* sehingga harus dihapus karena tidak dapat digunakan dalam perhitungan. Contoh data yang *inkonsisten* dalam kasus ini yaitu data rata-rata nilai raport kelas 1,2 dan 3. Banyak sekali data yang kosong atau null sehingga tidak dapat digunakan dalam proses perhitungan data mining dan pembuatan program nantinya. Dikatakan data yang *inkonsisten* karena data tersebut tidak memiliki nilai sehingga harus dilakukan *filtering* data pada program untuk menemukan data yang memiliki nilai atau data yang sebenarnya[7]. Terdapat beberapa langkah yang dilakukan dalam proses *cleaning* data profil mahasiswa yaitu sebagai berikut :

1. Melakukan *filter* data untuk menemukan data yang dapat digunakan dalam proses perhitungan data mining dengan menuliskan query sebagai berikut :

```
SELECT id,progdi,provinsi_kota_asal,jurusan_slta,ipk,year(angkatan) as
\"angkatan\",tanggal_yudisium FROM `mahasiswa` where progdi is not null and
provinsi_kota_asal is not null and jurusan_slta is not null and
tanggal_yudisium is not null
```

Hasil proses *filter* data dapat dilihat pada gambar 3 dibawah yang hanya menampilkan data yang memiliki nilai dan dapat digunakan dalam proses perhitungan data mining sehingga dapat menghasilkan rule yang valid.

	A	B	C	D	E	F	G	H	I	J	K
1	id	progdi	angkatan	ipk	tgl_yudisium	kota_asal	jurusan	rata_nilai_1_1	rata_nilai_1_2	rata_nilai_2_1	rata_nilai_2_2
3	24833	A14	2011	3.19	2015-08-11	KOTA SEMARANG	IPA	69.25	73.86	73.61	75.53
5	24843	A14	2011	3.24	2015-11-05	KABUPATEN PEMALANG	IPS	79.25	75.87	75.85	75.55
6	24847	A14	2011	3.09	2015-08-11	KOTA SEMARANG	IPA	64.46	64.93	70.53	69.92
7	24858	A14	2011	3.27	2016-02-26	KABUPATEN GROBOGAN	IPA	71.37	72.18	75.64	76.45
8	24875	A14	2011	3.19	2015-08-11	KABUPATEN BANJARNEGARA	IPA	72.06	73.41	72.97	74.71
9	24902	A14	2011	3	2016-08-08	KABUPATEN KENDAL	TKJ	74.20	74.58	74.00	75.15
12	24991	A14	2011	3.58	2015-11-05	KABUPATEN SEMARANG	IPA	79.69	80.50	80.38	31.23
13	24999	A14	2011	3.08	2015-08-11	KOTA SEMARANG	IPS	64.81	67.83	69.23	72.00
14	25017	A14	2011	3.31	2015-11-05	KOTA SEMARANG	IPA	71.43	72.50	76.25	77.00
15	25018	A14	2011	3.2	2015-08-11	KABUPATEN SEMARANG	IPA	70.47	72.37	70.92	77.92
16	25033	A14	2011	3.4	2015-11-05	KABUPATEN SEMARANG	MULTIMEDIA	78.00	76.13	78.05	79.23
18	25050	A14	2011	3.11	2015-08-11	KABUPATEN PEMALANG	IPS	74.06	75.38	73.77	75.00
19	25054	A14	2011	3.11	2016-02-26	KABUPATEN WONOGIRI	IPS	73.38	74.38	75.85	75.46
22	25102	A14	2011	3.2	2016-02-26	KOTA SEMARANG	IPS	71.87	75.25	74.00	75.61
23	25112	A14	2011	3.24	2015-11-05	KABUPATEN SEMARANG	IPS	70.44	72.12	77.54	76.40
24	25114	A14	2011	3.06	2016-08-08	KOTA SEMARANG	MULTIMEDIA	68.81	69.77	67.64	69.25
25	25118	A14	2011	3.53	2015-08-11	\N	IPA	71.25	72.50	74.90	79.10
26	25140	A14	2011	3.56	2015-08-11	KABUPATEN WONOGIRI	IPA	75.62	78.68	79.00	80.15
28	25175	A14	2011	2.99	2016-02-26	KABUPATEN REMBANG	OTOMOTIF	72.30	72.30	74.25	74.53
30	25213	A14	2011	3.38	2016-11-30	KOTA SEMARANG	IPS	72.57	70.85	72.76	72.47
32	25280	A14	2011	3.29	2015-11-05	KOTA SEMARANG	MULTIMEDIA	75.00	76.66	79.61	79.41
33	25283	A14	2011	2.92	2016-08-08	KABUPATEN PATI	ANIMASI	74.86	74.04	79.02	78.75

Gambar 3 Hasil Proses Filter Data

2. Mengkategorikan jurusan SMA menjadi 3 kategori yaitu IPA, IPS, dan lain-lain. Hal ini dilakukan karena jumlah jurusan selain IPA dan IPS terlalu banyak sehingga untuk memperkecil cakupan dibuat kategori yaitu lain-lain. Berikut adalah *query* untuk mengkategorikan jurusan.

```
if($Jurusan == 'IPA')$NewJurusan='IPA';
else if($Jurusan == 'IPS')$NewJurusan='IPS';
else if($Jurusan == '')$NewJurusan='-';
else $NewJurusan='lain-lain';
```

3. Mengkategorikan asal kota mahasiswa ke dalam wilayah provinsinya sehingga cakupannya menjadi lebih luas dan dapat lebih dipahami.
4. Mengkategorikan rata-rata nilai raport kelas 1 dan 2
Sebelum melakukan proses perhitungan data mining menggunakan perhitungan *apriori* perlu dilakukan pengkategorian nilai raport untuk memudahkan dalam proses perhitungan *apriori*. Adapun pengkategorian nilai raport kelas 1 dan 2 dapat dilihat pada tabel 1 berikut :

Tabel 1 Kriteria Rata-Rata Nilai Raport

No	Nama Kriteria	Kriteria
1	≥ 85	A
2	$75 - < 85$	B

Kategori yang digunakan untuk mendefinisikan rata-rata nilai raport adalah A jika ≥ 85 dan B jika rata-rata nilai raport antara 75– 84. Dengan adanya pengkategorian tersebut selain memudahkan dalam perhitungan data mining juga dapat memudahkan dalam membaca rule yang nantinya akan terbentuk. Berikut adalah query untuk mengkategorikan nilai raport menjadi kriteria yang telah ditentukan.

```

$Nhuruf1=($row['rata_nilai_1_1']+$row['rata_nilai_1_2']+$row['rata_nilai_2_1']+$row
['rata_nilai_2_2']/4;
/*Mencari Nilai Huruf Grade dari Nilai total */
if ($Nhuruf1 >= 85) $grade = "A";
elseif ($Nhuruf1 < 85 AND $Nhuruf1 > 75) $grade = "B";
else $grade = "-";

```

Hasil pengkategorian nilai raport sesuai dengan query diatas dapat dilihat pada gambar 4 dibawah pada kolom nilai huruf.

5. Membuat *labeling* untuk dijadikan acuan dalam proses perhitungan data mining nantinya. *Labeling* dilakukan dengan tujuan untuk mendapatkan status kelulusan tiap mahasiswa yaitu dengan menghitung selisih antara angkatan dan tahun yudisium dengan query sebagai berikut :

```
(year(Tanggal_Yudisium)-year(Angkatan))
```

Dari query tersebut didapatkan lama studi mahasiswa kemudian mencari status kelulusan tiap mahasiswa dengan query sebagai berikut:

```

if ($TampilLama == 4 AND $NewIPK > 3) $status = "SUKSES";
else if ($TampilLama > 4 AND $NewIPK < 3) $status = "TIDAK SUKSES";
else if ($TampilLama == 4 AND $NewIPK < 3) $status = "POTENSIAL";
else $status = "POTENSIAL";

```

Status kelulusan dibagi menjadi 3 kategori antara lain sukses, potensial, dan tidak sukses. Apabila lama studisama dengan 4 dan ipk ≥ 3 maka mahasiswa tersebut sukses, apabila lama studi sama dengan 4 dan ipk < 3 maka mahasiswa tersebut potensial, apabila lama studi lebih dari 4 dan ipk ≥ 3 maka mahasiswa tersebut potensial, apabila lama studi lebih dari 4 dan ipk < 3 maka mahasiswa tersebut tidak sukses.

Gambar 3.4 dibawah menunjukkan hasil proses *labeling* data profil mahasiswa dan didapatkan kolom baru yang merupakan hasil selisih antara angkatan dan tahun yudisium. Hal tersebut dilakukan guna untuk mengetahui mahasiswa yang sukses, potensial atau tidak sukses yang nantinya dapat digunakan sebagai perbandingan dalam proses perhitungan lebih lanjut.

Batasan *minimum support* dan *minimum confidence* dalam penelitian ini adalah 40% sehingga rule akan terbentuk apabila terdapat kombinasi dengan nilai *support* dan *confidence* sama dengan 40 % atau lebih dan kombinasi yang memiliki nilai *support* dan *confidence* dibawah 40% maka kombinasi tersebut tidak akan ditampilkan.

Data Mentah					Data Kategori	
No	ID MHS	Progdi	Provinsi_kota ASAL	Jurusan SLTA	Status	N huruf
1	31043	A11	Jawa Tengah	IPA	SUKSES	B
2	31076	A11	Jawa Tengah	IPA	SUKSES	B
3	31080	A11	Jawa Tengah	IPA	SUKSES	-
4	31081	A11	Jawa Tengah	IPA	SUKSES	-
5	31120	A11	Jawa Tengah	IPS	SUKSES	B
6	31141	A11	Jawa Tengah	lain-lain	SUKSES	A
7	31152	A11	Jawa Tengah	IPA	SUKSES	-
8	31155	A11	Jawa Tengah	IPA	SUKSES	-
9	31166	A11	Jawa Tengah	IPA	SUKSES	B
10	31236	A11	Jawa Tengah	IPA	SUKSES	B
11	31237	A11	Jawa Tengah	IPA	SUKSES	B
12	31251	A11	Jawa Tengah	IPA	SUKSES	-
13	31259	A11	Jawa Tengah	IPA	SUKSES	-
14	31262	A11	Jawa Tengah	lain-lain	SUKSES	B
15	31268	A11	Jawa Tengah	lain-lain	SUKSES	B
16	31272	A11	Jawa Tengah	IPA	SUKSES	B
17	31278	A11	Jawa Tengah	lain-lain	SUKSES	B
18	31283	A11	Jawa Tengah	IPA	SUKSES	B
19	31284	A11	Jawa Tengah	IPA	SUKSES	B
20	31296	A11	Jawa Tengah	IPA	SUKSES	B
21	31303	A11	Jawa Tengah	IPS	SUKSES	B
22	31319	A11	Jawa Tengah	IPA	SUKSES	-
23	31335	A11	Jawa Tengah	IPA	SUKSES	B
24	31336	A11	Jawa Tengah	IPA	SUKSES	B

Gambar 4 Hasil Proses Labeling

3.1 Implementasi Metode Apriori

Proses perhitungan apriori dilakukan pada data sejumlah 2125 data. Namun pada perhitungan manual tidak semua data tersebut akan dihitung, melainkan data sampel yang diambil dari 2125 data dalam database. Berikut merupakan implementasi perhitungan manual yang dilakukan pada program studi Sistem Informasi S-1.

Tabel 2 Tabel Sampel Profil Mahasiswa

No	Provinsi	Jurusan	Nilai Raport SMA	Status Kelulusan
1.	Jawa Tengah	IPS	A	Sukses
2.	Jawa Tengah	Lain – lain	B	Potensial
3.	Jawa Tengah	IPS	B	Sukses
4.	Sulawesi Selatan	IPA	A	Potensial
5.	Jawa Tengah	IPA	A	Sukses
⋮	⋮	⋮	⋮	⋮
82.	Banten	IPA	A	Sukses

Tabel 2 menunjukkan data profil mahasiswa yang kemudian akan dibuat calon 2 - *Itemset* yang masing-masing atribut akan berpasangan dengan status kelulusan. Namun sebelumnya harus ditentukan terlebih dahulu batasan untuk frekuensi kemunculan *itemset*, dalam penelitian ini batasan frekuensi kemunculan *itemset* yang berupa *minimum support* dan *minimum confidence* adalah 0,4 atau 40%. Kombinasi calon 2 - *Itemset* ini hanya dilakukan pada atribut jurusan dengan status kelulusan dan nilai raport dengan status kelulusan sedangkan untuk provinsi dengan status kelulusan tidak dibuat kombinasi calon 2 – *Itemset* karena berdasarkan

data yang didapat mayoritas mahasiswa yang lulus berasal dari provinsi Jawa Tengah sehingga relasi dari provinsi lain sangat sedikit.

Untuk menghitung nilai *support* dari kedua item dapat dilakukan dengan rumus sebagai berikut :

$$Support(A, B) = \frac{\sum \text{Transaksi mengandung } A \text{ dan } B}{\sum \text{Transaksi}}(3,1)$$

Dari rumus *support* diatas didapatkan hasil dari kombinasi calon 2 - *Itemset* dengan nilai *support* seperti pada tabel 3 berikut.

Tabel 3 Kombinasi Calon 2 – Itemset

No.	Karakteristik	Jumlah	Support
1.	IPA, Sukses	33	40,24 %
2.	IPA, Potensial	2	2,44 %
3.	IPS, Sukses	26	31,71 %
4.	IPS, Potensial	2	2,44 %
5.	Lain - lain, Sukses	13	15,85 %
6.	Lain - lain, Potensial	3	3,66 %
7.	B, Sukses	45	54,88 %
8.	B, Potensial	2	2,44 %

Karena sebelumnya sudah ditetapkan *minimum support* adalah 40 % maka *itemset* yang memiliki nilai *support* kurang dari 40 % akan dihilangkan dan didapatkan hasil kombinasi 2 - *Itemset* seperti tabel 4 berikut.

Tabel 4 Hasil Kombinasi 2 – Itemset

No.	Karakteristik	Jumlah	Support
1.	IPA, Sukses	33	40,24 %
2.	B, Sukses	45	54,88 %

Tabel 4 merupakan hasil dari kombinasi 2 – *Itemset* yang memiliki frekuensi kemunculan sama dengan 40% atau lebih dari 40% yang hanya menghasilkan 2 kombinasi, kemudian selanjutnya akan dihitung nilai *confidence* nya berdasarkan tabel 4 diatas. Untuk menghitung nilai *confidence* dapat dilakukan dengan rumus sebagai berikut :

$$Confidence P(B|A) = \frac{\sum \text{Transaksi mengandung } A \text{ dan } B}{\sum \text{Transaksi mengandung } A}$$

Dari rumus *confidence* tersebut didapatkan tabel yang berisi nilai *confidence* sebagai berikut :

Tabel 5 Hasil Confidence 2 – Itemset

No.	Karakteristik	Jumlah	Confidence
1.	IPA, Sukses	33	33/35 = 94,29 %
2.	B, Sukses	45	45/47 = 95,74 %

Setelah didapatkan kombinasi 2 – *Itemset* langkah selanjutnya adalah membuat calon 3 – *Itemset*. Namun dalam hal ini pembuatan calon 3 – *Itemset* tidak berdasarkan hasil kombinasi 2 – *Itemset* dikarenakan pada calon kombinasi 2 – *Itemset* atribut provinsi tidak dipasangkan dengan status kelulusan sedangkan pada calon 3 – *Itemset* atribut provinsi dipasangkan status kelulusan begitu juga dengan atribut yang lain. Dalam pembuatan 3 – *Itemset* terdapat 3 pasangan yang harus dicari. Pertama adalah provinsi, nilai, dan status kelulusan. Yang kedua

adalah provinsi, jurusan, dan status kelulusan. Dan yang ketiga adalah nilai, jurusan, dan status kelulusan. Setelah semua atribut tersebut dipasang-pasangkan maka dicari nilai *support* nya untuk mengetahui kombinasi yang memenuhi batasan *minimum support*. Berikut perhitungan calon kombinasi 3 – Itemset yang dilakukan.

Tabel 6 Kombinasi Calon 3 – Itemset

No.	Karakteristik	Jumlah	Support
1.	Jawa Tengah,IPA,Sukses	33	40,24 %
2.	Jawa Tengah,IPA,POTENSIAL	2	2,44 %
3.	Jawa Tengah,IPS,SUKSES	21	25,61 %
4.	Jawa Tengah,IPS,POTENSIAL	2	2,44 %
5.	Jawa Timur,IPS,SUKSES	1	1,22 %
6.	Jawa Barat,IPS,SUKSES	1	1,22%
7.	DIY,IPS,SUKSES	1	1,22 %
8.	Kalimantan Timur,IPS,SUKSES	1	1,22 %
9.	Banten,IPS,SUKSES	1	1,22 %
10.	Jawa Tengah,LAIN-2,SUKSES	13	15,85 %
11.	Jawa Tengah,LAIN-2,POTENSIAL	3	3,66 %
12.	Jawa Tengah,B,SUKSES	42	51,22 %
13.	Jawa Tengah,B,POTENSIAL	2	2,44 %
14.	Kalimantan Timur,B,SUKSES	1	1,22 %
15.	Banten,B,SUKSES	1	1,22 %
16.	B, IPA,SUKSES	23	28,05 %
17.	B, IPS,SUKSES	14	17,07 %
18.	B, IPS,Potensial	1	1,22 %
19.	B, Lain-lain,SUKSES	8	9,76 %
20.	B, Lain-lain,Potensial	1	1,22 %

Tabel 6 merupakan seluruh calon kombinasi 3 – *Itemset* yang merupakan kombinasi dari provinsi, nilai raport, jurusan, dan status kelulusan. Dari 20record yang terbentuk menghasilkan kombinasi 3 – *Itemset* seperti pada tabel 7 berikut.

Tabel 7 Hasil Kombinasi 3 – Itemset

No.	Karakteristik	Jumlah	Support	Confidence
1.	Jawa Tengah, IPA, Sukses	33	40,24 %	33/35 = 94,29 %
2.	Jawa Tengah, B, Sukses	42	51,22 %	42/44 = 95,45 %

Tabel 7 menunjukkan hasil dari kombinasi 3 - *Itemset* yang didapatkan 2 kombinasi berdasarkan batasan *minimum support* yang ditentukan sebelumnya. Kemudian kombinasi tersebut akan dihitung nilai *confidencenya* seperti tabel 8 berikut

Setelah didapatkan kombinasi 3 – *Itemset* seperti tabel 7 diatas langkah selanjutnya adalah membuat calon kombinasi 4 – *Itemset*. Pembuatan calon kombinasi 4 – *Itemset* berdasarkan kombinasi 3 – *Itemset*. Setelah semua atribut dipasang-pasangkan maka dicari nilai *support* nya untuk mengetahui kombinasi yang memenuhi batasan *minimum support*.

Tabel 8 Kombinasi Calon 4 – Itemset

No.	Karakteristik	Jumlah	Support
1.	Jawa Tengah,IPA,B,SUKSES	23	28,05 %
2.	Jawa Tengah,IPS,B,SUKSES	11	13,41 %

3.	Kalimantan Timur,IPS,B,SUKSES	1	1,22 %
4.	Banten,IPS,B,SUKSES	1	1,22 %
5.	Jawa Tengah,LAIN-2,B,SUKSES	8	9,76 %
6.	Jawa Tengah,LAIN-2,B,POTENSIAL	1	1,22 %

Berdasarkan tabel 8 menunjukkan bahwa tidak ada kombinasi yang memenuhi batasan *support* yang telah ditentukan sehingga tidak ada kombinasi 4 – *Itemset* yang dihasilkan.

Setelah semua aturan pola frekuensi tinggi ditemukan, maka langkah selanjutnya adalah mencari aturan asosiasi yang memenuhi syarat minimum untuk confidence dengan menghitung *confidence*, aturan asosiatif $A \rightarrow B$ [7].

Dari tabel 3 sampai 7 yang telah ditemukan, maka dapat terlihat besarnya support dan confidence dari masing-masing jumlah item set dari calon aturan asosiasi seperti dalam tabel berikut.

Tabel 9 Tabel Final Aturan Asosiasi

Aturan	Support	Confidence	Support x Confidence
Jika jurusan SMA IPA, maka sukses	$33/82 = 0.40243 = 40,24\%$	$33/35 = 0.94286 = 94,29\%$	38%
Jika nilai raport B, maka sukses	$45/82 = 0.54878 = 54,88\%$	$45/47 = 0.95745 = 95,74\%$	53%
Jika provinsi jawa tengah dan jurusan IPA, maka sukses	$33/82 = 0.40243 = 40,24\%$	$33/35 = 0.94285 = 94,29\%$	38%
Jika provinsi jawa tengah dan nilai raport B, maka sukses	$42/82 = 0.51219 = 51,22\%$	$42/44 = 0,95454 = 95,45\%$	49%

Tabel 9 merupakan penjabaran tabel final aturan asosiasi menjadi sebuah *rule* yang nantinya digunakan sebagai acuan dalam penerimaan mahasiswa baru di UDINUS.

3.2 Evaluasi dan Validasi

Untuk membuktikan keakuratan atau kevalidan kombinasi yang terbentuk yaitu dengan menggunakan *lift ratio* [8]. *Lift ratio* dapat dihitung dengan menggunakan rumus berikut :

$$Lift\ ratio = \frac{Support(A \cap B)}{Support(A) * Support(B)}$$

Apabila *lift ratio* > 1 maka kombinasi tersebut benar-benar valid dan ada manfaatnya. Dari hasil kombinasi 2 – *Itemset* dan 3 – *Itemset* yang terbentuk didapatkan nilai *lift ratio* seperti dibawah ini [9].

Tabel 10 Validasi menggunakan Lift Ratio

Aturan	Lift Ratio
Jika jurusan SMA IPA, maka sukses	$0,40243/0,37477 = 1,0738$
Jika nilai raport B, maka sukses	$0.54878 /0,50327 = 1,0904$
Jika provinsi jawa tengah dan jurusan IPA, maka sukses	$0.40243 /0,35192 = 1,1435$
Jika provinsi jawa tengah dan nilai raport B, maka sukses	$0.51219 /0,47258 = 1,0838$

Dapat dilihat pada tabel 10 diatas adalah hasil dari *lift ratio* dari masing-masing kombinasi yang dihasilkan dan semuanya memiliki nilai *lift ratio* lebih dari 1 sehingga karakteristik tersebut dapat dikatakan valid atau akurat dan dapat dimanfaatkan sebagai acuan dalam penerimaan mahasiswa baru di UDINUS sehingga nantinya dapat menyaring mahasiswa baru yang berpotensi lulus tepat waktu dan berprestasi.

4. KESIMPULAN

Dari penelitian yang telah dilakukan dapat ditarik kesimpulan seperti berikut :

1. Penerapan data mining dapat dimanfaatkan untuk mencari karakteristik profil calon mahasiswa yang berpotensi lulus tepat waktu dengan menggunakan algoritma *apriori*. Data yang digunakan adalah data mahasiswa yang telah lulus dengan atribut provinsi, jurusan SMA, nilai raport, ipk, angkatan, dan yudisium. Hasil dari pengolahan data tersebut menghasilkan 4 *rule* atau karakteristik antara lain jika jurusan SMA IPA maka sukses dengan nilai *confidence* 94,29% jika nilai raport B maka sukses dengan nilai *confidence* 95,74%, jika provinsi jawa tengah dan jurusan IPA maka sukses dengan nilai *confidence* 94,29%, jika provinsi jawa tengah dan nilai raport B maka sukses dengan nilai *confidence* 95,45%.
2. Dengan menggunakan *lift / improvement ratio* sebagai validasi *association rule* semua *rule* atau karakteristik yang terbentuk memiliki nilai *lift ratio* > 1 sehingga keakuratannya tidak dapat diragukan lagi dan memiliki manfaat dalam penelitian ini.
3. Semakin tinggi nilai *support* dan nilai *confidence* yang dihasilkan maka semakin kuat nilai hubungan antar atribut. Karakteristik tersebut dapat dijadikan acuan dalam penerimaan mahasiswa baru karena telah mengetahui karakteristik yang cocok untuk masing-masing program studi sehingga dapat menyaring calon mahasiswa yang berpotensi lulus tepat waktu ketika kuliah dan berprestasi dengan mendapatkan IPK lebih dari sama dengan 3.

5. SARAN

Saran yang dapat diberikan dalam penelitian ini adalah :

1. Hasil penelitian ini dapat dimanfaatkan secara optimal oleh BIAK UDINUS dan dapat membantu dalam pengambilan keputusan berkaitan dengan penerimaan mahasiswa baru di UDINUS sehingga dapat menyeleksi calon mahasiswa baru yang berpotensi lulus tepat waktu.
2. Pengembangan model data mining agar jauh lebih baik dan lebih *user friendly*.
3. Dapat menggunakan data yang jumlahnya lebih besar untuk menghasilkan *rule* yang lebih akurat.

DAFTAR PUSTAKA

- [1] R. M. Yuli Asriningtias, "Aplikasi Data Mining Untuk Menampilkan Informasi Tingkat Kelulusan Mahasiswa," *Junal Informatika*, vol. 8, p. 839, 2014.
- [2] A. Harton Rohul Meisa Tambun, "Rancang Bangun Aplikasi Data Mining Untuk Menampilkan Informasi Tingkat Kelulusan Mahasiswa Dengan Algoritma Apriori," *Jurnal Rekayasa dan Manajemen Sistem Informasi*, vol. 1, p. 31, 2015.
- [3] R. F. H. Dewi Setianingsih, "Penerapan Data Mining Dalam Analisis Kejadian Tanah Longsor Di Indonesia Dengan Menggunakan Association Rule Algoritma Apriori," *Prosiding Seminar Nasional Matematika dan Pendidikan Matematika UMS*, p. 731, 2015.
- [4] A. D. Nurjoko, "Penerapan Data Mining Menggunakan Association Rule Untuk Mendukung Strategi Pemasaran Calon Mahasiswa Baru (Studi Kasus IBI Darmajaya)," *Jurnal TIM DARMAJAYA*, vol. 01, p. 17, 2015.
- [5] N. M. Huda, "Aplikasi Data Mining Untuk Menampilkan Informasi Tingkat Kelulusan Mahasiswa (Studi Kasus di Fakultas MIPA Universitas Diponegoro)," *Jurnal Teknik Informatika Jurusan MIPA*, p. 1, 2010.
- [6] Y. Zhao, *Association Rule Mining With R**, University Of Canberra, 2016.
- [7] W. A. Triyanto, "Association Rule Mining Untuk Penentuan Rekomendasi Promosi Produk," *Jurnal SIMETRIS*, vol. 5, p. 121, 2014.
- [8] H. S. B. R. Kennedi Tampubolon, "Implementasi Data Mining Algoritma Apriori Pada Sistem Persediaan Alat-Alat Kesehatan," *Informasi dan Teknologi Ilmiah (INTI)*, vol. 1, p. 93, 2013.
- [9] A. F. Fajri, "Implementasi Algoritma Apriori Dalam Menentukan Program Studi Yang Diambil Mahasiswa," *Jurnal Ipteks Terapan*, vol. 10, pp. 81-85, 2016