

Penerapan Algoritma Apriori Untuk Membantu Calon Mahasiswa Dalam Memilih Program Studi Di Fakultas Ilmu Komputer Universitas Dian Nuswantoro

Apriori Algorithm Application For Helping Students In Choosing A Study Program In Faculty Of Computer Science University Dian Nuswantoro

Marshela Dinda Amalia¹, Lalang Erawan²

^{1,2}Fakultas Ilmu Komputer, Universitas Dian Nuswantoro Semarang
e-mail: marshela.dinda@mhs.dinus.ac.id, lalang.erawan@dsn.dinus.ac.id

Abstrak

Menuut Educational Psychologist dari Integrity Development Flexibility (IDF) Irene Guntur, M.Psi., CGA, sebanyak 87% mahasiswa di Indonesia salah jurusan. Sering terjadi ketidakseimbangan antara jumlah mahasiswa yang diterima dengan jumlah mahasiswa yang lulus tepat waktu. Pada data Fakultas Ilmu Komputer Udinus angkatan 2008 – 2012 sebanyak 25% mahasiswa lulus tepat waktu sedangkan mahasiswa yang lulus terlambat sebanyak 75%. Data mining adalah proses penemuan pola yang terdapat dalam suatu data dengan jumlah yang besar. Dengan memanfaatkan data mahasiswa yang sudah lulus maka akan menghasilkan satu set item acuan untuk membantu mahasiswa dalam memilih program studi dan pada penelitian ini calon mahasiswa dapat mengetahui status kelulusannya kelak. Penelitian ini menggunakan minimal support 20% dan minimal confidence 50%. Jika kombinasi itemset tidak memenuhi syarat minimal support dan minimal confidence maka itemset tersebut akan dieliminasi. Hasil yang di peroleh untuk program studi Teknik Informatika satu itemset acuan, Sistem Informasi 2 itemset acuan, Desain Komunikasi Visual 2 itemset acuan, Teknik Informatika-D3 satu itemset acuan, dan broadcasting-D3 satu itemset acuan.

Kata kunci : Algoritma Apriori, prediksi, Mahasiswa, Rekomendasi

Abstract

According Educational Psychologist from Integrity Development Flexibility (IDF) Irene Guntur, M.Psi., CGA, as many as 87% of students in Indonesia are wrong majors. The factor of the graduating student is one of the majors. There is often an imbalance between the number of students received and the number of students who graduate on time. In the data of the Faculty of Computer Science Udinus class of 2008 to 2012 as many as 25% of students graduated exactly while the passing of students is late 75%. Data mining is the process of finding patterns in a large number of data. By utilizing the data of students who have graduated it will produce a set item reference to help mahasiswa in choosing a course of study and in this study prospective mahasiswa can know the status of his graduation later. This research has minimum support 20% and minimum confidence 50%. If the combination of itemset is not qualified it will be eliminated. The results obtained are Informatics Engineering has one reference itemset, Information System has 2 reference itemset, Visual Communication Design has 2 reference itemset, Informatics Engineering - D3 has one reference itemset and D3 broadcasting has one reference itemset.

Keywords : Apriori Algorithm, prediction, Student, Recommendation

1. PENDAHULUAN

Badan Pusat Statistik adalah Lembaga Pemerintah NonKementerian yang bertanggung jawab langsung kepada Presiden. Sebelumnya, BPS merupakan Biro Pusat Statistik, yang dibentuk berdasarkan UU Nomor 6 Tahun 1960 tentang Sensus dan UU Nomor 7 Tahun 1960 tentang Statistik. Sebagai pengganti kedua UU tersebut ditetapkan UU Nomor 16 Tahun 1997 tentang Statistik. Berdasarkan UU ini yang ditindaklanjuti dengan peraturan perundangan dibawahnya, secara formal nama Biro Pusat Statistik diganti menjadi Badan Pusat Statistik. Badan Pusat Statistik merilis laporan jumlah angka pengangguran pada Februari 2016 yaitu sebanyak 944.666 mahasiswa. Angka tersebut terdiri dari 249.362 mahasiswa D3 sedangkan sisanya yaitu 695.304 mahasiswa sarjana. [1] Jika masalah pengangguran tidak segera diselesaikan maka angka pengangguran di Indonesia bisa terus meningkat setiap tahunnya.

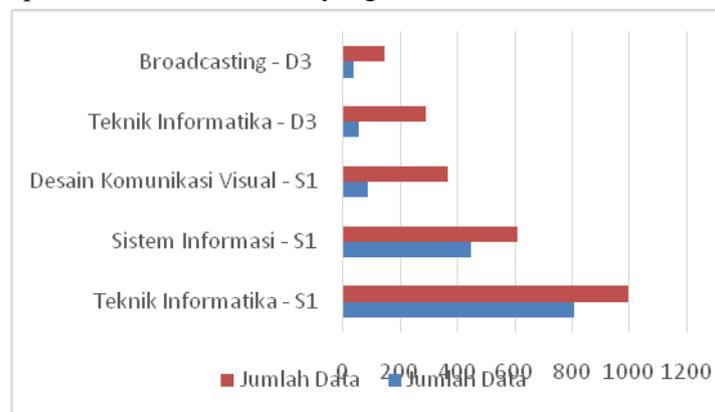
Menurut Educational Psychologist dari Integrity Development Flexibility (IDF) Irene Guntur, M.Psi., Psi., CGA, sebanyak 87% mahasiswa di Indonesia salah jurusan. "Salah jurusan bisa memicu pada pengangguran. Supaya tidak ada pengangguran lagi mahasiswa jangan sampai salah jurusan saat kuliah" [2]

Universitas Dian Nuswantoro Semarang (UDINUS) adalah salah satu perguruan tinggi swasta di Indonesia yang memiliki jumlah mahasiswa yang cukup banyak dan terus meningkat tiap tahunnya. Fakta di lapangan menunjukkan rendahnya jumlah mahasiswa yang dapat lulus tepat waktu. Dalam Peraturan Menteri Pendidikan dan Kebudayaan (Pendidikbud) Republik Indonesia tahun 2014, pada bagian standar kompetensi lulusan pasal 17 disebutkan bahwa untuk memenuhi capaian pembelajaran lulusan, mahasiswa wajib menempuh beban belajar paling sedikit 144 sks dengan masa studi 4 (empat) tahun untuk program sarjana. [3]

Table 1 **Error! No text of specified style in document.** Perbandingan Jumlah Mahasiswa yang Lulus Terlambat dan Lulus Tepat Waktu tahun 2008 – 2012 di Fakultas Ilmu Komputer [4]

Jurusan	Jumlah Data	
	Tepat Waktu	Terlambat
Teknik Informatika - S1	811	996
Sistem Informasi - S1	449	611
Desain Komunikasi Visual - S1	89	370
Teknik Informatika - D3	57	291
Broadcasting - D3	40	148

Tabel di atas adalah tabel jumlah mahasiswa Fakultas Ilmu Komputer yang lulus tepat waktu dan lulus terlambat. Grafik berikut menunjukkan grafik perbandingan antara mahasiswa yang lulus dengan tepat waktu dan mahasiswa yang lulus terlambat.



Gambar 1 Grafik Perbandingan Jumlah Mahasiswa yang Lulus Terlambat dan Lulus Tepat waktu Tahun 2008 – 2012 Fakultas Ilmu Komputer [4]

Terlihat pada gambar 1 mahasiswa yang lulus dengan tepat waktu lebih sedikit dari pada mahasiswa yang lulus terlambat. Faktor mahasiswa lulus dengan terlambat adalah salah jurusan. Mahasiswa yang salah jurusan prestasinya tidak akan optimal, kesulitan memahami materi serta kesulitan memecahkan persoalan sehingga berpengaruh pada nilai indeks prestasi. Hal itu juga menyebabkan mahasiswa banyak mengulang mata kuliah dan akan menyebabkan mahasiswa tidak bisa lulus tepat waktu serta bertambahnya biaya kuliah. [5] Maka dari itu diperlukan informasi untuk calon mahasiswa dalam pemilihan program studi. Informasi yang dimaksud adalah informasi yang dapat memberi rekomendasi untuk pemilihan program studi. Prediksi kecocokan dapat diperoleh berdasarkan hasil perbandingan mahasiswa yang sudah lulus yang memiliki kemiripan data demografis calon mahasiswa. Data mahasiswa yang sudah lulus dipakai untuk membantu calon mahasiswa untuk memilih jurusan. Jika calon mahasiswa yang memiliki kemiripan data dengan mahasiswa yang sudah lulus maka hasil dari penelitian ini dapat dijadikan acuan untuk calon mahasiswa mengambil keputusan dalam memilih program studi.

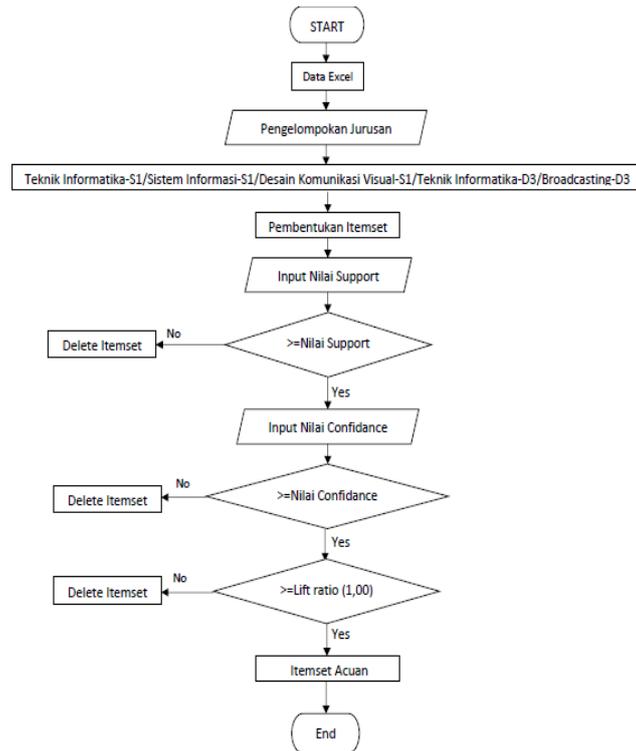
Algoritma Apriori dipilih karena mampu menyimpulkan data dengan proses yang sangat efisien dan dapat mempercepat proses pembentukan kecenderungan pola kombinasi *itemset*. Dengan menggunakan Algoritma Apriori didapatkan hasil berupa aturan yang merupakan kumpulan frequent item set. Dalam penelitian yang sebelumnya ditemukan bahwa minimal support 20% dan minimal confidence 50% dapat menghasilkan perhitungan yang dilakukan optimal. [6] Maka dalam penelitian ini ditetapkan minimal support-nya adalah 20% dan minimal Confidence-nya adalah 50%.

2. METODE PENELITIAN

Perhitungan dengan menggunakan algoritma apriori dalam penelitian ini memerlukan pengolahan data untuk mendapatkan data yang diinginkan. Data primer penelitian adalah data yang diperoleh dari bagian PSI Udinus. Data tersebut adalah data mahasiswa fakultas ilmu komputer Udinus tahun 2008 – 2012 yang sudah lulus tepat waktu. Data diperoleh melalui softcopy dan dikirim melalui email mahasiswa.

Berikut adalah langkah – langkah dalam proses penerapan algoritma apriori dalam penelitian ini:

1. Mencari 1-itemset (himpunan item yang terdiri dari 1 item) dan menghitung nilai supportnya dengan cara men-scan database. Lalu bandingkan nilai support yang didapatkan dengan nilai minimum support yang ditentukan terlebih dahulu.
2. Hanya Itemset termasuk dalam large itemset yang diikuti dalam iterasi selanjutnya.
3. Hasil large itemset pada iterasi pertama (L1) digunakan pada iterasi kedua untuk membentuk kandidat itemset kedua (L2). Hasil large itemset pada iterasi selanjutnya digunakan untuk iterasi selanjutnya menggunakan hasil large item set pada iterasi sebelumnya (Lk-1) untuk membentuk kandidat item set berikut (Lk). Sistem akan menggabungkan (join) Lk-1 dengan Lk-a untuk mendapatkan Lk, seperti pada iterasi sebelumnya hanya Item set yang termasuk dalam large item set yang diikuti dalam iterasi selanjutnya.
4. Itemset hasil join kemudian dihitung supportnya.
5. Proses join akan membentuk kandidat yang akan terus dilakukan hingga himpunan kandidat item setnya null, atau sampai kandidat sudah tidak ada yang bisa dibentuk lagi.
6. Kemudian, association rule yang memenuhi nilai support dan confidence yang telah ditentukan dibentuk dari *frequent item set* tersebut
7. Itemset yang terbentuk harus diuji validitasnya dengan lift ratio



Gambar 2 Proses Algoritma Apriori dalam Penelitian

Analisis asosiasi merupakan dasar dari Teknik data mining lainnya. Analisis pola frekuensi tinggi (*frequent pattern mining*) adalah salah satu tahap Analisis asosiasi yang banyak digunakan oleh para peneliti untuk menghasilkan beberapa algoritma yang efisien. 2 tahap analisis asosiasi adalah :

1. Analisis polanilai penunjang (*support*)

Dalam mencari kombinasi item yang memenuhi syarat minimum dari nilai pendukung yang ada didalam database dilakukan pada tahap ini.

Rumus nilai support :

$$\text{Support}(A) = \frac{\text{jumlahTransaksiMengandungA}}{\text{NilaiTransaksi}}$$

Rumus nilai support 2 item :

$$\text{Support}(A, B) = \text{Support}(A \cap B) = \frac{\text{jumlahTransaksiMengandungA dan B}}{\text{NilaiTransaksi}}$$

2. Perhitungan nilai kepastian (*confidence*)

Aturan asosiasi ditentukan setelah pola frekuensi tinggi ditemukan. Aturan asosiasi yang dicari adalah aturan assosiatif yang memenuhi syarat minimum untuk confidence dengan menghitung confidence aturan asosiasi “ jika A maka B “. Rumus mencari Nilai confidence dari aturan “ jika A maka B “ :

$$\text{confidence} = \text{Support}(A|B) = \frac{\text{jumlahTransaksiMengandungA dan B}}{\text{jumlahTransaksi mengandungA}}$$

Lift ratio adalah parameter penting setelah mendapatkan nilai *support* dan *confidence* dalam *association rules* yang digunakan untuk mengevaluasi sebuah aturan asosiasi. *Lift ratio*

adalah nilai yang menunjukkan kevalidan proses transaksi dan memberikan informasi apakah benar produk A dibeli bersamaan dengan produk B. *Lift Ratio* dapat dihitung dengan rumus sebagai berikut :

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

Data yang digunakan pada penerapan algoritma apriori

Tabel 2 Contoh data yang digunakan dalam penelitian

JURUSAN	NILAI	STATUS
IPA	B	BERHASIL
IPA	C	BERHASIL
IPS	C	BERHASIL
IPS	B	BERHASIL
IPA	B	BERHASIL
IPA	B	BERHASIL
IPA	B	BERHASIL
IPS	C	BERHASIL
IPS	C	BERHASIL
IPA	C	BERHASIL
IPA	B	BERHASIL
IPA	B	BERHASIL
IPA	B	BERHASIL

a. Atribut Jurusan

Atribut ini adalah atribut yang menjelaskan jurusan mahasiswa sebelum kuliah. Atribut ini penting karena untuk mengetahui jurusan sekolah menengah pada mahasiswa yang sudah lulus untuk di jadikan acuan calon mahasiswa untuk memilih program studi.

b. Atribut Nilai Akhir

Atribut ini adalah rata – rata nilai kelas 1 semester 1 dan 2. Cara menghitungnya adalah rata-rata nilai kelas 1 semester 1 dan 2 di jumlahkan lalu di bagi 2. Adapun pengkatagorian rata-rata nilai tersebut adalah :

Tabel 3 Pengkatagorian atribut nilai akhir

No	Nama Kriteria	Kriteria
1.	≥ 85	A
2.	$70 - < 85$	B
3.	$60 - < 70$	C

Kategori yang digunakan untuk mendefinisikan rata-rata nilai raport adalah A jika ≥ 85 , B jika rata- rata nilai raport antara 70 – 84 dan C jika rata – rata raport antara 60 – 69. Dengan adanya pengkategorian tersebut dapat memudahkan perhitungan data mining dan dapat memudahkan untuk membaca rule.

c. Status

Status kelulusan dibagi menjadi 3 kategori yaitu :

Tabel 4 Pengkategorian Status

No	Status	Kuliah Tempuh	IPK Kelulusan
1	Sangat berhasil	3,5 tahun	$\geq 3,0$
2	Berhasil	4 tahun	$\geq 3,0$
3	Potensial	4 tahun	$< 3,0$

Dengan pengkategorian status perhitungan apriori akan lebih mudah dan dapat memudahkan dalam membaca rule.

3. HASIL DAN PEMBAHASAN

Setelah melakukan proses perhitungan dengan menggunakan algoritma apriori pada data mahasiswa yang sudah lulus tahun 2008 – 2012 maka di peroleh hasil :

Tabel 5 Itemset yang terbentuk dalam penelitian

Jurusan	Itemset	Support	Confidance	Lift Ratio
Teknik Informatika - S1	IPA , B , Berhasil	34,38 %	81,76%	1,00
Sistem Informasi - S1	IPS , B	27,07 %	80,52%	1,13
	IPA , Sangat Berhasil	32,46%	51,75%	1,13
Desain Komunikasi Visual - S1	IPA , B , Sangat Berhasil	37,60%	73,44 %	1,04
	IPS , B , Sangat Berhasil	24,80%	83,78%	1,03
Teknik Informatika - D3	IPS , B , Sangat berhasil	20,59%	73,68%	1,13
Broadcasting - D3	Animasi , B , Berhasil	24,33%	83,33%	1,02

Item set yang terbentuk pada tabel 4 dapat di lihat nilai support nya >20% , nilai confidance nya >50% dan nilai lift ratio nya >1,00. Maka Itemset yang dihasilkan dalam penelitian bersifat valid dan dapat dijadikan acuan.

4. KESIMPULAN

Berdasarkan pembahasan yang telah dilakukan dengan algoritma *apriori* dan dilakukannya pengujian maka dalam penelitian ini dapat ditarik kesimpulan:

1. Jurusan Teknik Informatika – S1
 Dalam penerapan algoritma apriori pada data mahasiswa yang sudah lulus tepat waktu menghasilkan 1 itemset yaitu mahasiswa yang berasal dari IPA dan B memiliki nilai rata-rata B pada penelitian ini di peroleh hasil nilai support 34,38% nilai confidance 81,76% dan nilai lift ratio 1,00. Maka bila calon mahasiswa yang memiliki kemiripan data dengan mahasiswa yang sudah lulus tersebut dalam penelitian ini direkomendasikan mengambil Jurusan Teknik Informatika – S1 dengan status kelulusan berhasil.
2. Jurusan Sistem Informasi – S1
 Dalam penerapan algoritma apriori pada data mahasiswa yang sudah lulus tepat waktu menghasilkan 2 itemset. Itemset yang dihasilkan pertama yaitu mahasiswa yang berasal dari jurusan IPS dan memiliki nilai rata – rata B pada penelitian ini diperoleh nilai support 27,07% nilai confidance 80,52% dan nilai lift ratio 1,13. Maka bila calon mahasiswa yang memiliki kemiripan data dengan mahasiswa yang sudah lulus tersebut dalam penelitian ini di rekomendasikan untuk mengambil jurusan Sistem Informasi – S1. Pada itemset pertama kelulusan calon mahasiswa tidak ditemukan karena pada perhitungan kandidat ketiga nilai nya tidak memenuhi syarat. Itemset yang dihasilkan kedua yaitu mahasiswa yang berasal dari jurusan IPA kelak akan lulus dengan status kelulusan Sangat Berhasil pada penelitian ini diperoleh nilai support 32,46% nilai confidance 51,75% dan nilai lift ratio 1,13. Maka bila calon mahasiswa yang memiliki kemiripan data dengan mahasiswa yang sudah lulus tersebut dalam penelitian ini direkomendasikan untuk mengambil jurusan Sistem Informasi – S1. Pada itemset kedua rata – rata nilai calon mahasiswa tidak diketahui karena pada perhitungan kandidat ketiga nilainya tidak memenuhi syarat..
3. Jurusan Desain Komunikasi Visual – S1

Dalam penerapan algoritma apriori pada data mahasiswa yang sudah lulus tepat waktu menghasilkan 2 itemset. Itemset pertama yang dihasilkan yaitu mahasiswa yang berasal dari jurusan IPA dan memiliki nilai rata – rata B pada penelitian ini di peroleh nilai support 37,60% nilai confidence 73,44% dan nilai lift ratio 1,04. Maka bila calon mahasiswa yang memiliki kemiripan data dengan mahasiswa yang sudah lulus tersebut direkomendasikan untuk mengambil jurusan Desain Komunikasi Visual – S1 dengan status kelulusan sangat berhasil. Itemset kedua yang dihasilkan yaitu mahasiswa yang berasal dari jurusan IPS dan memiliki rata nilai B pada penelitian ini diperoleh nilai support 24,80% nilai confidence 83,78% dan nilai lift ratio 1,03. Maka bila calon mahasiswa yang memiliki kemiripan data dengan mahasiswa yang sudah lulus tersebut dalam penelitian dengan status kelulusan sangat berhasil.

4. Jurusan Teknik Informatika – D3

Dalam penerapan algoritma apriori pada data mahasiswa yang sudah lulus tepat waktu menghasilkan 1 itemset yaitu mahasiswa yang berasal dari jurusan IPS dan memiliki nilai rata – rata nilai B pada penelitian ini diperoleh nilai support 20,59% nilai confidence 73,68% dan nilai lift ratio 1,13. Maka bila calon mahasiswa yang memiliki kemiripan data dengan mahasiswa yang sudah lulus tersebut dalam penelitian ini di rekomendasikan untuk mengambil jurusan Teknik Informatika – D3 dengan status kelulusan sangat berhasil.

5. Broadcasting – D3

Dalam penerapan algoritma apriori pada data mahasiswa yang sudah lulus tepat waktu menghasilkan 1 itemset yaitu mahasiswa yang berasal dari jurusan Animasi dan memiliki nilai rata – rata B pada penelitian ini diperoleh nilai support 24,33% nilai confidence 83,33% dan nilai lift ratio 1,02. Maka bila calon mahasiswa yang memiliki kemiripan data dengan mahasiswa yang sudah lulus tersebut dalam penelitian ini direkomendasikan untuk mengambil jurusan Broadcasting – D3 dengan status kelulusan berhasil.

5. SARAN

Saran yang dapat diberikan sehubungan dengan hasil penelitian yang di lakukan adalah sebagai berikut :

1. Menggunakan data lebih banyak untuk menghasilkan jurusan pada Sekolah Menengah lebih banyak sehingga jurusan yang di rekomendasikan lebih bermacam-macam.
2. Menggunakan data lebih banyak untuk menghasilkan itemset/rule yang lebih akurat sehingga dapat dijadikan acuan untuk memilih program studi.

DAFTAR PUSTAKA

- [1] “www.bps.go.id,” [Online].
- [2] R. F. Harahap, “okezone.com,” MNC, 25 February 2014. [Online]. Available: <http://news.okezone.com/read/2014/02/24/373/945961/duh-87-mahasiswa-indonesia-salah-jurusan>. [Diakses 2017].
- [3] “<http://www.banpt.kemdiknas.go.id>,” Kementrian Pendidikan Nasional dan Kebudayaan Republik Indonesia, maret 2014. [Online].
- [4] “PSI UDINUS,” 2008 - 2012. [Online].
- [5] I. T. Sumatera, “ITERA,” 2016. [Online]. Available: <http://usm.itera.ac.id/dampak-salah-memilih-jurusan/>.
- [6] M. Fauzi, K. R. S. W dan I. Asror, “Penerapan Metode Association Rule Menggunakan Algoritma Apriori Pada Simulasi Prediksi Hujan Wilayah Kota Bandung,” *Jurnal Ilmiah Teknologi Informasi Terapan*, vol. II, 2016.