

Algoritma A Priori Dalam Pengelompokan Data Pendaftaran Mahasiswa Baru

Fina Nasari¹

¹ Sistem Informasi Universitas Potensi Utama
Jl. K.L Yos Sudarso KM. 6.5 No. 3A Tanjung Mulia Medana
¹finanasari@gmail.com

Intisari— Pendaftaran mahasiswa baru merupakan proses penerimaan mahasiswa baru. Proses penerimaan mahasiswa baru yang dilakukan setiap tahunnya menghasilkan data yang semakin besar. Data yang besar jika tidak dimanfaatkan dengan baik, hanya akan memenuhi memory penyimpanan. Data mining menjadi salah satu solusi untuk mendapatkan sebuah informasi baru dari pengolahan data-data lama. Ada beberapa fungsi data mining diantaranya asosiasi, klasifikasi, clustering, prediksi serta pengenalan pola. Algoritma Apriori adalah salah satu algoritma dalam data mining yang melihat hubungan antara satu item set dengan item set yang lain (asosiasi antara satu item set dengan item set yang lain). Hubungan yang akan dilihat dalam penelitian ini adalah hubungan antara asal sekolah menengah atas dengan program studi yang akan dipilih. Hasil yang diperoleh dari penelitian ini adalah calon mahasiswa yang memilih program studi sistem informasi adalah calon mahasiswa yang berasal dari SMK (Sekolah menengah Kejuruan) dengan nilai confidence 83 % begitu juga pada program studi teknik informatika, calon mahasiswa berasal dari SMK (Sekolah menengah Kejuruan) dengan nilai confidence 83 %.

Keywords— A Priori, Data Mining

Abstract—Registration of new students is the process of admitting new students. The process of admitting new students which is carried out every year produces more and more data. Large data if not utilized properly, will only fill up storage memory. Data mining is one of the solutions to obtain new information from processing old data. There are several data management functions including association, classification, clustering, prediction and pattern recognition. The Apriori algorithm is one of the algorithms in data mining that looks at the relationship between one item set and another item set (association between one item set and another item set). The relationship that will be seen in this study is the relationship between high school origin and the study program to be chosen. The results obtained from this study are prospective students who choose the information systems study program are prospective students from SMK (Vocational High School) with a confidence value of 83% as well as in the informatics engineering study program, prospective students come from SMK (Vocational High School). with a confidence value of 83%

Keywords— A Priori, Data Mining

I. PENDAHULUAN

Pendaftaran mahasiswa baru merupakan kegiatan tahunan yang dilakukan untuk menerima mahasiswa baru di lingkungan Universitas Potensi Utama. Pada proses pendaftaran mahasiswa baru, setiap calon mahasiswa wajib mengisi formulir pendaftaran mahasiswa baru yang berisi data-data calon mahasiswa. Data-data calon mahasiswa tersebut yang apabila dibiarkan terus menerus akan menjadi data yang beku, namun akan berbeda ketika data tersebut diolah atau diproses untuk menghasilkan data yang baru.

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan [1].

Salah satu algoritma dalam data mining adalah algoritma a priori. Algoritma a priori bertujuan untuk menentukan frequent itemsets dijalankan pada sekumpulan

data. Analisis a priori didefinisikan suatu proses untuk menemukan semua aturan a priori yang memenuhi syarat minimum untuk support dan syarat minimum untuk confidence [2]. Algoritma a priori dalam penelitian ini digunakan untuk melihat hubungan antara asal sekolah menengah atas dengan program studi yang dipilih di Universitas, dengan begitu bisa dilihat kecenderungan dari masing-masing jurusan dalam pemilihan program studi.

Tujuan dari penelitian ini adalah mengelompokkan asal sekolah dari jurusan yang dipilih mahasiswa di Universitas Potensi Utama.

II. METODOLOGI PENELITIAN

1.1 Analisa Penelitian Terdahulu

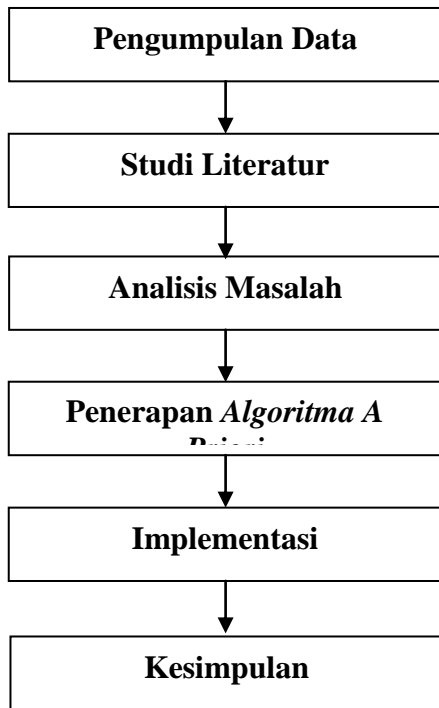
Menurut Robi Yanto & Riri Khoiriah dalam Jurnal yang berjudul Implementasi data mining dengan metode algoritma a priori dalam menentukan pola pembelian obat menerangkan bahwa dengan menerapkan algoritma a priori telah memenuhi kebutuhan dan lebih efektifitas dari sistem sebelumnya dalam penataletakan obat dan memudahkan

dalam mengetahui keberadaan obat yang diketahui dari 2 *itemset* obat[3].

Menurut Kennedy dkk dalam jurnal Implementasi data mining *algoritma a priori* pada sistem persediaan alat-alat kesehatan menjelaskan bahwa data mining dengan *algoritma a priori* sangat efisien dan mempercepat proses pembentukan kecenderungan pola kombinasi *itemset* hasil penjualan alat-alat kesehatan di Apotek Kelambir 2 Medan[4].

1.2 Alur Analisis

Untuk menganalisa data pendaftaran, alur analisa data yang dilakukan dapat dilihat pada gambar 1.



Gambar 1. Alur Analisis

1.2.1 Pengumpulan Data

Metode yang digunakan dalam rangka pengumpulan data-data yang diperlukan dalam penelitian ini adalah sebagai berikut:

1. Studi literatur dari sumber-sumber kepustakaan sebagai landasan dalam menganalisis permasalahan yang disusun dalam penelitian ini.
2. Pengumpulan data yang diperoleh dari data pendaftaran atau penerimaan mahasiswa baru di Universitas Potensi Utama.

2.2.2 Studi Literatur

Pada tahap ini mempelajari sumber-sumber keperustakaan, baik karya ilmiah/jurnal ilmiah maupun dari buku bacaan yang berkaitan dengan data mining atau *algoritma a priori*.

2.2.3 Analisis Masalah

Proses pendaftaran mahasiswa baru yang dilakukan setiap tahun menghasilkan data pendaftaran yang semakin bertambah tiap tahunnya, jika untuk satu atau dua tahun data tidak cukup memberikan dampak yang signifikan dalam mempengaruhi ruang penyimpanan data, namun ketika data dibiarkan maka akan membutuhkan ruang penyimpanan yang cukup besar, namun data tersebut tidak memberikan kontribusi yang berarti atau tidak menghasilkan informasi baru. Dengan adanya data mining memberikan satu solusi dalam penggalian pengetahuan baru dari data-data yang sudah ada, dalam hal ini adalah data pendaftaran mahasiswa baru untuk melihat kecenderungan kebersamaan antar satu itemset dengan itemset yang lain dengan menerapkan sebuah *algoritma a priori* untuk melihat apakah ada kecenderungan asal sekolah menengah atas dengan pemilihan program studi pada saat mendaftar di Universitas.

2.2.4 Penerapan *algoritma A Priori*

Pada tahap ini menjelaskan terkait dengan penerapan *algoritma a priori* pada data pendaftaran mahasiswa baru di universitas potensi utama.

2.2.4.1 Data Mining

Datamining merupakan proses menemukan korelasi baru yang bermanfaat, pola dan *trend* dengan menambang sejumlah repositori data dalam jumlah besar, menggunakan teknologi pengenalan pola seperti statistik dan teknik matematika[5].

Datamining semakin menyebar dan berkembang dengan pesat belakangan ini karena kemampuannya dalam menambang pola bermanfaat dan *trend* dari basis data yang sudah ada. Perusahaan-perusahaan telah menghabiskan dana milyaran untuk mengumpulkan data dalam jumlah *megabytes* atau *terabytes* tapi tidak mendapatkan keuntungan yang bernilai di dalamnya, padahal didalamnya terdapat informasi yang berharga namun tersembunyi pada repositori data[5].

Terdapat enam fungsi dalam data mining, enam fungsi tersebut diantaranya:

1. Fungsi deskripsi (*description*)
2. Fungsi estimasi (*estimation*)
3. Fungsi prediksi (*prediction*)
4. Fungsi klasifikasi (*classification*)
5. Fungsi pengelompokan (*clafification*)
6. Fungsi asosiasi (*association*) [5].

2.2.4.2 *Algoritma A Priori*

Algoritma a priori biasa digunakan dalam menentukan *frequent itemsets* untuk aturan *assosiasi boolean*[4]. *Algoritma a priori* mampu melihat *assosiasi* kombinasi dari dua itemset pada pada penjualan Sparepart Mobil pada PT. IDK 1 dengan nilai *Confident* diatas 70 %[6].

Interestingness measure yang dapat digunakan dalam data mining adalah:

1. *Support* digunakan untuk melihat dominasi suatu item atau itemset dalam sebuah transaksi data.

2. *Confidence* digunakan untuk melihat tingkat keyakinan dari suatu hubungan dari dua itemset atau lebih[3].

Formula yang digunakan untuk menghitung nilai *support* terhadap item A adalah:

$$Support(A) = \frac{Jumlah\ Transaksi\ yang\ mengandung\ item\ A}{Total\ Transaksi} \quad (1)[6]$$

Sementara formula yang digunakan untuk menghitung nilai *Support* dari 2 item dapat adalah sebagai berikut:

$$Support(A, B) = \frac{Jumlah\ Transaksi\ yang\ mengandung\ item\ A\ dan\ B}{Total\ Transaksi} \quad (2) \quad [6]$$

Setelah nilai kombinasi antar item ditemukan berdasarkan hasil perhitungan nilai *support*, selanjutnya menghitung nilai *confident* dengan menggunakan rumus berikut ini:

$$Confidence(A \rightarrow B) = P(B|A) = \frac{Jumlah\ Transaksi\ yang\ mengandung\ item\ A\ dan\ B}{Jumlah\ Transaksi\ yang\ mengandung\ A} \quad (3)$$

2.2.4.3 Penerapan *Algoritma a priori* pada data Pendaftaran Mahasiswa Baru

Pada Tahap ini dilakukan penerapan *algoritma a priori* pada data pendaftaran Mahasiswa baru. Variabel yang akan digunakan adalah variabel program studi dan variabel asal sekolah.

III. HASIL DAN PEMBAHASAN

Hasil Pembahasan

1. Data

Data yang dipakai dalam penelitian ini adalah data mahasiswa baru stambuk 2015 Universitas Potensi Utama. Sample data dapat dilihat pada Tabel 1.

Tabel. 1 Data Pendaftaran Mahasiswa Baru

No. USM	Ujian	Nama Calon Mahasiswa	Program Studi	Asal Sekolah
15330006		Afni Sari Ayu	Desain Interior	SMK
15320002		Rizky Maulana	Desain Komunikasi Visual	SMK
15110038		Sri Rezeki Widya Ritonga	Ekonomi Syariah	SMA
15100006		Febry Alfaruzi Sitorus	Ilmu Hubungan Internasional	SMA
15310017		Muhammad Rian Maulana	Manajemen Informatika	SMK
15930013		Hamdan Pelani	Pend. Bahasa Inggris	SMA
15120013		Novita Pertiwi	Perbankan Syariah	SMA
15830004		Riza Pahlawan	Psikologi	SMA

No. USM	Ujian	Nama Calon Mahasiswa	Program Studi	Asal Sekolah
15210014		Vina Monica	Sistem Informasi	SMA
15230027		Tuti Susilawati.B	Sistem Informasi	SMK
15210022		Muhammad Rizki	Sistem Informasi	SMK
15210081		Rudi Leo Patra Sihotang	Sistem Informasi	SMK
15220011		Khairuna Nisrin	Sistem Informasi	SMK
15230076		Sri Hartini	Sistem Informasi	SMK
15130108		Muhammad Iqbal Khadafi	Teknik Informatika	SMK
15130008		Muhammad Nasri Gea	Teknik Informatika	SMK
15110084		Syukri Adi Yahya Siregar	Teknik Informatika	SMK
15110085		Khairrin Nisa Sitompul	Teknik Informatika	SMK
15130119		Achmad Darmanto	Teknik Informatika	SMK
15130117		Herri Juliyanto	Teknik Informatika	SMK

2. Transformasi Data

Proses transformasi data adalah proses perubahan data kedalam bentuk lain, tujuan dari transformasi data membantu dan mempermudah pengolahan data.

Data yang akan diolah selanjutnya transformasikan untuk memudahkan dalam pengolahan data. Hasil transformasi data dapat dilihat pada Tabel. 2

Tabel. 2 Hasil Transformasi Data Program Studi

Program Studi	Transformasi
Desain Interior	DI
Desain Komunikasi Visual	DKV
Ekonomi Syariah	ES
Ilmu Hubungan Internasional	HI
Manajemen Informatika	MI
Pend. Bahasa Inggris	PBI
Perbankan Syariah	PS
Psikologi	Psi
Sistem Informasi	SI
Teknik Informatika	TIF

3. Tabulasi Frekuensi

Tabulasi frekuensi berisi frekuensi untuk masing-masing itemset. Hasil tabulasi frekuensi dari data penelitian diatas dapat dilihat pada tabel. 3

Tabel. 3 Hasil Tabulasi Frekuensi

No. Ujian USM	DI	DKV	ES	HI	MI	PBI	PS	Psi	SI	TIF	SMA	SMK
15330006	1	0	0	0	0	0	0	0	0	0	0	1
15320002	0	1	0	0	0	0	0	0	0	0	0	1
15110038	0	0	1	0	0	0	0	0	0	0	1	0
15100006	0	0	0	1	0	0	0	0	0	0	1	0
15310017	0	0	0	0	1	0	0	0	0	0	0	1
15930013	0	0	0	0	0	1	0	0	0	0	1	0
15120013	0	0	0	0	0	0	1	0	0	0	1	0
15830004	0	0	0	0	0	0	0	1	0	0	1	0
15210014	0	0	0	0	0	0	0	0	1	0	1	0
15230027	0	0	0	0	0	0	0	0	1	0	0	1
15210022	0	0	0	0	0	0	0	0	1	0	0	1
15210081	0	0	0	0	0	0	0	0	1	0	0	1
15220011	0	0	0	0	0	0	0	0	1	0	0	1
15230076	0	0	0	0	0	0	0	0	1	0	0	1
15130108	0	0	0	0	0	0	0	0	0	1	0	1
15130008	0	0	0	0	0	0	0	0	0	1	0	1
15110084	0	0	0	0	0	0	0	0	0	1	0	1
15110085	0	0	0	0	0	0	0	0	0	1	0	1
15130119	0	0	0	0	0	0	0	0	0	1	0	1
15130117	0	0	0	0	0	0	0	0	0	1	0	1
Σ	1	1	1	1	1	1	1	1	6	6	6	14

4. Menghitung Nilai *Support* dari masing-masing *itemset*

Menghitung nilai *support* dari masing-masing *itemset* dilakukan dengan menggunakan rumus.

$$Support(A) = \frac{Jumlah\ Transaksi\ yang\ mengandung\ item\ A}{Total\ Transaksi}$$

Nilai *Support* dari masing-masing *itemset* dapat dilihat pada tabel.4

Tabel.4 Nilai *Support* masing-masing *itemset*

Nama Item	Jumlah	Support
DI	1	5 %
DKV	1	5 %
ES	1	5 %

Nama Item	Jumlah	Support
HI	1	5 %
MI	1	5 %
PBI	1	5 %
PS	1	5 %
Psi	1	5 %
SI	6	30 %
TIF	6	30 %
SMA	6	30 %
SMK	14	70 %

Nilai *minimum support* 25%, *items* yang memenuhi syarat *minimum support* akan dihitung kembali untuk mencari nilai *support* dua *itemset*. *Item* yang memenuhi syarat dapat dilihat pada tabel. 5

Tabel. 5 *Itemset* yang memenuhi syarat *minimum support*

Nama Item	Jumlah	Support
SI	6	30 %
TIF	6	30 %
SMA	6	30 %
SMK	14	70 %

Selanjutnya dari *itemset* yang memenuhi syarat *minimum support* dihitung kembali untuk mencari *support* dua *itemset*. Untuk menghitung nilai *support* dari dua *itemset* menggunakan rumus:

$$Support(A) = \frac{Jumlah\ Transaksi\ yang\ mengandung\ item\ A}{Total\ Transaksi}$$

Hasil perhitungan nilai *support* dari dua *itemset* dapat dilihat pada tabel.6

Tabel.6 Nilai *Support* dua *Itemset*

Nama Item	Jumlah	Support
SI, TIF	0	0 %
SI, SMA	1	5 %
SI, SMK	5	25 %
TIF, SMA	0	0 %
TIF, SMK	6	30 %

Dari hasil perhitungan dua *itemset* diperoleh dua item yang memenuhi syarat dan selanjutnya dilakukan perhitungan mencari nilai *support* untuk tiga *itemset*. Hasil perhitungan nilai *support* dari tiga *item set* dapat dilihat pada tabel.7

Tabel. 7 Nilai *support* tiga *itemset*

Nama Item	Jumlah	Support
SI, TIF, SMK	0	0

Hasil perhitungan nilai *support* dari tiga *itemset* tidak memenuhi nilai *minimum support*, oleh karena itu proses perhitungan nilai *support* berhenti pada *support* dua *itemset*. Dua *itemset* tersebut selanjutnya akan dicari nilai *confidence* dari hubungan masing-masing *itemset*.

5. Menghitung Nilai *Confidence*

Dalam mencari nilai *confidence* dari masing-masing kombinasi *itemset* akan ditentukan terlebih dahulu nilai *minimum confidence*. Nilai *minimum confidence* 75 %. Untuk mencari nilai *confidence* dilakukan dengan menggunakan rumus berikut ini:

$$Confidence(A \rightarrow B) = P(B|A) = \frac{Jumlah\ Transaksi\ yang\ mengandung\ item\ A\ dan\ B}{Jumlah\ Transaksi\ yang\ mengandung\ A}$$

Hasil perhitungan nilai *confidence* dapat dilihat pada tabel.8

Tabel. 8 Hasil Perhitungan nilai *confidence*

No. Aturan	Aturan	Confidence
A1	Jika Program Studi Yang Dipilih SI atau TIF, Maka Asal sekolah SMK	0
A2	Jika Program Studi Yang Dipilih TIF, Maka Asal sekolah SMK	100%
A3	Jika Program Studi Yang Dipilih SI, Maka Asal sekolah SMK	83%

Dari tiga aturan yang memenuhi syarat *minimum confidence* hanya aturan A2 dengan nilai *confidence* 100% dan aturan A3 dengan nilai *confidence* 83 %, Dengan demikian hubungan antara kombinasi dua *itemset* memilihi hubungan yang mempengaruhi namun tidak sebaliknya.

3.2 Hasil

Implementasi dilakukan dengan melakukan perhitungan *algorithm a priori* menggunakan aplikasi. Hasil perhitungan *algorithm a priori* dengan sistem yang ada dapat dilihat pada gambar. 2



Gambar 2. Hasil Perhitungan *algorithm a priori* dengan sistem

Dari hasil perhitungan algoritma apriori diperoleh informasi sebagai berikut:

- a. Jika Program Studi yang dipilih adalah SI dan TIF, Maka Asal Sekolah SMK tidak memiliki nilai *confidence* atau tidak dapat dipastikan ketika mahasiswa memilih program studi SI dan TIF maka belum tentu mahasiswa berasal dari SMK
- b. Jika Program Studi yang dipilih adalah TIF, Maka Asal sekola adalah SMK memiliki nilai *confidence* 93 % dan memenuhi syarat minimum *confidence*, sehingga bisa diambil kesimpulan bahwa mayoritas mahasiswa yang memilih program studi TIF berasal dari SMK
- c. Jika Program Studi yang dipilih adalah SI, Maka Asal Sekolah adalah SMK memiliki nilai *confidence* 83 % dan memenuhi syarat minimum *confidence*, sehingga bisa diambil kesimpulan bahwa mayoritas mahasiswa yang memilih program studi SI berasal dari SMK
- d. Jika Calon Mahasiswa berasal dari SMK, Maka memilih program studi SI memiliki nilai *confidence* 36 % belum memenuhi syarat minimum *confidece* sehingga bisa disimpulkan belum bisa dipastikan jika asal sekolah SMK pasti memilih program studi SI
- e. Jika Calon Mahasiswa berasal dari SMK, Maka memilih program studi TIF memiliki nilai *confidence* 46 % belum memenuhi syarat minimum *confidence* sehingga dapat disimpulkan belum bisa dipastikan jika asal sekolah SMK pasti memilih program studi TIF

IV. KESIMPULAN

Kesimpulan yang dapat diambil dari penelitian ini adalah:

- a. Mayoritas yang memilih program studi Sistem Informasi(SI) adalah mahasiswa yang berasal dari Sekolah Menengah Kejuruan(SMK) dengan memiliki nilai *confidence* 83 % dan mayoritas yang memilih program studi Teknik Informatika(TIF) adalah mahasiswa yang berasal dari Sekolah Menengah Kejuruan(SMK) dengan memiliki nilai *confidence* 93 %, Namun tidak berlaku sebaliknya dimana mahasiswa yang berasal dari Sekolah Menengah Kejuruan (SMK) belum tentu memilih program studi Sistem Informasi(SI) maupun Teknik Informatika (TIF).
- b. Pada *algoritma a priori* nilai *confidence* dari hubungan antar *itemset* belum tentu berlaku sebaliknya. Misalnya Jika $A \rightarrow B$ terpehuni belum tentu $B \rightarrow A$ terpenuhi.
- c. *Algoritma a priori* dipengaruhi oleh kombinasi *itemset* untuk setiap transaksi dan intensitas kebersamaan antar satu *itemset* dengan *itemset* yang lain. Semakin besar intensitas kebersamaan, maka semakin besar juga kemungkinan nilai *confidence* dari kombinasi *itemset* tersebut.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Unviersita Potensi Utama yang telah menyediakan tempat riset dan mendukung terlaksananya penelitian ini.

DAFTAR PUSTAKA

- [1] Rajesh, K., & Anand, S. (2012). Analysis of SEER dataset for breast cancer diagnosis using C4. 5 classification algorithm. *International Journal of Advanced Research in Computer and Communication Engineering*, 1(2), 2278-1021.
- [2] Sari, E. N. (2013). Analisa Algoritma Apriori Untuk Menentukan Merek Pakaian Yang Paling Diminati Pada Mode Fashion Group Medan. *Pelita Informatika: Informasi dan Informatika*, 4(3).
- [3] Yanto, R., & Khoiriah, R. (2015). Implementasi Data Mining dengan Metode Algoritma Apriori dalam Menentukan Pola Pembelian Obat. *Creative Information Technology Journal*, 2(2), 102-113.
- [4] Tampubolon, K., Saragih, H., & Reza, B. (2013). Implementasi Data Mining Algoritma Apriori pada sistem persediaan alat-alat kesehatan. *Majalah Ilmiah Informasi dan Teknologi Ilmiah (INTI)*, 1(1).
- [5] Larose, Daniel, 2005, *Discovery Knowledge in Data, A Jhon Wiley & Sons, Inc Publication*. Canada
- [6] Umami, K. (2016). Analisa Data Mining Dalam Penjualan Sparepart Mobil Dengan Menggunakan Metode Algoritma Apriori (Studi Kasus: DI PT. IDK 1 Medan). *CSRID (Computer Science Research and Its Development Journal)*, 8(3), 155-164