

PENERAPAN ASSOCIATION RULE DENGAN ALGORITMA APRIORI UNTUK MENAMPILKAN INFORMASI TINGKAT KELULUSAN MAHASISWA TEKNIK INFORMATIKA S1 FAKULTAS ILMU KOMPUTER UNIVERSITAS DIAN NUSWANTORO

Riko Adhi Saputro

Abstract— Dengan adanya pertumbuhan yang pesat dari akumulasi data telah menciptakan kondisi yang kaya akan data tapi minim akan informasi. Data mining adalah penemuan informasi baru dengan mencari pola atau aturan dari sejumlah data dalam jumlah besar untuk mendapatkan hubungan yang dapat memberikan indikasi tertentu yang berguna. Dengan memanfaatkan data kelulusan dan data induk dari mahasiswa UDINUS diharapkan dapat menghasilkan informasi tentang tingkat kelulusan mahasiswa melalui teknik data mining. Kategori tingkat kelulusan diukur dari lama studi dan IPK. Algoritma yang digunakan adalah algoritma Apriori, informasi yang ditampilkan berupa nilai *support* dan *confidence* dari masing-masing kategori tingkat kelulusan.

Kata kunci: data mining, algoritma Apriori, tingkat kelulusan, data induk mahasiswa.

INTRODUCTION

anyaknya jumlah mahasiswa yang masuk tentu menjadi kebanggaan dari suatu Universitas, namun dengan jumlah mahasiswa yang banyak dimana jumlah mahasiswa masuk tidak sama dengan mahasiswa yang lulus. Akan mengakibatkan banyak masalah antara lain turunnya akreditasi pada perguruan tinggi tersebut, kualitas dari lulusan mahasiswa, serta menurunnya kinerja program studi. Dengan mengetahui faktor faktor tingkat kelulusan maka hal tersebut dapat diminimalkan. Oleh karena itu, pada penelitian ini penulis akan menerapkan Algoritma Apriori untuk menampilkan informasi tingkat kelulusan mahasiswa Fakultas Ilmu Komputer dengan aplikasi SPMF untuk menentukan nilai *support* dan nilai *confiden* dari data mahasiswa yang telah diolah.

I. TINJAUAN PUSTAKA

Berbagai penelitian yang relevan telah dilakukan oleh peneliti terdahulu sebagai berikut:

Nuqson Masykur Huda [1] dengan judul “*Aplikasi Data Mining Untuk Menampilkan Informasi Tingkat Kelulusan Mahasiswa (Studi Kasus di Fakultas MIPA Universitas Diponegoro)*”. Dalam penelitiannya membahas mengenai penggunaan algoritma *A-Priori* untuk menentukan tingkat kelulusan mahasiswa dengan memanfaatkan data induk dan data kelulusan mahasiswa yang ada. Pada penelitian yang dilakukan, digunakan data dari buku wisuda ke-115 yang

dimana 48 dari 80 peserta wisuda Program Sarjana (S1) regular di Fakultas MIPA menempuh masa studi lebih dari 8 semester. Tingkat kelulusan mahasiswa dapat dilihat dari lama studi dan IPK (Indeks Prestasi Kumulatif). Dari dua parameter tersebut data diubah menjadi tipe data yang memudahkan untuk diproses. Tingkat kelulusan diukur dari lama studi dan IPK, lama studi dikategorikan berdasarkan peraturan akademik BAB I pasal 1 ayat 2 yang berbunyi “Program sarjana (S1) regular adalah program pendidikan akademik setelah pendidikan menengah, yang memiliki beban studi sekarang-kurangnya 144 (seratus empat puluh empat) sks dan sebanyak-banyaknya 160 (seratus enam puluh) sks yang dijadwalkan untuk 8 (delapan) semester dan dapat ditempuh dalam waktu kurang dari 8 (delapan) semester dan paling lama 14 (empat belas) semester.” Batas minimum *support* yang diberikan sebesar 1 hingga 3. Setelah proses *mining* akan disajikan hasil dari *data mining* berupa tabel hubungan kekuatan dengan nilai *support* dan *confidence* masing-masing atribut serta *threshold* yang digunakan. Semakin tinggi nilai *confidence* dan *support* maka semakin kuat nilai hubungan antar atribut.

Benni R Siburian [2] dengan judul “*Aplikasi Data Mining Untuk Menampilkan Tingkat Kelulusan Mahasiswa Dengan Algoritma Apriori*”. Penelitian ini mengacu pada bagaimana menentukan faktor-faktor tingkat kelulusan yang efektif dan merancang aplikasi data mining untuk menampilkan informasi kelulusan mahasiswa. Table-table beserta isinya adalah contoh data kelulusan mahasiswa dengan menggunakan data fiktif. Hal ini dimaksudkan untuk memberikan gambaran bagaimana melakukan mining sehingga menghasilkan aturan asosiasi yang sesuai.

Friski Tri Cahyono [3] judul “*Penerapan Algoritma Naïve Bayes Untuk Memprediksi Masa Studi Mahasiswa Teknik Informatika S1 Fakultas Ilmu Komputer Universitas Dian Nuswantoro*”. Dalam penelitiannya membahas tentang penggunaan algoritma *Naïve Bayes* untuk memprediksi masa studi mahasiswa dengan fitur seleksi menggunakan Algoritma Genetika untuk menentukan variabel yang akan digunakan. Kelebihan Algoritma *Naive Bayes Classifier* diataranya adalah tingkat akurasi yang tinggi dan error rate yang minimum. Algoritma Genetika digunakan untuk mencari *optimasi* terbaik dari variabel – variabel yang digunakan. *Optimasi* adalah proses menyelesaikan suatu masalah tertentu supaya berbeda pada kondisi yang paling menguntungkan dari suatu sudut pandang. Prinsip dasar dari algoritma genetika sendiri adalah bagaimana mendapatkan keturunan yang lebih baik dengan melalui proses seleksi layaknya seleksi alam. Pada

penelitiannya akan digunakan data mahasiswa yang telah lulus sebagai *data training* dan *data testing*. Dengan penerapan algoritma ini diharapkan dapat membantu untuk memprediksi masa studi mahasiswa yang lulus tepat waktu atau lulus tidak tepat waktu.

II. DATA MINING

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar [1]. Menurut Gartner Group data mining didefinisikan sebagai suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika. Selain definisi di atas beberapa definisi juga diberikan seperti tertera dibawah ini : “Data mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual.” “Data mining merupakan [4] bidang dari beberapa bidang keilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, database, dan visualisasi untuk pengenalan permasalahan pengambilan informasi dari database yang besar.” Kemajuan luar biasa yg terus berlanjut dalam bidang data mining didorong oleh beberapa faktor, antara lain [4] :

1. Pertumbuhan yang cepat dalam pengumpulan data.
2. Penyimpangan data dalam data warehouse, sehingga seluruh perusahaan memiliki akses kedalam database yang handal.
3. Adanya peningkatan akses data melalui navigasi web dan intranet.
4. Tekanan kompetisi bisnis untuk meningkatkan penguasaan pasar dalam globalisasi ekonomi.
5. Perkembangan teknologi perangkat lunak untuk data mining (ketersediaan teknologi).
6. Perkembangan yang hebat dalam kemampuan komputasi dan pengembangan kapasitas media penyimpanan.

Dari definisi-definisi yang telah disampaikan, hal penting yang terkait dengan data mining adalah:

1. Data mining merupakan suatu proses otomatis terhadap data yang sudah ada.
2. Data yang akan diproses berupa data yang sangat besar.
3. Tujuan data mining adalah mendapatkan hubungan atau pola yang mungkin memberikan indikasi yang bermanfaat.

Untuk melakukan penggalian data, ada beberapa tahapan. Tahap-tahap tersebut bersifat interaktif di mana pemakai terlibat langsung atau dengan perantaraan knowledge base [5].

a. Pembersihan data

Pada umumnya data yang diperoleh dari perusahaan memiliki data yang tidak sempurna seperti data yang hilang, data yang tidak valid. Sebaiknya data-data yang tersebut lebih baik dibuang karena keberadaannya dapat mengurangi mutu atau akurasi dari hasil data mining nantinya.

b. Integrasi data

Tidak jarang data yang diperlukan untuk data mining tidak hanya berasal dari satu database tetapi juga berasal dari beberapa database atau file teks. Integrasi data perlu dilakukan secara hati-hati dikarenakan kesalahan pada integrasi data dapat terjadi penyimpangan pada data keluaran proses data mining.

c. Transformasi data

Beberapa teknik data mining membutuhkan format data yang khusus sebelum bisa diaplikasikan. Disini juga dilakukan pemilihan data yang diperlukan oleh teknik data mining yang dipakai. Transformasi dan pemilihan data ini juga menentukan kualitas dari hasil data mining nantinya karena ada beberapa karakteristik dari teknik-teknik data mining tertentu yang tergantung pada tahapan ini.

d. Aplikasi teknik data mining

Aplikasi teknik data mining sendiri hanya merupakan salah satu bagian dari proses data mining. Gunakan teknik data mining yang sesuai dengan hasil yang diinginkan.

e. Evaluasi pola yang ditemukan

Dalam tahap ini hasil dari teknik data mining berupa polapola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai hipotesa ada beberapa alternatif yang dapat diambil seperti: menjadikannya umpan balik untuk memperbaiki proses data mining, mencoba teknik data mining lain yang lebih sesuai, atau menerima hasil ini sebagai suatu hasil yang di luar dugaan yang mungkin bermanfaat.

f. Presentasi pola

Tahap terakhir dari proses data mining adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisa yang didapat. Dalam presentasi ini, visualisasi juga bisa membantu mengkomunikasikan hasil data mining.

Berikut ini adalah enam tahap siklus hidup pengembangan data mining CRISP-DM [6] :

1. Fase Pemahaman Bisnis
2. Fase Pemahaman Data
3. Fase pengolahan
4. Fase Pemodelan
5. Fase Evaluasi
6. Fase Penyebaran

III. ALGORITMA A-PRIORI

Algoritma a-priori adalah suatu algoritma dasar yang diusulkan oleh Agrawal dan Srikant pada tahun 1994 untuk menentukan Frequent itemsets untuk aturan asosiasi Boolean. Algoritma A-priori termasuk jenis Aturan Asosiasi pada data mining [7]. Aturan yang menyatakan asosiasi antara beberapa atribut sering disebut affinity analysis atau market basket analysis. Analisis asosiasi atau association rule mining adalah teknik data mining untuk menemukan aturan suatu kombinasi item. Salah satu tahap analisis asosiasi yang menarik perhatian banyak peneliti untuk menghasilkan algoritma yang efisien adalah analisis pola frekuensi tinggi (frequent pattern mining). Penting tidaknya suatu asosiasi dapat diketahui dengan dua tolok ukur, yaitu : support dan confidence. Support (nilai penunjang) adalah persentase kombinasi item tersebut dalam

database, sedangkan confidence (nilai kepastian) adalah kuatnya hubungan antara-item dalam aturan asosiasi.

1. Pola frekuensi tinggi

Adalah tahap dimana pembentukan kombinasi antar item yang telah memenuhi syarat *minimum support* yang telah ditentukan sebelumnya. Nilai *support* pada satu itemset dapat diperoleh dengan menggunakan sebuah rumus di bawah ini:

$$\text{Support}(A) = \frac{\sum \text{Transaksi yang mengandung } A}{\sum \text{transaksi yang telah terjadi}} \quad (1)$$

Rumus diatas memiliki arti untuk menentukan nilai *support* pada satu itemset jumlah transaksi yang mengandung item A dibagi dengan jumlah transaksi yang ada pada database.

Sedangkan pada dua itemset atau lebih diperoleh dengan rumus:

$$\text{Support}(A, B) = \frac{\sum \text{Transaksi mengandung } A \text{ dan } B}{\sum \text{Total transaksi yang terjadi}} \quad (2)$$

Pada rumus diatas untuk menentukan nilai *support* pada dua itemset atau lebih, jumlah transaksi yang mengandung item A dan B dibagi dengan total transaksi yang terjadi pada dataset.

2. Pembentukan Aturan Asosiasi

Langkah yang dilakukan setelah menentukan nilai *support* pada itemset berfrekuensi tinggi lalu dibentuk aturan asosiasi yang menyatakan kuatnya hubungan kombinasi itemset pada transaksi. Untuk menentukan aturan asosiasi yang terbentuk minimal itemset harus memiliki dua kandidat A dan B. Pada rules yang terbentuk berlaku hukum asosiatif $A \rightarrow B$ tidak berlaku $B \rightarrow A$. Untuk menentukan aturan $A \rightarrow B$ digunakan rumus:

$$\text{Confidence}(A \rightarrow B) = \frac{\sum \text{Transaksi mengandung } W \text{ dan } b}{\sum \text{Transaksi mengandung } a}$$

Pada rumus diatas untuk menentukan nilai confidence pada aturan asosiasi, jumlah transaksi yang mengandung item A dan B dibagi dengan jumlah transaksi yang mengandung item A.

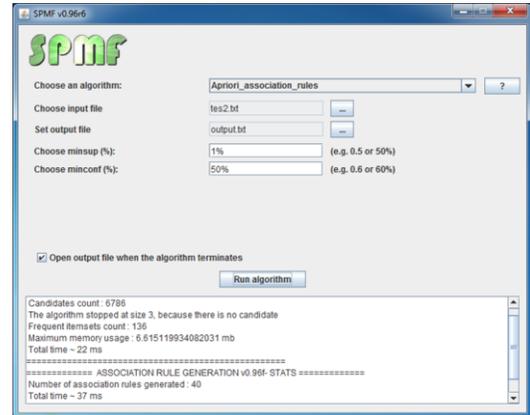
IV. SEQUENTIAL PATTERN MINING FRAMEWORK (SPMF)

Sequential Pattern Mining Framework (SPMF) adalah fungsi pustaka (library) yang ditulis pada bahasa pemrograman Java untuk menangani tugas data mining dengan lisensi open-source GPL v3 [8]. SPMF memiliki 52 koleksi algoritma yang dapat digunakan untuk:

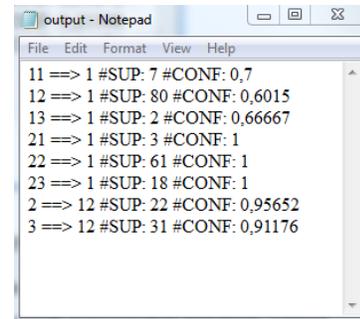
- *Sequential Pattern Mining.*
- *Association Rule Mining.*
- *Frequent Itemset Mining.*
- *Sequential Rule Mining*
- *Clustering*

Source code untuk tiap algoritma yang tersedia pada fungsi pustaka dapat diintegrasikan pada program Java yang dibuat oleh pengembang software [8].

V. PENGUJIAN SPMF



Gambar 1 Pengujian SPMF



Gambar 2 Hasil Training

VI. KESIMPULAN

Dari hasil penerapan *Association rule* data mining induk mahasiswa dan kelulusan mahasiswa TI S1 UDINUS, maka dapat disimpulkan bahwa:

1. *Association rule mining* menggunakan metode *apriori* berhasil diimplementasikan menemukan 8 rule menggunakan software SPMF dengan menentukan nilai minimum *support* 0.2% dan minimum *confidence* 50% dari keterkaitan antara data induk mahasiswa dan data kelulusan mahasiswa.
2. Pola data yang ditemukan pada data induk mahasiswa dan data kelulusan mahasiswa yang memuat atribut jalur masuk dengan kategori *Reguler* memiliki kecenderungan asosiasi yang kuat dengan memuat 6 rule dan atribut predikat prestasi dengan kategori *lama studi 4 tahun atau kurang dari 4 tahun dan IPK 2,76 – 3,50* memuat 3 rule.
3. Kecenderungan pola yang terbentuk dari *association rule mining* data induk dan kelulusan mahasiswa dapat disimpulkan bahwa rata-rata mahasiswa dari jalur masuk *Reguler*, *PMDK*, dan *Khusus* memiliki tingkat kelulusan dengan *lama studi 4 tahun atau kurang dari 4 tahun dan IPK 2,76 – 3,50*.

VII. SARAN

Untuk pengembangan penelitian ini lebih lanjut, diharapkan dapat menggunakan data lebih banyak dan lingkup yang lebih luas. Misal semua jurusan atau bahkan semua fakultas. Dan diharapkan adanya pembuatan aplikasi untuk menampilkan informasi yang lebih lengkap dan dapat digunakan secara maksimal bagi pihak Universitas.

Beberapa ide gagasan yang dapat digunakan untuk penelitian selanjutnya adalah sebagai berikut:

1. Penelitian selanjutnya dapat dikembangkan pada jenis data yang sama dengan menggunakan metode yang lebih baik seperti algoritma FP-GROWTH.
2. Hasil aturan asosiasi yang terbentuk dapat dikembangkan menjadi basis pengetahuan untuk membuat aplikasi yang lebih bagus mengenai hubungan antara data induk dan data kelulusan mahasiswa.

Tabel 1 hasil Pengujian

Kode Predikat Prestasi dengan Jalur Masuk Aturan Asosiasi	Nilai Support	Nilai Confidence
11→1 Mahasiswa dengan Jalur Masuk 1 (Reguler) memiliki tingkat kelulusan 11 (lama studi 4 tahun atau kurang dari 4 tahun dan IPK 3,51 – 4,00)	7/228 (3,07%)	0,70 (70%)
12→1 Mahasiswa dengan Jalur Masuk 1 (Reguler) memiliki tingkat kelulusan 12 (lama studi 4 tahun atau kurang dari 4 tahun dan IPK 2,76 – 3,50)	80/228 (35,08%)	0,6015 (60%)
13→1 Mahasiswa dengan Jalur Masuk 1 (Reguler) memiliki tingkat kelulusan 13 (lama studi 4 tahun atau kurang dari 4 tahun dan IPK 2,00 – 2,75)	2/228 (0,8771%)	0,6666 (66%)
21→1 Mahasiswa dengan Jalur Masuk 1 (Reguler) memiliki tingkat kelulusan 21 (lama studi lebih dari 4 tahun dan IPK 3,51 – 4,00)	3/228 (1,31%)	1 (100%)
22→1 Mahasiswa dengan Jalur Masuk 1 (Reguler) memiliki tingkat kelulusan 22 (lama studi lebih dari 4 tahun dan IPK 2,76 – 3,50)	61/228 (26,75%)	1 (100%)
23→1 Mahasiswa dengan Jalur Masuk 1 (Reguler) memiliki tingkat kelulusan 23 (lama studi lebih dari 4 tahun dan IPK 2,00 – 2,75)	18/228 (7,89%)	1 (100%)
2→12 Mahasiswa dengan Jalur Masuk 2 (Khusus) memiliki tingkat kelulusan 12 (lama studi 4 tahun atau kurang dari 4 tahun dan IPK 2,76 – 3,50)	22/228 (9,64%)	0,95652 (57,143%)
3→12 Mahasiswa dengan Jalur Masuk 3 (PMDK) memiliki tingkat kelulusan 12 (lama studi 4 tahun atau kurang dari 4 tahun dan IPK 2,76 – 3,50)	31/228 (13,59%)	0,91176 (91,176%)

REFERENCES

- [1] 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. 90, 2010.
- [2] B. R. Siburian, "Aplikasi Data Mining Untuk Menampilkan Tingkat Kelulusan Mahasiswa Dengan Algoritma Apriori," *Pelita Informatika Budi Darma*, vol. VII, pp. 1-6, 2014.
- [3] F. T. Cahyono, "Penerapan Algoritma Naive Bayes Untuk Memprediksi Masa Studi Mahasiswa Teknik Informatika S1 Fakultas Ilmu Komputer Universitas Dian Nuswantoro," Skripsi Teknik Informatika Universitas Dian Nuswantoro, Semarang, 2014.
- [4] D. T. Larose, *Discovering Knowledge in Data: An Introduction to Data Mining*, Hoboken: John Wiley & Sons Inc., 2005.
- [5] M., Kamber and Han, *Data Mining concepts and Techniques second Edition*, San Francisco: Morgan Kauffman, 2006.
- [6] P. Chapman, *CRISP-DM 1.0: Step-by-step Data Mining Guide*, SPSS, 2006.
- [7] Kusrini and E.T. Luthfi, *Algoritma Data Mining*, Yogyakarta: Andi Offset, 2009.
- [8] Philippe-Fournier-Viger, "SPMF : An Open-Source Data Mining Library," Philippe-Fournier-Viger, Februari 2010. [Online]. Available: [Http://www.philippe-fournier-viger.com/spmf/index.php](http://www.philippe-fournier-viger.com/spmf/index.php). [Diakses Januari 2015].