

ANALISIS POLA HUBUNGAN DALAM DATA INDEKS PRESTASI KOMULATIF DAN DATA DASAR SNMPTN DENGAN ALGORITMA APRIORI

ANALYSIS OF RELATION TERM BETWEEN CUMMULATIVE PERFORMANCE INDEX AND SNMPTN DATA BASE USING APPRIORITY ALGORITHM

Kuswari Hernawati, Nur Insani, Bambang Sumarno, Nurhadi Waryanto

Jurusan Pendidikan Matematika, FMIPA, Universitas Negeri Yogyakarta

*email: kuswari@uny.ac.id

diterima 18 September 2015, disetujui 30 September 2015

Abstrak

Dalam pengelolaan Perguruan tinggi, selain sumber daya sarana, prasarana dan manusia, sistem informasi adalah salah satu sumber daya yang dapat dimanfaatkan untuk meningkatkan keunggulan bersaing dan menyediakan data yang akurat untuk kepentingan para pengambil kebijakan, misalnya informasi tentang nilai ujian SNMPTN, daerah asal mahasiswa, IPK mahasiswa, lama studi mahasiswa. Universitas Negeri Yogyakarta menerima mahasiswa baru dengan rata-rata kurang lebih 6.000 orang setiap tahunnya, baik melalui jalur Seleksi Nasional Mahasiswa Perguruan Tinggi Negeri (SNMPTN), SBMPTN (Seleksi Bersama Mahasiswa Perguruan Tinggi Negeri) maupun ujian Seleksi Mandiri (SM). Dengan semakin meningkatnya jumlah calon mahasiswa melalui SBMPTN, maka semakin bertambah pula data dasar calon mahasiswa di database setiap tahunnya. Dengan memanfaatkan data dasar mahasiswa pada SBMPTN dan Indeks Prestasi Kumulatif (IPK), penelitian ini bertujuan untuk menerapkan data mining dengan teknik association rule dengan menggunakan algoritma Apriori untuk mencari pola hubungan antara data dasar SBMPTN dan nilai IPK mahasiswa UNY. Data dasar SBMPTN mahasiswa yang akan diproses mining meliputi data asal sekolah, data kabupaten sekolah asal, data penghasilan orang tua, data tingkat pendidikan orang tua, data nilai rata-rata UAN, dan data nilai tes potensi akademik(TPA). Hasil yang diperoleh adalah tidak ada data dalam data dasar SNMPTN yang secara signifikan mempengaruhi perolehan IPK. Hal ini terlihat dari aturan asosiasi yang diperoleh dari 50 aturan asosiasi terbaik tidak terlihat munculnya itemset IPK yang disertai dengan munculnya itemset yang lainnya.

Kata kunci: *Data Mining, Association Rule, Algoritma Apriori, SNMPTN*

Abstract

In the university management, in addition to resources infrastructure, facilities and people, information systems is one of the resources that can be utilized to enhance the competitive advantage and provide accurate data for the benefit of policy makers, for example, information about the test scores SNMPTN, region of origin students, GPA student, students study duration. Yogyakarta State University to accept new students with an average of approximately 6,000 people annually, through the National Selection of State Universities Student (SNMPTN), SBMPTN (Joint Student Selection State University) and the Independent Selection exam (SM). With the increasing number of prospective students through SBMPTN, then increasingly also the basic data in a database of prospective students annually. By utilizing basic data on SBMPTN students and grade point average (GPA), the study aims to apply data mining techniques using the association rule Apriori algorithm to look for patterns of association between baseline SBMPTN and UNY students GPA. Basic data to be processed SBMPTN mining student data origin include school, home school district data, earnings data parent, parental education level data, the average value data UAN, and data values academic potential test (TPA). The results obtained are no data in the data base SNMPTN that significantly affect the acquisition of GPA. This is evident from the association rules derived from the 50 best asosiasi rules not seen the emergence of itemset GPA accompanied by the emergence of other itemset.

Keywords: Data Mining, Association Rule, Algoritma Apriori, SNMPTN

Pendahuluan

Perguruan tinggi saat ini dituntut untuk memiliki keunggulan bersaing dengan memanfaatkan sumber daya yang dimiliki. Selain sumber daya sarana, prasarana dan manusia,

informasi merupakan kebutuhan penting yang harus dipenuhi oleh setiap organisasi termasuk perguruan tinggi. Informasi adalah hasil dari pengolahan data, sedangkan data merupakan

kenyataan yang menggambarkan suatu kejadian nyata. Dalam suatu organisasi sebuah basis data digunakan untuk menyimpan data penting yang ada dan merupakan bahan baku informasi yang selanjutnya diolah dan disajikan sebagai dasar pengambilan keputusan kepada pihak-pihak yang membutuhkan. Peningkatan kualitas keputusan manajerial merupakan salah satu upaya untuk meningkatkan kualitas mutu organisasi/perguruan tinggi.

Salah satu bentuk basis data yang digunakan oleh organisasi yang memiliki data berskala besar adalah data warehouse. Informasi penting dan potensial di dalam data warehouse dapat dianalisis menggunakan teknik yang dikenal dengan istilah data mining. Data mining dapat membantu organisasi dalam mencari dan memahami pola tersembunyi (*hidden pattern*) dari data yang dimiliki. Informasi yang dihasilkan dari penerapan teknik data mining digunakan untuk menggali dan memprediksi potensi-potensi yang ada dalam suatu organisasi. Data mining merupakan sebuah proses untuk menemukan hubungan, pola dan trend baru yang bermakna dengan menyaring data yang sangat besar, yang tersimpan dalam penyimpanan, menggunakan teknik pengenalan pola seperti teknik statistik dan matematika [1]. *Association rule* (aturan asosiatif) adalah salah satu teknik utama dalam data mining dan merupakan bentuk yang paling umum dipakai dalam menemukan pattern atau pola dari suatu kumpulan data [2]. *Association rule* (aturan asosiatif) berusaha menemukan aturan-aturan tertentu yang mengasosiasikan data yang satu dengan data yang lain. Salah satu algoritma yang dapat digunakan untuk menemukan *association rule* adalah algoritma Apriori [3].

Hasil pengetahuan atau informasi yang ditemukan dari data mining dapat digunakan untuk memprediksi berbagai macam hal seperti pendaftaran mahasiswa pada suatu mata kuliah tertentu, keterasingan model pengajaran kelas tradisional, pendeteksian penggunaan alat yang tidak diperbolehkan ketika mengikuti ujian online, pendeteksian nilai abnormal pada lembar jawab mahasiswa, prediksi kinerja / performa mahasiswa, dan sebagainya.

Universitas Negeri Yogyakarta menerima mahasiswa baru dengan rata-rata kurang lebih 6.000 orang setiap tahunnya, baik melalui jalur Seleksi Nasional Mahasiswa Perguruan Tinggi Negeri (SNMPTN) (sebelumnya disebut PBUD), Seleksi Bersama Mahasiswa Perguruan Tinggi Negeri (SBMPTN) (sebelumnya disebut

SNMPTN), maupun ujian Seleksi Mandiri (SM). Pada tahun 2010, enam puluh persen (60%) dari 6000 orang tersebut akan diambil melalui seleksi jalur SNMPTN baik ujian tertulis (utul) maupun melalui jalur undangan, sedangkan 40% lainnya diambil dari ujian SM. Di lain pihak, jumlah calon mahasiswa baru yang mendaftar di tiga universitas negeri di DIY melalui jalur Ujian Tulis pada SNMPTN 2012 meningkat 15-20% dibanding dengan pendaftar pada tahun 2011. Dengan semakin meningkatnya jumlah calon mahasiswa melalui SNMPTN, maka semakin bertambah pula data dasar calon mahasiswa di database setiap tahunnya. Jumlah data yang begitu besar justru dapat membuat suatu permasalahan bagi lembaga pendidikan jika tidak bisa dimanfaatkan. Semakin banyak data, maka lembaga tinggi tersebut semakin memerlukan usaha untuk memilah data mana yang dapat diolah menjadi informasi. Jika data tersebut dibiarkan begitu saja, maka data tersebut hanya akan menjadi gudang atau sampah data yang tidak berarti.

Oleh karena itu, dengan memanfaatkan data dasar mahasiswa pada SBMPTN dan data kelulusan mahasiswa yaitu Indeks Prestasi Kumulatif (IPK) penelitian ini bertujuan untuk mengetahui informasi melalui data mining yang menggunakan teknik aturan asosiatif (*association rule*) dengan algoritma Apriori. Data dasar SBMPTN mahasiswa yang akan diproses mining meliputi data asal sekolah, data kabupaten sekolah asal, data penghasilan orang tua, data tingkat pendidikan orang tua, data nilai rata-rata UAN, data nilai tes potensi akademik (TPA) dan data program studi.

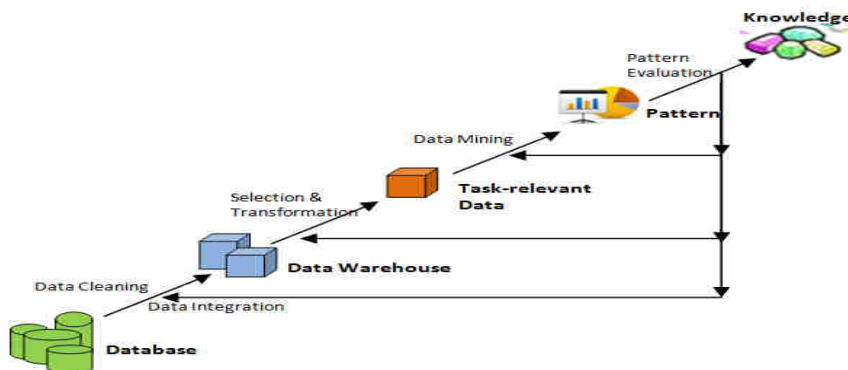
Data Mining

Data mining merupakan sebuah proses untuk menemukan hubungan, pola dan trend baru yang bermakna dengan menyaring data yang sangat besar, yang tersimpan dalam penyimpanan, menggunakan teknik pengenalan pola seperti teknik statistik dan matematika. Hubungan yang dicari dalam data mining dapat berupa hubungan antara dua atau lebih dalam satu dimensi, misalnya dalam dimensi produk, dapat dilihat keterkaitan pembelian suatu produk dengan produk yang lain. Selain itu hubungan juga dapat dilihat antara 2 atau lebih atribut dan 2 atau lebih obyek [4]. Dengan menggunakan teknik data mining, data historik dan data operasional dalam sistem informasi suatu lembaga pendidikan dapat dimanfaatkan untuk

menunjang proses pengambilan keputusan manajemen di masa yang akan datang.

Istilah data mining merujuk pada ekstraksi atau penambangan pengetahuan dari sejumlah besar data [5]. Data Mining, juga dikenal sebagai Knowledge Discovery in Database (KDD),

mengacu pada ekstraksi implisit informasi dari data dalam database yang sebelumnya tidak diketahui dan potensial dalam penemuan pengetahuan. Langkah penemuan pengetahuan dalam data mining disajikan dalam Gambar 1 [7].



Gambar 1 Proses penemuan Pengetahuan dalam Data mining

Proses KDD terdiri dari beberapa langkah yang mengarah dari koleksi data mentah ke beberapa bentuk pengetahuan baru. Proses berulang terdiri dari langkah-langkah berikut:

1. Data cleaning / pembersihan data adalah fase di mana noise data dan data yang tidak relevan akan dihapus dari koleksi.
2. Data Integration/Integrasi data, pada tahap ini, berbagai sumber data sering heterogen, akan dikombinasikan dalam sumber yang sama.
3. Data selection /Pemilihan Data, pada langkah ini, dipilih data yang relevan untuk dan dipisahkan dari kumpulan data.
4. Data transformation /transformasi data, juga dikenal sebagai konsolidasi data, adalah fase di mana data yang dipilih ditransformasikan ke dalam bentuk yang sesuai untuk prosedur penambangan.
5. Data mining, merupakan langkah penting di mana teknik yang sesuai diterapkan untuk mengekstrak pola yang berguna.
6. Pattern evaluation /Evaluasi pola, pada langkah ini, pola yang mewakili pengetahuan diidentifikasi berdasarkan langkah-langkah yang diberikan.
7. Knowledge representation /representasi pengetahuan, adalah tahap akhir di mana penemuan pengetahuan dinyatakan secara visual kepada pengguna. Langkah penting ini menggunakan teknik visualisasi untuk membantu pengguna memahami dan menginterpretasikan hasil data mining.

[7]

Association Rule

Analisis asosiasi atau association rule mining adalah teknik data mining untuk menemukan aturan asosiatif antara suatu kombinasi item. Contoh dari aturan asosiatif dari analisa pembelian di suatu pasar swalayan adalah dapat diketahuinya berapa besar kemungkinan seorang pelanggan membeli roti bersamaan dengan susu. Dengan pengetahuan tersebut pemilik pasar swalayan dapat mengatur penempatan barangnya atau merancang kampanye pemasaran dengan memakai kupon diskon untuk kombinasi barang tertentu. Analisis asosiasi menjadi terkenal karena aplikasinya untuk menganalisa isi keranjang belanja di pasar swalayan, sedemikian sehingga analisis asosiasi juga sering disebut dengan istilah market basket analysis [3]. Analisis asosiasi dikenal juga sebagai salah satu teknik data mining yang menjadi dasar dari berbagai teknik data mining lainnya. Khususnya salah satu tahap dari analisis asosiasi yang disebut analisis pola frekuensi tinggi (frequent pattern mining) menarik perhatian banyak peneliti untuk menghasilkan algoritma yang efisien. Penting tidaknya suatu aturan asosiatif dapat diketahui dengan dua parameter: support (nilai penunjang) yaitu persentase kombinasi item tersebut dalam database dan confidence (nilai kepastian) yaitu kuatnya hubungan antar item dalam aturan asosiatif.

Menurut Zaki [8] association rule menghasilkan rules yang menentukan seberapa besar hubungan antar X dan Y, dan diperlukan dua ukuran untuk rules ini, yakni support dan confidence.

Support merupakan kemungkinan X dan Y muncul bersamaan yang dinotasikan:

$$\text{Support } (x \rightarrow Y) = \frac{\text{Jumlah transaksi yang mengandung X dan Y}}{\text{Jumlah Transaksi}}$$

Sedangkan confidence merupakan kemungkinan munculnya Y ketika X juga muncul, dinotasikan :

$$\text{Confidence } (x \rightarrow Y) = \frac{\text{Jumlah transaksi yang mengandung X dan Y}}{\text{Jumlah Transaksi yang mengandung X}}$$

Ada beberapa algoritma yang sudah dikembangkan mengenai aturan asosiasi, namun ada satu algoritma klasik yang sering dipakai yaitu algoritma apriori. Ide dasar dari algoritma ini adalah dengan mengembangkan frequent itemset. Dengan menggunakan satu item dan secara rekursif mengembangkan frequent itemset dengan dua item, tiga item dan seterusnya hingga frequent itemset dengan semua ukuran. Untuk mengembangkan frequent set dengan dua item, dapat menggunakan frequent set item [9].

Algoritma Apriori

Algoritma Apriori adalah sebuah algoritma pencarian pola yang sangat populer dalam teknik penambangan data (data mining). Algoritma ini ditujukan untuk mencari kombinasi item-set yang mempunyai suatu nilai keseringan tertentu sesuai kriteria atau filter yang diinginkan. Hasil dari algoritma ini dapat digunakan untuk membantu dalam pengambilan keputusan pihak manajemen. Berikut diberikan contoh untuk membantu mendapatkan gambaran mengenai cara kerja dari sebuah algoritma Apriori untuk mendapatkan aturan asosiasi (*Association Rule*) dari sejumlah transaksi item.

Algoritma Apriori dibagi menjadi beberapa tahap yang disebut iterasi atau pass. Tiap iterasi menghasilkan pola frekuensi tinggi dengan panjang yang sama dimulai dari pass pertama yang menghasilkan pola frekuensi tinggi dengan panjang satu. Di iterasi pertama ini, support dari setiap item dihitung dengan men-scan database. Setelah support dari setiap item didapat, item yang memiliki support diatas minimum support dipilih

sebagai pola frekuensi tinggi dengan panjang 1 atau sering disingkat 1-itemset. Singkatan k-itemset berarti satu set yang terdiri dari k item.

Iterasi kedua menghasilkan 2-itemset yang tiap set-nya memiliki dua item. Pertama dibuat kandidat 2-itemset dari kombinasi semua 1-itemset. Lalu untuk tiap kandidat 2-itemset ini dihitung support-nya dengan men-scan database. Support disini artinya jumlah transaksi dalam database yang mengandung kedua item dalam kandidat 2-itemset. Setelah support dari semua kandidat 2-itemset didapatkan, kandidat 2-itemset yang memenuhi syarat minimum support dapat ditetapkan sebagai 2-itemset yang juga merupakan pola frekuensi tinggi dengan panjang 2.

Untuk selanjutnya pada iterasi ke-k dapat dibagi lagi menjadi beberapa bagian:

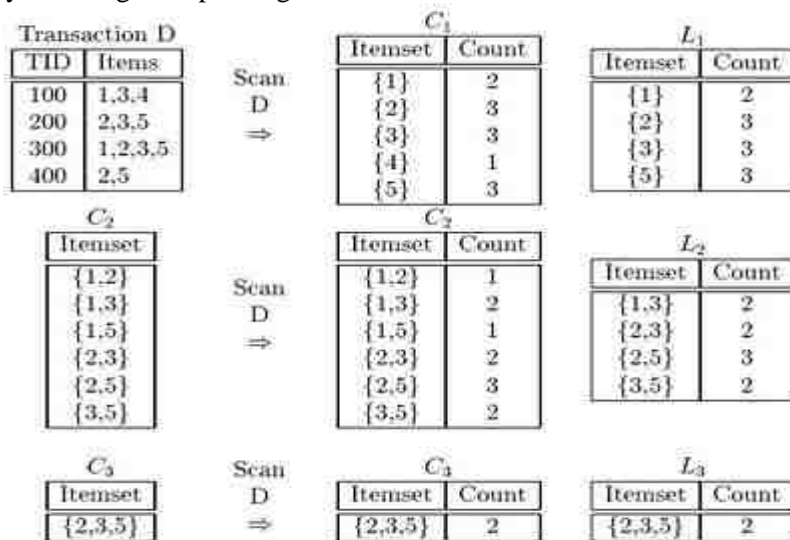
1. Pembentukan kandidat itemset,
Kandidat k-itemset dibentuk dari kombinasi (k-1)-itemset yang didapat dari iterasi sebelumnya. Satu ciri dari algoritma Apriori adalah adanya pemangkasan kandidat k-itemset yang subset-nya yang berisi k-1 item tidak termasuk dalam pola frekuensi tinggi dengan panjang k-1.
2. Penghitungan support dari tiap kandidat k-itemset.
Support dari tiap kandidat k-itemset didapat dengan men-scan database untuk menghitung jumlah transaksi yang memuat semua item di dalam kandidat k-itemset tsb. Ini adalah juga ciri dari algoritme Apriori dimana diperlukan penghitungan dengan scan seluruh database sebanyak k-itemset terpanjang.
3. Tetapkan pola frekuensi tinggi. Pola frekuensi tinggi yang memuat k item atau k-itemset ditetapkan dari kandidat k-itemset yang support-nya lebih besar dari minimum support.
4. Bila tidak didapat pola frekuensi tinggi baru maka seluruh proses dihentikan. Bila tidak, maka k ditambah satu dan kembali ke bagian 1.

Di sini minimum support adalah 50% atau minimal support-nya adalah 2. Pada iterasi pertama, item yang support-nya atau count-nya dibawah 2 dieliminasi dari 1-itemset L1. Kemudian kandidat 2-itemset C2 dari iterasi kedua dibentuk dari cross product item-item yang ada di L1. Setelah kandidat 2-itemset itu dihitung dari database, ditetapkan 2-itemset L2. Proses serupa berulang di iterasi ketiga, tetapi perhatikan bahwa selain {2,3,5} yang menjadi kandidat 3-itemset C3 sebenarnya ada juga itemset {1,2,3} dan {1,3,5} yang dapat diperoleh dari kombinasi item-item di L2, tetapi kedua itemset itu dipangkas karena {2,3}

dan {1,5} tidak ada di L2. Proses ini berulang sampai tidak ada lagi kandidat baru yang dapat dihasilkan di iterasi ke 4.

Dalam contoh ini bisa dilihat bahwa Apriori dapat mengurangi jumlah kandidat yang harus dihitung support-nya dengan pemangkasan.

Misalnya kandidat 3-itemset dapat dikurangi dari 3 menjadi 1 saja. Pengurangan jumlah kandidat ini merupakan sebab utama peningkatan performa Apriori.



Gambar 2. Salah satu contoh algoritma apriori

Pencarian Pola Hubungan dalam Data IPK dan Data Dasar SNMPTN dengan Algoritma Apriori

Data yang diambil dalam penelitian ini adalah data SNMPTN tahun 2010 (sekarang disebut jalur SBMPTN). Sedangkan data IPK yang diambil adalah IPK mahasiswa semester 3. Pemilihan data IPK pada semester 3 ini dengan alasan bahwa mahasiswa pada semester ini, belum banyak mengulang mengambil mata kuliah yang sama, sehingga asumsi IPK yang diperoleh masih heterogen, sesuai dengan kemampuan mahasiswa.

Penelitian ini dilakukan dalam beberapa tahapan.

1. Data cleaning/pembersihan data

Data cleaning/pembersihan data adalah fase di mana noise data dan data yang tidak relevan akan dihapus dari koleksi. Pada tahap ini dilakukan pembersihan dan penyesuaian data pada data mahasiswa, yaitu menghapus data mahasiswa yang diterima namun tidak terdapat pada data IPK pada semester pada saat pengambilan data. Hal ini disebabkan oleh beberapa hal diantaranya adalah, mahasiswa diterima tetapi tidak melakukan registrasi atau mahasiswa yang pindah karena diterima di perguruan tinggi lain

2. Data Integration/Integrasi data

Data Integration/Integrasi data, pada tahap ini, berbagai sumber data sering heterogen, akan dikombinasikan dalam sumber yang sama. Data-data yang terpisah yaitu pada data IPK yang diperoleh dari berbagai program studi diintegrasikan ke dalam satu sumber.

3. Data selection/Pemilihan Data

Data selection/Pemilihan Data, pada langkah ini, dipilih data yang relevan untuk dan dipisahkan dari kumpulan data. Setelah dilakukan pembersihan dan pemilihan data, diperoleh data seperti pada Tabel 1.

4. Data transformation /transformasi data

Data transformation/transformasi data, juga dikenal sebagai konsolidasi data, adalah fase di mana data yang dipilih ditransformasikan ke dalam bentuk yang sesuai untuk prosedur penambangan. Transformasi yang dilakukan dalam penelitian adalah sebagai berikut.

- Tingkat pendidikan orang tua, terdiri dari pendidikan ayah dan pendidikan ibu, yang dikategorikan dalam 8 kategori yaitu: Doktor, Magister, Sarjana, Diploma, Tamat_SLTA, Tamat_SLTP, Tamat_SD, Tidak_Tamat_SD.

- Pekerjaan Orang Tua, terdiri dari pekerjaan ayah dan pekerjaan ibu, yang dikategorikan menjadi 8 kategori, disajikan dalam Tabel 2

Tabel 1 Data Dasar SNMPTN tahun 2010

No	Pendidikan Ayah	Pekerjaan Ayah	Pendidikan Ibu	Pekerjaan Ibu	Penghasilan/ bln	Kab. (SLTA)	Rata UAN	Nilai TPA	IPK
1	Diploma	Pegawai/ Karyawan	Tamat SLTA	Tidak Bekerja	Rp. 1.000.001 - 5.000.000	Kota Surakarta	6.50	84.41	3.57
2	Diploma	Guru	Sarjana/S1	Guru	Rp. 1.000.001 - 5.000.000	Kabupaten Purworejo	6.67	72.25	3.55
3	Diploma	Pegawai/ Karyawan	Tamat SLTA	Tidak Bekerja	Sampai dengan Rp. 1.000.000	Kabupaten Sleman	6.67	88.71	2.78
.....

Tabel 2. Kategori pekerjaan orang tua

Pekerjaan	Kode
Pegawai/Karyawan	K1
Guru	K2
Wiraswasta	K3
Petani/Nelayan	K4
Dosen	K5
Buruh	K6
Tentara/Polisi	K7
Tidak Bekerja	K8

- Penghasilan Orang tua, dikategorikan menjadi 4 kategori, yang disajikan dalam Tabel 3. Nilai rata-rata UAN, dikategorikan dalam 4 Kategori, yang disajikan dalam Tabel 4.

Tabel 3 Kategori penghasilan orang tua

Penghasilan	Kode
Sampai dengan Rp. 1.000.000	P1
Rp. 1.000.001 - Rp. 5.000.000	P2
Rp. 5.000.001 - Rp. 10.000.000	P3
Lebih dari Rp. 10.000.000	P4

Tabel 4 Kategori Nilai UAN

Rata-rata nilai UAN	Kode
Sampai dengan 5,5	1
5,5 ≤ Rata-rata < 7	2

7 ≤ Rata-rata < 8,5 3
 ≥ 8,5 4

- Asal Kabupaten, terdiri dari 127 kabupaten, diantaranya Kab Bangka, Kab Bangka Barat, Kab Banjarnegara, Kab Bantul, Kab Banyuasin, Kab Banyumas, Kab Banyuwangi, Kab Barito Utara, Kab Batang, Kab Belitung, Kab Belitung Timur, Kab Bengkulu Utara, Kab Berau, Kab Blitar, Kab Blora, Kab Boyolali, Kab Brebes, Kab Ciamis, Kab Cilacap, Kab Demak, Kab Flores Timur, Kab Gresik, dll.
- IPK dikelompokkan dalam 4 kategori. Pengelompokan ini didasarkan pada predikat IPK menurut peraturan akademik yang berlaku di UNY, yang disajikan pada Tabel 5. Sedangkan untuk IPK yang kurang dari 2,5 dikategorikan sebagai D.

Tabel 5 Kategori IPK

IPK	Predikat	Kategori
IPK > 3,5	Dengan Pujian (Cumlaude)	A
3 < IPK ≤ 3,5	Sangat Memuaskan	B
2,5 < IPK ≤ 3	Memuaskan	C

- Nilai Tes Potensi Akademik(TPA), dikategorikan dalam 5 kategori, yaitu

Tabel 6 Kategori Nilai TPA

Nilai TPA	Kategori
-----------	----------

TPA > 80	T5
60 < TPA ≤ 80	T4
40 < TPA ≤ 60	T3
20 < TPA ≤ 40	T2
TPA < 20	T1

- Asal SMA, terdiri dari 484 SMA, yang meliputi SMAN 3 Surakarta, SMAN 2 Purworejo, SMAN 1 Depok, SMA Muhammadiyah Mlati, SMA Negeri 6, SMA Negeri 10, SMA Negeri 7, SMA Darul Hikmah, SMA Muhammadiyah 1 Klaten SMAN 1 Kasihan, MAS Sunan Pandanaran, SMAN 1 Kejobong dll.

Hasil Data Mining

Data mining, merupakan langkah penting di mana teknik yang sesuai diterapkan untuk mengekstrak pola yang berguna. Proses mining dalam penelitian ini digunakan untuk mengetahui pola hubungan data dasar SNMPTN dengan Indeks Prestasi Mahasiswa (IPK).

Data dasar SNMPTN meliputi Tingkat pendidikan orang tua, Pekerjaan Orang Tua, Penghasilan Orang tua, Asal SMA, Asal Kabupaten (SMA asal), Nilai Tes Potensi Akademik, dan IPK mahasiswa. Data dianalisis dengan menggunakan software WEKA. Data diubah dalam format .arff,

dalam kasus ini disimpan sebagai **snmpndata.arff** sebagai berikut.

Dari proses transformasi data dan pengaturan penulisan data yang telah disesuaikan dengan format dalam software WEKA, diperoleh format seperti pada Gambar 2. Selanjutnya file snmpndata.arff dipanggil dalam area kerja WEKA. Visualisasi dari masing-masing variabel data disajikan pada Gambar 3.

Selanjutnya, data yang telah dimasukkan diolah untuk mencari assosiation rulenya dengan algoritma apriori. Dalam penelitian ini dipilih minimum Support 10%, artinya bahwa *itemset* yang nilainya kurang dari 10% (150) akan dieliminasi. Konfigurasi minimum support 10%, dan *minimum confidence* = 75%, artinya bahwa hasil assosiation rule yang diharapkan atau ditampilkan hasilnya adalah memiliki minimum konfidensi 75%, pada software WEKA disajikan pada Gambar 7. Pada data asal SMA, terlihat tidak ada data yang nilai *itemset*nya lebih besar atau sama dengan 150, sehingga data SMA akan dieliminasi, dan tidak diikutkan dalam pencarian assosiation rule.

Dengan konfigurasi yang telah ditetapkan sebelumnya, setelah dijalankan dan dicari 50 aturan terbaik yang memiliki konfidensi lebih besar atau sama dengan 75%.

```

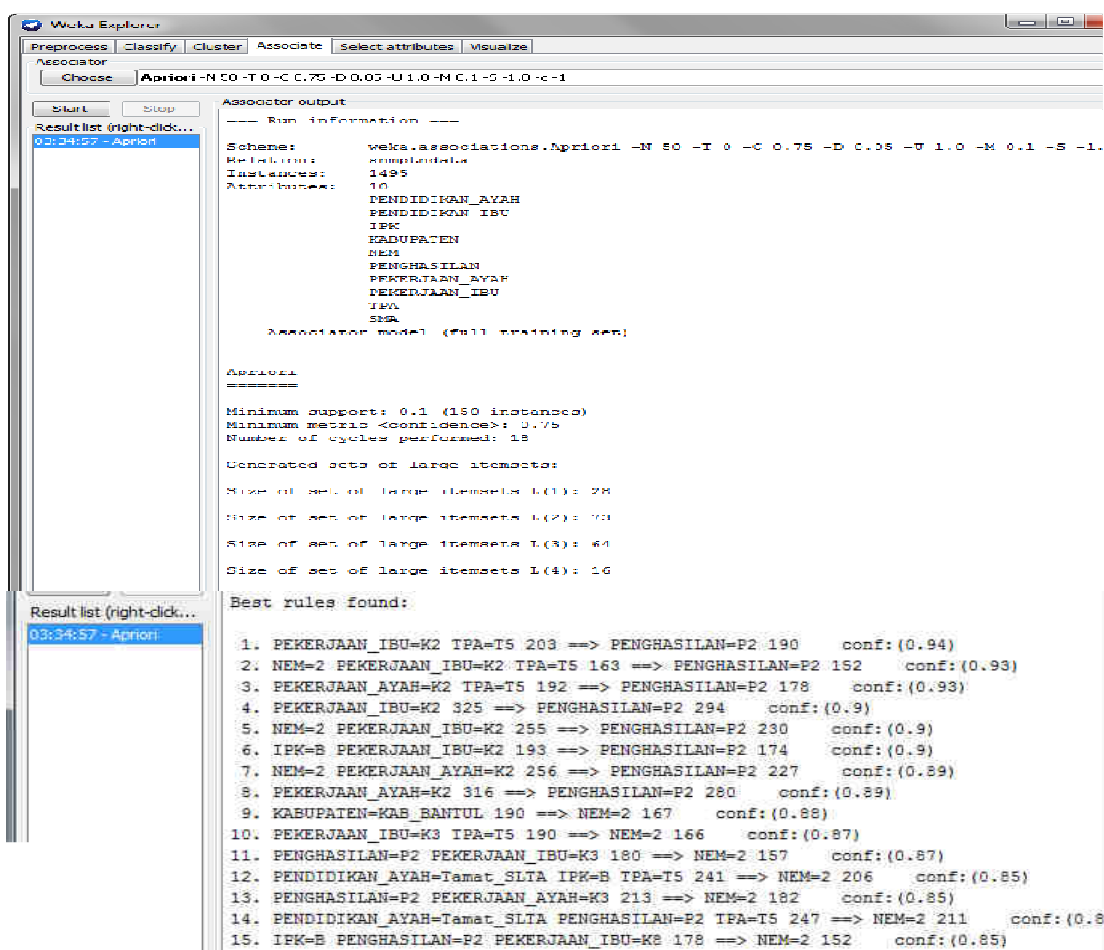
1 @relation snmpndata
2
3 @attribute PENDIDIKAN_AYAH { Diploma , Doktor , Magister , Sarjana , Tamat_SD , Tamat_SLTA , Tamat_SLTP , Tidak_tamat_SD }
4 @attribute PENDIDIKAN_IBU { Diploma , Doktor , Magister , Sarjana , Tamat_SD , Tamat_SLTA , Tamat_SLTP , Tidak_tamat_SD }
5 @attribute IPK { A , B , C , D }
6 @attribute KABUPATEN { KAB_BANGKA , KAB_BANGKA_BARAT , KAB_BANTARNEGARA , KAB_BANTUL , KAB_BANYUASIN , KAB_BANYUMAS , KAB_BANYUWANGI , KAB_BARITO_UTARA , KAB_BATANG ,
7 @attribute NEM { 1 , 2 , 3 , 4 }
8 @attribute PENGHASILAN { P1 , P2 , P3 , P4 }
9 @attribute PEKERJAAN_AYAH { K1 , K2 , K3 , K4 , K5 , K6 , K7 , K8 , K9 }
10 @attribute PEKERJAAN_IBU { K1 , K2 , K3 , K4 , K5 , K6 , K7 , K8 , K9 }
11 @attribute TPA { T1 , T2 , T3 , T4 , T5 }
12 @attribute SMA { mahasanjufri , mamaahid , mamualimmin , mamuallimeat , maniboyolali , mantulungayung , manboanjarneyara , man2kebumen , man2madiun , man2pekalongan ,
13
14
15 @data
16 Tamat_SLTA , Tamat_SLTA , A , KOTA_SURAKARTA , 1 , P2 , K1 , K3 , T5 , sman3surakarta
17 Tamat_SLTP , Tamat_SLTA , A , KAB_PURNOREJO , 1 , P2 , K2 , K2 , T4 , sman2purworejo
18 Tamat_SLTA , Tamat_SLTA , A , KAB_SLEMAN , 1 , P1 , K1 , K3 , T5 , sman1depok
19 Tamat_SLTP , Tamat_SD , A , KAB_SLEMAN , 1 , P1 , K1 , K3 , T4 , smanhammadiyahmlati

```

Gambar 2 Format data .arff



Gambar 3 Visualisasi semua data terhadap IPK



Gambar 4 Output pencarian Association Rule

Dari output yang dihasilkan, diperoleh 50 aturan terbaik, yaitu :

1. PEKERJAAN_IBU=K2 TPA=T5 203 ==> PENGHASILAN=P2 190 conf:(0.94)
2. NEM=2 PEKERJAAN_IBU=K2 TPA=T5 163 ==> PENGHASILAN=P2 152 conf:(0.93)
3. PEKERJAAN_AYAH=K2 TPA=T5 192 ==> PENGHASILAN=P2 178 conf:(0.93)
4. PEKERJAAN_IBU=K2 325 ==> PENGHASILAN=P2 294 conf:(0.9)
5. NEM=2 PEKERJAAN_IBU=K2 255 ==> PENGHASILAN=P2 230 conf:(0.9)
6. IPK=B PEKERJAAN_IBU=K2 193 ==> PENGHASILAN=P2 174 conf:(0.9)
7. NEM=2 PEKERJAAN_AYAH=K2 256 ==> PENGHASILAN=P2 227 conf:(0.89)
8. PEKERJAAN_AYAH=K2 316 ==> PENGHASILAN=P2 280 conf:(0.89)
9. KABUPATEN=KAB_BANTUL 190 ==> NEM=2 167 conf:(0.88)
10. PEKERJAAN_IBU=K3 TPA=T5 190 ==> NEM=2 166 conf:(0.87)
11. PENGHASILAN=P2 PEKERJAAN_IBU=K3 180 ==> NEM=2 157 conf:(0.87)
12. PENDIDIKAN_AYAH=Tamat_SLTA IPK=B TPA=T5 241 ==> NEM=2 206 conf:(0.85)
13. PENGHASILAN=P2 PEKERJAAN_AYAH=K3 213 ==> NEM=2 182 conf:(0.85)
14. PENDIDIKAN_AYAH=Tamat_SLTA PENGHASILAN=P2 TPA=T5 247 ==> NEM=2 211 conf:(0.85)
15. IPK=B PENGHASILAN=P2 PEKERJAAN_IBU=K8 178 ==> NEM=2 152 conf:(0.85)
16. PEKERJAAN_AYAH=K1 TPA=T5 239 ==> NEM=2 204 conf:(0.85)
17. PENDIDIKAN_AYAH=Tamat_SLTA PENDIDIKAN_IBU=Tamat_SLTA PENGHASILAN=P2 195 ==> NEM=2 166 conf:(0.85)
18. PENDIDIKAN_AYAH=Tamat_SLTA PENGHASILAN=P2 374 ==> NEM=2 317 conf:(0.85)
19. PENGHASILAN=P2 PEKERJAAN_IBU=K8 308 ==> NEM=2 261 conf:(0.85)
20. PENDIDIKAN_IBU=Tamat_SLTA IPK=B PENGHASILAN=P2 216 ==> NEM=2 183 conf:(0.85)
21. PENDIDIKAN_AYAH=Tamat_SLTA IPK=B PENGHASILAN=P2 209 ==> NEM=2 177 conf:(0.85)
22. PENDIDIKAN_IBU=Sarjana 248 ==> NEM=2 210 conf:(0.85)
23. IPK=B PENGHASILAN=P2 TPA=T5 351 ==> NEM=2 297 conf:(0.85)
24. IPK=B TPA=T5 597 ==> NEM=2 505 conf:(0.85)
25. PENDIDIKAN_AYAH=Tamat_SLTA TPA=T5 388 ==> NEM=2 328 conf:(0.85)
26. IPK=C 217 ==> NEM=2 183 conf:(0.84)
27. IPK=B PENGHASILAN=P1 TPA=T5 229 ==> NEM=2 193 conf:(0.84)
28. PENDIDIKAN_IBU=Tamat_SLTA PENGHASILAN=P2 TPA=T5 248 ==> NEM=2 209 conf:(0.84)
29. PEKERJAAN_IBU=K8 TPA=T5 363 ==> NEM=2 305 conf:(0.84)
30. PENGHASILAN=P2 TPA=T5 613 ==> NEM=2 515 conf:(0.84)
31. PENDIDIKAN_AYAH=Tamat_SLTA PEKERJAAN_IBU=K8 225 ==> NEM=2 189 conf:(0.84)
32. IPK=B PEKERJAAN_IBU=K8 TPA=T5 225 ==> NEM=2 189 conf:(0.84)
33. TPA=T5 966 ==> NEM=2 810 conf:(0.84)
34. PENGHASILAN=P2 PEKERJAAN_IBU=K8 TPA=T5 210 ==> NEM=2 176 conf:(0.84)
35. PENGHASILAN=P2 TPA=T4 222 ==> NEM=2 186 conf:(0.84)
36. PEKERJAAN_IBU=K3 317 ==> NEM=2 265 conf:(0.84)
37. IPK=B PEKERJAAN_IBU=K3 206 ==> NEM=2 172 conf:(0.83)
38. PENGHASILAN=P1 TPA=T5 327 ==> NEM=2 273 conf:(0.83)
39. PENDIDIKAN_IBU=Tamat_SLTA PENGHASILAN=P2 368 ==> NEM=2 307 conf:(0.83)
40. PENDIDIKAN_IBU=Tamat_SLTP 205 ==> NEM=2 171 conf:(0.83)
41. IPK=A TPA=T5 220 ==> NEM=2 183 conf:(0.83)
42. PENDIDIKAN_AYAH=Tamat_SLTA IPK=B 378 ==> NEM=2 314 conf:(0.83)
43. PENDIDIKAN_AYAH=Tamat_SLTA 616 ==> NEM=2 511 conf:(0.83)
44. PENGHASILAN=P2 932 ==> NEM=2 773 conf:(0.83)
45. IPK=B PENGHASILAN=P2 550 ==> NEM=2 456 conf:(0.83)
46. PENDIDIKAN_IBU=Tamat_SLTA IPK=B TPA=T5 228 ==> NEM=2 189 conf:(0.83)
47. IPK=A PENGHASILAN=P2 TPA=T5 181 ==> NEM=2 150 conf:(0.83)
48. PEKERJAAN_AYAH=K2 TPA=T5 192 ==> NEM=2 159 conf:(0.83)
49. IPK=B PEKERJAAN_AYAH=K1 244 ==> NEM=2 202 conf:(0.83)
50. PEKERJAAN_IBU=K8 534 ==> NEM=2 442 [conf:\(0.83\)](#)

Pattern evaluation /Evaluasi pola

Pada langkah ini, pola yang mewakili pengetahuan diidentifikasi berdasarkan langkah-langkah yang diberikan. Dari aturan asosiasi yang dihasilkan pada langkah Data Mining, terlihat bahwa Aturan assosiasi yang terbaik mempunyai konfidensi 94% muncul pada pola hubungan tingkat penghasilan tipe P2 sebanyak 109, yang muncul bersamaan dengan Pekerjaan Ibu Tipe K2 dan TPA tipe T5 sebanyak 203 data.

Pada aturan assosiasi yang diperoleh, pada pola hubungan yang mengandung variabel IPK (Tipe A, B dan C), sebagian besar muncul

bersamaan variabel NEM tipe 2 ($5,5 \leq \text{Rata-rata UAN} < 7$), lebih khususnya adalah NEM tipe 2 muncul ketika variabel IPK disertai dengan variabel yang lain muncul.

Dari data tersebut tidak terdapat aturan yang menyatakan IPK muncul ketika variabel lain muncul, sehingga dapat dikatakan bahwa tidak ada data dalam data dasar SNMPTN yang secara signifikan mempengaruhi perolehan IPK. Hal ini terlihat dari aturan assosiasi yang ditampilkan pada Gambar 8, dari 50 aturan assosiasi terbaik tidak terlihat munculnya itemset IPK yang disertai dengan munculnya itemset yang lainnya.

Lengkapnya sebagai berikut

15. IPK=B PENGHASILAN=P2 PEKERJAAN_IBU=K8 178 ==> NEM=2 152 conf:(0.85)
23. IPK=B PENGHASILAN=P2 TPA=T5 351 ==> NEM=2 297 conf:(0.85)
24. IPK=B TPA=T5 597 ==> NEM=2 505 conf:(0.85)
26. IPK=C 217 ==> NEM=2 183 conf:(0.84)
27. IPK=B PENGHASILAN=P1 TPA=T5 229 ==> NEM=2 193 conf:(0.84)
32. IPK=B PEKERJAAN_IBU=K8 TPA=T5 225 ==> NEM=2 189 conf:(0.84)
41. IPK=A TPA=T5 220 ==> NEM=2 183 conf:(0.83)
45. IPK=B PENGHASILAN=P2 550 ==> NEM=2 456 conf:(0.83)
47. IPK=A PENGHASILAN=P2 TPA=T5 181 ==> NEM=2 150 conf:(0.83)
49. IPK=B PEKERJAAN_AYAH=K1 244 ==> NEM=2 202 [conf:\(0.83\)](#)

Simpulan

Aturan assosiasi yang terbaik mempunyai konfidensi 94% muncul pada pola hubungan tingkat penghasilan tipe P2 sebanyak 109, yang muncul bersamaan dengan Pekerjaan Ibu Tipe K2 dan TPA tipe T5 sebanyak 203 data.

Pada aturan assosiasi yang diperoleh, tidak ada data dalam data dasar SNMPTN yang secara signifikan mempengaruhi perolehan IPK. Hal ini terlihat dari aturan assosiasi yang dihasilkan tidak terlihat munculnya itemset IPK yang disertai dengan munculnya itemset yang lainnya.

Pada analisis selanjutnya dapat menggunakan data yang lebih besar lagi, misalnya dengan mengambil data mahasiswa yang masuk melalui jalur selain SNMPTN, juga pada angkatan yang lain dan semester yang lain sehingga rules yang dihasilkan lebih beragam dan lebih berguna untuk pengambilan keputusan di perguruan tinggi, karena semakin besar data semakin berguna informasi yang dihasilkan.

Ucapan Terima Kasih

Terima kasih penulis ucapkan kepada Dekan FMIPA UNY atas bantuan dana dan fasilitasi pelaksanaan penelitian. Terima kasih juga penulis sampaikan kepada Yuliati, M.Kes, Dr. Tien Aminatun yang ikut berperan dalam sumbangan penelitian di desa Donotirto, Kretek, Kabupaten Bantul.

Pustaka

- [1] Larose, D. T. (2005). *Discovering Knowledge in Data: An Introduction to Data mining*. New Jersey: John Wiley & Sons. Inc.
- [2] Abidi, & Yu-N. (2000). A Convergence of Knowledge Management and Data Mining: Toward 'Knowledge-Driven' Strategic Services. 3rd International Convergence on the Practical Applications of Knowledge Management (PAKEM 2000). Manchester.
- [3] Agrawal, R., & Srikant, R. (1994). Fast Algorithms For Mining Association Rules.

- In Proc.1994 . *Proc. International Conference Very Large DataBases (VLDB)*.
- [4] Jan, J., & Kamber, M. (2006). *Data Mining: Concepts and Techniques, Second Edition*. San Francisco: Morgan Kaufmann Publishers.
 - [5] Wandu, N., Hendrawan, R. A., & Mukhlason, A. (2012). Pengembangan Sistem Rekomendasi Penelusuran Buku dengan Penggalan Association Rule Menggunakan Algoritma Apriori (Studi Kasus Badan Perpustakaan dan Kearsipan Provinsi Jawa Timur). *JURNAL TEKNIK POMITS Vol. 1, No. 1* , 1-5
 - [6] Ponniah, P. (2001). *Datawarehouse Fundamentals: A comprehensive Guide for IT Professional*. New York: John Willey & Sons. Inc.
 - [7] Zaïane, O. R. (1999). *Principles of Knowledge Discovery in Databases*. Canada: University of Alberta.
 - [8] Santosa, B. (2007). *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*”.. Yogyakarta.: Graha Ilmu
 - [9] Gorunescu, F. (2011). *Data Mining Concepts, Models and Techniques*. Verlag Berlin: Springer.
 - [10] Kantardzic, M. (2003). *Data Mining: Concepts, Models, Methods, and Algorithms*. New Jersey: John Wiley & Sons.