

# Penggunaan *Association Rule Data Mining* Untuk Menentukan Pola Lama Studi Mahasiswa F-MIPA UNSRAT

<sup>1</sup>Muhamad Zainal M, <sup>1</sup>Altien J Rindengan, <sup>1</sup>Winsy Ch. D. Weku

<sup>1</sup>Program Studi Matematika, F-MIPA, UNSRAT, al\_intermilan@yahoo.co.id

<sup>1</sup>Program Studi Matematika, F-MIPA, UNSRAT, altien@unsrat.ac.id

<sup>1</sup>Program Studi Matematika, F-MIPA, UNSRAT, winsyweku@gmail.com

## Abstract

The requirement of highest information sometimes is not balance with the provision of adequate information, so that the information must be re-excavated in large data. By using the technique of association rule we can obtain information from large data such as the college data. The purposes of this research is to determine the patterns of study from student in F-MIPA UNSRAT by using association rule method of data mining algorithms and to compare in the apriori method and a hash-based algorithms. The major's student data of F-MIPA UNSRAT as a data were processed by association rule method of data mining with the apriori algorithm and a hash-based algorithm by using support and confidence at least 1 %. The results of processing data with apriori algorithms was same with the processing results of hash-based algorithms is as much as 49 combinations of 2-itemset. The pattern that formed between 7,5% of graduates from mathematics major that studied for more 5 years with confidence value is 38,5%.

*Keywords:* Apriori algorithm, hash-based algorithm, association rule, data mining.

## Abstrak

Kebutuhan informasi yang sangat tinggi terkadang tidak diimbangi dengan pemberian informasi yang memadai, sehingga informasi tersebut harus kembali digali dalam data yang besar. Dengan menggunakan teknik *association rule* kita dapat memperoleh informasi dari data yang besar seperti data yang ada di perguruan tinggi. Tujuan penelitian ini adalah menentukan pola lama studi mahasiswa F-MIPA UNSRAT dengan menggunakan metode *association rule data mining* serta membandingkan algoritma *apriori* dan algoritma *hash-based*. Data yang digunakan adalah data induk mahasiswa F-MIPA UNSRAT yang diolah menggunakan teknik *association rule data mining* dengan algoritma *apriori* dan algoritma *hash-based* dengan *minimum support* 1% dan *minimum confidence* 1%. Hasil pengolahan data dengan algoritma *apriori* sama dengan hasil pengolahan data dengan algoritma *hash-based* yaitu sebanyak 49 kombinasi *2-itemset*. Pola yang terbentuk antara lain 7,5% lulusan yang berasal dari jurusan matematika menempuh studi selama lebih dari 5 tahun dengan nilai *confidence* 38,5%.

Kata kunci : *Association rule data mining*, algoritma *apriori*, algoritma *hash-based*

## 1. Pendahuluan

Informasi menjadi sebuah kebutuhan yang sangat penting bagi manusia untuk menunjang segala aktivitas yang dilakukan. Terkadang, kebutuhan informasi yang sangat tinggi tidak diimbangi dengan pemberian informasi yang memadai. Akibatnya informasi tersebut harus kembali digali dalam data yang besar. Salah satu metode yang biasa digunakan dalam mencari informasi dalam data yang besar yaitu *data mining*. Dalam *data mining* terdapat beberapa teknik salah satunya yaitu *association rule*. Dengan menggunakan teknik *association rule* kita dapat memperoleh informasi dari data yang besar seperti data mahasiswa yang ada di perguruan tinggi.

Setiap perguruan tinggi selalu berusaha meningkatkan kualitas yang dimiliki agar dapat bersaing dengan perguruan tinggi lainnya. Salah satu cara yang dilakukan perguruan tinggi untuk meningkatkan kualitas yang dimiliki yaitu melakukan evaluasi, dengan evaluasi tersebut perguruan tinggi dapat memperbaiki kekurangan-kekurangan yang sebelumnya terjadi. Salah satu bahan evaluasi adalah mengenai lama studi mahasiswa.

UNSRAT memiliki banyak mahasiswa yang terbagi kedalam beberapa fakultas. F-MIPA merupakan salah satu fakultas yang terdapat di UNSRAT dan telah menghasilkan banyak lulusan dengan lama studi yang bervariasi. Oleh karena itu, dengan memanfaatkan data induk mahasiswa dan

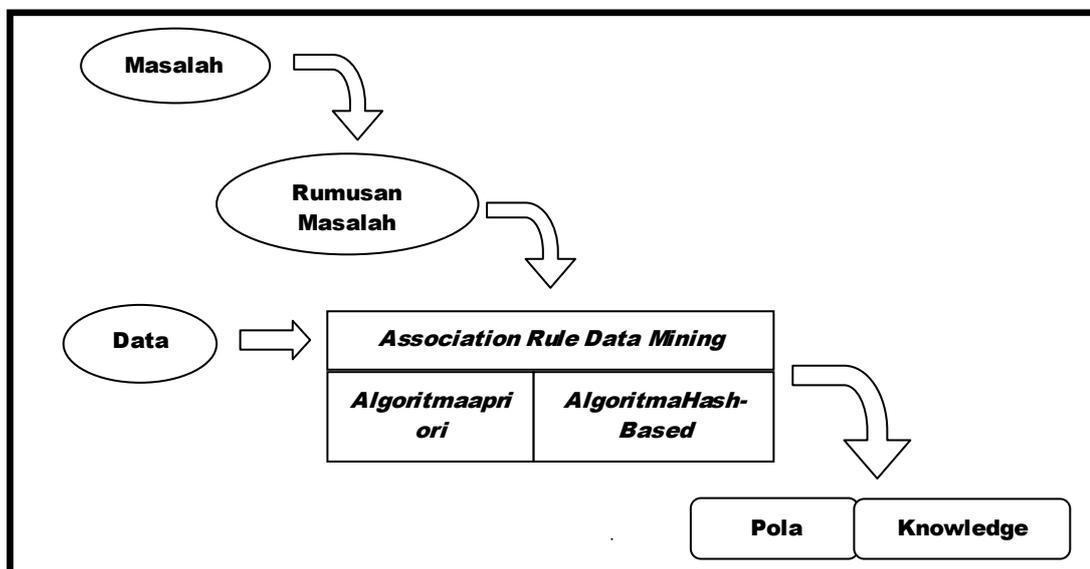
data kelulusan mahasiswa, dapat diketahui informasi pola lama studi mahasiswa melalui teknik *data mining*.

## 2. Metode Penelitian

Pada penelitian ini akan digunakan teknik *association rule data mining* dengan dua algoritma yaitu algoritma *apriori* dan algoritma *hash-based*. Menurut Turban *et al* (2005) *Data mining* merupakan proses yang menggunakan teknik statistik, matematika, kecerdasan buatan dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terakit dari berbagai *database* besar. Langkah-langkah pada *association rule* (Ulmer, 2002) :

1.  $SetK=1$
2. Hitung semua itemset berukuran  $K$
3. Hitung *support* untuk semua kandidat itemset - menyaring itemset berdasarkan minimum *support*
4. Gabungkan semua itemset berukuran  $K$  untuk menghasilkan calon itemset ukuran  $K+1$
5.  $SetK=K+1$
6. Ulangi langkah 3 sampai 5 sampai tidak ada itemset yang lebih besar dapat dibentuk
7. Menghasilkan set akhir itemset dengan menciptakan persatuan semua itemset berukuran  $K$ .

### 2.1. Desain Arsitektur Penelitian



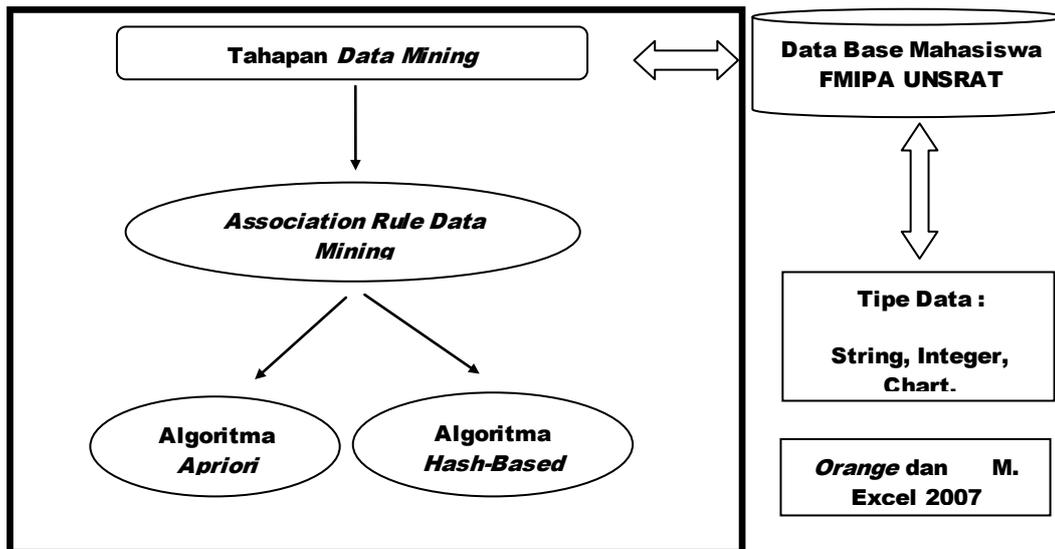
Gambar 1. Model Penelitian Berdasarkan *Waterfall*

Berdasarkan masalah yang ada, rumusan masalah yang akan di bahas pada penelitian ini yaitu menentukan pola lama studi mahasiswa F-MIPA UNSRAT, serta melihat perbandingan 2 algoritma yang digunakan yaitu algoritma *apriori* dan algoritma *hash-based*. Penelitian ini menggunakan teknik *association rule data mining* untuk mengelola data dengan menggunakan dua algoritma yaitu algoritma *apriori* dan algoritma *hash-based*. Hasil yang akan diperoleh dari pengolahan data ini adalah pola lama studi.

Menurut Han (2006), ada beberapa tahap *data mining* yaitu :

1. Pembersihan data (*data cleaning*)
2. Integrasi data (*data integration*)
3. Seleksi data (*data selection*)
4. Transformasi data (*data transformation*)
5. Proses *mining*: proses *mining* akan menggunakan dua algoritma yaitu algoritma *apriori* dan algoritma *hash-based*.
6. Evaluasi pola (*pattern evaluation*)
7. Presentasi pengetahuan (*knowledge presentation*)

## 2.2. Desain Arsitektur Metode



Gambar 2. Proses Pengolahan Data Dalam *Data Mining*

*Data base* mahasiswa F-MIPA UNSRAT dibagi menjadi beberapa kategori antara lain, IPK dibagi menjadi  $IPK > 3,5$ ,  $3 \leq IPK \leq 3,5$ ,  $2,5 \leq IPK < 3$ ,  $IPK < 2,5$ , lama lulus dibagi menjadi lama lulus  $\leq 4$  tahun,  $4 \text{ tahun} < \text{lama lulus} \leq 5$  tahun, lama lulus  $> 5$  tahun, jalur masuk dibagi menjadi jalur masuk SPMB, jalur masuk SUMIKOLA, jalur masuk T2, jalur masuk KEMITRAAN, jenis kelamin dibagi menjadi jenis kelamin laki-laki, jenis kelamin perempuan, jurusan dibagi menjadi jurusan biologi, jurusan matematika, jurusan kimia, jurusan fisika, jurusan farmasi, asal sekolah dibagi menjadi asal sekolah kabupaten, asal sekolah kecamatan. Pada tahap analisis data, data mahasiswa F-MIPA UNSRAT yang didapatkan akan diolah dengan *software Orange* dan Microsoft Office Excel 2007 menggunakan *data mining*.

## 3. Hasil dan Pembahasan

### 3.1. Algoritma *apriori*

Data yang diperoleh dari SIM (sistem informasi mahasiswa) F-MIPA (data sekunder) berupa nama, jalur masuk Universitas Sam Ratulangi, IPK, lama studi, jenis kelamin, dan jurusan sertawawancara (data primer) berupa asal sekolah diberikan simbol sebagai berikut :

Simbol	Keterangan	Simbol	Keterangan
IPK	: IPK	K	: Jalur masuk KEMITRAAN
A	: $IPK > 3,5$	J.K	: Jenis kelamin
B	: $3 \leq IPK \leq 3,5$	L	: Jenis kelamin laki-laki
C	: $2,5 \leq IPK < 3$	P	: Jenis kelamin perempuan
D	: $IPK < 2,5$	Jurusan	: Jurusan
LL	: Lama Lulus	Q	: Jurusan Biologi
E	: Lama lulus $\leq 4$ tahun	R	: Jurusan Matematika
F	: $4 \text{ tahun} < \text{lama lulus} \leq 5$ tahun	S	: Jurusan Kimia
G	: Lama lulus $> 5$ tahun	T	: Jurusan Fisika
J.M	: Jalur masuk	U	: Jurusan Farmasi
H	: Jalur masuk SPMB	SEKOLAH	: Asal sekolah
I	: Jalur masuk SUMIKOLA	X	: Asal sekolah Kabupaten
J	: Jalur masuk T2	Y	: Asal sekolah Kecamatan

Dengan menggunakan *software Orange* data tersebut diolah dan menghasilkan kombinasi *itemset* sampai kombinasi  $\delta$ -*itemset*. Berikut merupakan beberapa kombinasi *itemset* seperti pada tabel 1 dan tabel 2.

Tabel 1. Beberapa kombinasi 2-Itemset

2-Itemset	Support	Confidence
J.K=P -> LL=F	25,2%	36,9%
J.K=L -> LL=G	13,8%	43,5%
SEKOLAH=X -> LL=G	21,8%	38,3%
SEKOLAH=Y -> LL=G	16%	37,2%
J. M=J -> LL=G	33%	51,2%
J. M=I -> LL=F	4,2%	53,2%
J. M=H -> LL=F	3,2%	42,2%
J. M=K -> LL=E	18,3%	90,9%
JURUSAN=S -> LL=G	14,3%	44,8%
JURUSAN=Q -> LL=G	13,3%	43,2%
JURUSAN=R -> LL=G	7,5%	38,5%
JURUSAN=U -> LL=F	5%	52,6%
JURUSAN=T -> LL=E	3,8%	46,9%
IPK=C -> LL=G	18%	48,2%
IPK=B -> LL=F	17,5%	43,4%
IPK=D -> LL=G	6,5%	65%
IPK=A -> LL=E	5%	40,5%

Support dan Confidence dari kombinasi 2-itemset diatas diperoleh menggunakan rumus seperti berikut :

$$\text{Support}(X \cap G) = \frac{\text{Jumlah lulusan mengandung X dan G}}{\text{Total lulusan}}$$

$$\text{Support}(X \cap G) = \frac{131}{600} = 0,218 = 21,8\%$$

$$\text{Confidence} = P(G|X) = \frac{\text{Jumlah lulusan mengandung X dan G}}{\text{Jumlah lulusan mengandung X}}$$

$$\text{Confidence} = P(G|X) = \frac{131}{342} = 0,383 = 38,3\%$$

Rumus yang sama digunakan untuk mencari nilai Support dan confidence dari kombinasi 2-itemset lainnya.

1. Kombinasi 2-itemset berdasarkan asal sekolah
  - a) Jika X (asal sekolah kabupaten ) maka G (lama lulus > 5 tahun) dengan nilai support 21,8% dan confidence 38,3% artinya sebanyak 21,8% lulusan yang memiliki asal sekolah kabupaten menempuh studi selama lebih dari 5 tahun dengan nilai confidence 38,3%.
  - b) Jika Y (asal sekolah kecamatan ) maka G (lama lulus > 5 tahun) dengan nilai support 16% dan confidence 37,2% artinya sebanyak 16% lulusan yang memiliki asal sekolah kecamatan menempuh studi selama lebih dari 5 tahun dengan nilai confidence 37,2%.
  - c) Berdasarkan pola diatas dapat dilihat bahwa alumni F-MIPA yang berasal dari kabupaten dan kecamatan cenderung memiliki lama studi yang sama yaitu lebih dari 5 tahun.
2. Kombinasi 2-itemset berdasarkan jenis kelamin
  - a) Jika P (jenis kelamin perempuan) maka F (4 tahun < lama lulus ≤ 5 tahun) dengan nilai support 25,2% dan confidence 36,9% artinya sebanyak 25,2% lulusan yang berjenis kelamin perempuan menempuh studi selama lebih dari 4 tahun dengan nilai confidence 36,9%.

- b) Jika L (jenis kelamin laki-laki) maka G (lama lulus > 5 tahun) dengan nilai *support* 13,8% dan *confidence* 43,5% artinya sebanyak 13,8% lulusan yang berjenis kelamin laki-laki menempuh studi selama lebih dari 5 tahun dengan nilai *confidence* 43,5%
- c) Berdasarkan pola diatas dapat dilihat bahwa alumni F-MIPA yang berjenis kelamin perempuan cenderung lebih cepat lulus di dibandingkan dengan alumni F-MIPA yang berjenis kelamin laki-laki.
3. Kombinasi 2-*itemset* berdasarkan jalur masuk
- a) Jika J (jalur masuk T2) maka G (lama lulus > 5 tahun) dengan nilai *support* 33% dan *confidence* 51,2% artinya sebanyak 33% lulusan yang masuk melalui jalur masuk T2 menempuh studi selama lebih dari 5 tahun dengan nilai *confidence* 51,2%.
- b) Jika K (jalur masuk kemitraan) maka E (lama lulus  $\leq$  4 tahun) dengan nilai *support* 18,3% dan *confidence* 90,9% artinya sebanyak 18,3% lulusan yang masuk melalui jalur masuk kemitraan menempuh studi selama kurang dari 4 tahun dengan nilai *confidence* 90,9%.
- c) Jika I (jalur masuk sumikola) maka F (4 tahun < lama lulus  $\leq$  5 tahun) dengan nilai *support* 4,2% dan *confidence* 53,2% artinya sebanyak 4,2% lulusan yang masuk melalui jalur masuk sumikola menempuh studi selama lebih dari 4 tahun dengan nilai *confidence* 53,2%.
- d) Jika H (jalur masuk SPMB) maka F (4 tahun < lama lulus  $\leq$  5 tahun) dengan nilai *support* 3,2% dan *confidence* 42,2% artinya sebanyak 3,2% lulusan yang masuk melalui jalur masuk SPMB menempuh studi selama lebih dari 4 tahun dengan nilai *confidence* 42,2%.
- e) Berdasarkan pola diatas dapat dilihat bahwa alumni F-MIPA yang masuk melalui jalur masuk kemitraan cenderung lebih cepat lulus di ikuti dengan jalur masuk sumikola dan jalur masuk SPMB, kemudian di ikuti oleh jalur masuk T2.
4. Kombinasi 2-*itemset* berdasarkan jurusan
- a) Jika S (jurusan kimia) maka G (lama lulus > 5 tahun) dengan *support* 14,3% dan *confidence* 44,8% artinya sebanyak 14,3% lulusan yang berasal dari jurusan kimia menempuh studi selama lebih dari 5 tahun dengan nilai *confidence* 44,8%.
- b) Jika Q (jurusan biologi) maka G (lama lulus > 5 tahun) dengan *support* 13,3% dan *confidence* 43,2% artinya sebanyak 13,3% lulusan yang berasal dari jurusan biologi menempuh studi selama lebih dari 5 tahun dengan nilai *confidence* 43,2%.
- c) Jika R (jurusan matematika) maka G (lama lulus > 5 tahun) dengan *support* 7,5% dan *confidence* 38,5% artinya sebanyak 7,5% lulusan yang berasal dari jurusan matematika menempuh studi selama lebih dari 5 tahun dengan nilai *confidence* 38,5%.
- d) Jika U (jurusan farmasi) maka F (4 tahun < lama lulus  $\leq$  5 tahun) dengan *support* 5% dan *confidence* 52,6% artinya sebanyak 5% lulusan yang berasal dari jurusan farmasi menempuh studi selama lebih dari 4 tahun dengan nilai *confidence* 52,6%.
- e) Jika T (jurusan fisika) maka E (lama lulus  $\leq$  4 tahun) dengan *support* 3,8% dan *confidence* 46,9% artinya sebanyak 3,8% lulusan yang berasal dari jurusan fisika menempuh studi selama kurang dari 4 tahun dengan nilai *confidence* 46,9%.
- f) Berdasarkan pola diatas dapat dilihat bahwa alumni F-MIPA yang berasal dari jurusan fisika cenderung lebih cepat lulus diikuti jurusan farmasi, kemudian jurusan matematika, jurusan kimia dan jurusan biologi.
5. Kombinasi 2-*itemset* berdasarkan IPK
- a) Jika C ( $2,5 \leq$  IPK < 3) maka G (lama lulus > 5 tahun) dengan *support* 18% dan *confidence* 48,2% artinya sebanyak 18% lulusan yang memiliki IPK diantara 2,5 sampai kurang dari 3 menempuh studi selama lebih dari 5 tahun dengan nilai *confidence* 48,2%.
- b) Jika B ( $3 \leq$  IPK  $\leq$  3,5) maka F (4 tahun < lama lulus  $\leq$  5 tahun) dengan *support* 17,5% dan *confidence* 43,4% artinya sebanyak 17,5% lulusan yang memiliki IPK diantara 3 sampai 3,5 menempuh studi selama lebih dari 4 tahun dengan nilai *confidence* 43,4%.
- c) Jika D (IPK < 2,5) maka G (lama lulus > 5 tahun) dengan *support* 6,5% dan *confidence* 65% artinya sebanyak 6,5% lulusan yang memiliki IPK kurang dari 2,5 menempuh studi selama lebih dari 5 tahun dengan nilai *confidence* 65%.

- d) Jika A (IPK > 3,5) maka E (lama lulus ≤ 4 tahun) dengan *support* 5% dan *confidence* 40,5% artinya sebanyak 5% lulusan yang memiliki IPK lebih dari 3,5 menempuh studi selama kurang dari 4 tahun dengan nilai *confidence* 40,5%.
- e) Berdasarkan pola diatas dapat dilihat bahwa alumni F-MIPA yang memiliki IPK lebih dari 3,5 cenderung lebih cepat lulus diikuti IPK diantara 2,5 sampai kurang dari 3, kemudian alumninyang memiliki IPK diantara 3 sampai 3,5 dan alumni yang memiliki IPK kurang dari 2,5.

Tabel 2. Beberapa kombinasi 3-Itemset

3-itemset	Support	Confidence
JURUSAN=R J.K=L -> LL=G	3,5%	48,8%
JURUSAN=R J.K=P -> LL=E	5,3%	43,2%
JURUSAN=R SEKOLAH=Y -> LL=E	3,2%	61,3%
JURUSAN=R SEKOLAH=X -> LL=G	6,5%	45,3%
JURUSAN=R J. M=H -> LL=G	1,3%	57,1%
JURUSAN=R J. M=J -> LL=G	5,8%	58,3%
JURUSAN=R J. M=K -> LL=E	5,2%	96,9%
JURUSAN=R J. M=I -> LL=F	1%	54,5%
JURUSAN=R IPK=A -> LL=E	1,2%	58,3%
JURUSAN=R IPK=B -> LL=G	4,3%	51%
JURUSAN=R IPK=C -> LL=E	4,5%	52,9%

*Support* dan *Confidence* dari kombinasi 3-itemset diperoleh menggunakan rumus seperti berikut :

$$\text{Support}(R \cap K \cap G) = \frac{\text{Jumlah lulusan mengandung R K dan G}}{\text{Total lulusan}}$$

$$\text{Support}(R \cap K \cap G) = \frac{21}{600} = 0,035 = 3,5\%$$

$$\text{Confidence} = P(G|R \cap K) = \frac{\text{Jumlah lulusan mengandung R K dan G}}{\text{Jumlah lulusan mengandung R K}}$$

$$\text{Confidence} = P(G|J \cap X) = \frac{21}{43} = 0,488 = 48,8\%$$

Rumus yang sama digunakan untuk mencari nilai *Support* dan *confidence* dari kombinasi 3-itemset lainnya.

- Kombinasi 3-itemset berdasarkan jenis kelamin
  - Jika R (jurusan matematika)P (jenis kelamin perempuan) maka E (lama lulus ≤ 4 tahun) dengan nilai *support* 5,3% dan *confidence* 43,2% artinya sebanyak 5,3% lulusan yang berasal dari jurusan matematika dan berjenis kelamin perempuan maka menempuh studi selama kurang dari 4 tahun dengan nilai *confidence* 43,2%.
  - Jika R (jurusan matematika)L (jenis kelamin laki-laki) maka G (lama lulus > 5 tahun) dengan nilai *support* 3,5% dan *confidence* 48,8% artinya sebanyak 3,5% lulusan yang berasal dari jurusan matematika dan berjenis kelamin laki-laki maka menempuh studi selama lebih dari 5 tahun dengan nilai *confidence* 48,8%.
  - Berdasarkan pola diatas dapat dilihat bahwa alumni jurusan matematika yang berjenis kelamin perempuan cenderung lebih cepat lulus di bandingkan dengan laki-laki.
- Kombinasi 3-itemset berdasarkan asal sekolah
  - Jika R (jurusan matematika) Y (asal sekolah kecamatan ) maka E (lama lulus ≤ 4 tahun) dengan nilai *support* 3,2% dan *confidence* 61,3% artinya sebanyak 3,2% lulusan yang

- berasal dari jurusan matematika dan memiliki asal sekolah kecamatan makamenempuh studi selama kurang dari 4 tahun dengan nilai *confidence* 61,3%.
- b) Jika R (jurusan matematika) X (asal sekolah kabupaten) maka G (lama lulus > 5 tahun) dengan nilai *support* 6,5% dan *confidence* 45,3% artinya sebanyak 6,5% lulusan yang berasal dari jurusan matematika dan memiliki asal sekolah kabupaten makamenempuh studi selama lebih dari 5 tahun dengan nilai *confidence* 45,3%.
  - c) Berdasarkan pola diatas dapat dilihat bahwa alumni jurusan matematika yang asal sekolah kecamatan cenderung lebih cepat lulus di bandingkan dengan asal sekolah kabupaten.
3. Kombinasi *3-itemset* berdasarkan jalur masuk
    - a) Jika R (jurusan matematika)K (jalur masuk kemitraan) maka E (lama lulus  $\leq 4$  tahun) dengan nilai *support* 5,2% dan *confidence* 96,9% artinya sebanyak 5,2% lulusan yang berasal dari jurusan matematika dan masuk melalui jalur masuk kemitraan makamenempuh studi selama kurang dari 4 tahun dengan nilai *confidence* 96,9%.
    - b) Jika R (jurusan matematika)I (jalur masuk sumikola) maka F (4 tahun < lama lulus  $\leq 5$  tahun) dengan nilai *support* 1% dan *confidence* 54,5% artinya sebanyak 1% lulusan yang berasal dari jurusan matematika dan masuk melalui jalur masuk sumikola makamenempuh studi selama lebih dari 4 tahun dengan nilai *confidence* 54,5%.
    - c) Jika R (jurusan matematika) J (jalur masuk T2) maka G (lama lulus > 5 tahun) dengan nilai *support* 5,8% dan *confidence* 58,3% artinya sebanyak 5,8% lulusan yang berasal dari jurusan matematika dan masuk melalui jalur masuk T2 makamenempuh studi selama lebih dari 5 tahun dengan nilai *confidence* 58,3%.
    - d) Jika R (jurusan matematika)H (jalur masuk SPMB) maka G (lama lulus > 5 tahun) dengan nilai *support* 1,3% dan *confidence* 57,1% artinya sebanyak 1,3% lulusan yang berasal dari jurusan matematika dan masuk melalui jalur masuk SPMB makamenempuh studi selama lebih dari 5 tahun dengan nilai *confidence* 57,1%.
    - e) Berdasarkan pola diatas dapat dilihat bahwa alumni jurusan matematika yang masuk melalui jalur masuk kemitraan cenderung lebih cepat lulus di ikuti dengan jalur masuk sumikola kemudian di ikuti oleh jalur masuk T2 dan jalur masuk SPMB.
  4. Kombinasi *3-itemset* berdasarkan IPK
    - a) Jika R (jurusan matematika)A (IPK > 3,5) maka E (lama lulus  $\leq 4$  tahun) dengan nilai *support* 1,2% dan *confidence* 58,3% artinya sebanyak 1,2% lulusan yang berasal dari jurusan matematika dan memiliki IPK lebih dari 3,5 makamenempuh studi selama kurang dari 4 tahun dengan nilai *confidence* 58,3%.
    - b) Jika R (jurusan matematika)C ( $2,5 \leq \text{IPK} < 3$ ) maka E (lama lulus  $\leq 4$  tahun) dengan nilai *support* 4,5% dan *confidence* 52,9% artinya sebanyak 4,5% lulusan yang berasal dari jurusan matematika dan memiliki IPK diantara 2,5 sampai kurang dari 3 makamenempuh studi selama kurang dari 4 tahun dengan nilai *confidence* 52,9%.
    - c) Jika R (jurusan matematika)B ( $3 \leq \text{IPK} \leq 3,5$ ) maka G (lama lulus >5 tahun) dengan nilai *support* 4,3% dan *confidence* 51% artinya sebanyak 4,3% lulusan yang berasal dari jurusan matematika dan memiliki IPK diantara 3 sampai 3,5 makamenempuh studi selama lebih dari 5 tahun dengan nilai *confidence* 51%.
    - d) Berdasarkan pola diatas dapat dilihat bahwa alumni jurusan matematika yang memiliki IPK lebih dari 3,5 dan IPK diantara 2,5 sampai kurang dari 3 cenderung lebih cepat lulus dibandingkan alumniyang memiliki IPK diantara 3 sampai 3,5.

### 3.2. Algoritma *hash-based*

Data yang sama yang digunakan pada algoritma *apriori* diolah kembali menggunakan algoritma *hash-based* yang diolah dalam Microsoft Office Excel 2007. Pola yang dihasilkan oleh algoritma *hash-based* sama dengan pola yang dihasilkan oleh algoritma *apriori*. Pada algoritma *hash-based* pembangkitan *2-itemset* dihasilkan menggunakan tabel *hash* berbeda dengan pembangkitan *2-itemset* algoritma *apriori* yang dihasilkan menggunakan *minimum support*, namun tidak terdapat perbedaan *2-itemset* yang dihasilkan oleh kedua algoritma.

#### 4. Kesimpulan

1. Dengan menggunakan *association ruledata mining* dihasilkan pola lama studi mahasiswa F-MIPA UNSRAT. Beberapa pola lama studi mahasiswa F-MIPA UNSRAT :
  - a) Kombinasi *2-itemset* antara lain sebanyak 18,3% mahasiswa yang masuk F-MIPA UNSRAT melalui jalur masuk kemitraan menempuh studi di F-MIPA UNSRAT selama kurang dari 4 tahun dengan nilai *confidence* sebesar 90,9%.
  - b) Kombinasi *2-itemset* antara lain sebanyak 7,5% lulusan yang berasal dari jurusan matematika menempuh studi selama lebih dari 5 tahun dengan nilai *confidence* 38,5%.
  - c) Kombinasi *2-itemset* antara lain sebanyak 5% lulusan yang memiliki IPK lebih dari 3,5 menempuh studi selama kurang dari 4 tahun dengan nilai *confidence* 40,5%.
  - d) Kombinasi *3-itemset* antara lain sebanyak 1,2% lulusan yang berasal dari jurusan matematika dan memiliki IPK lebih dari 3,5 maka menempuh studi selama kurang dari 4 tahun dengan nilai *confidence* 58,3%.
  - e) Kombinasi *3-itemset* antara lain sebanyak 5,2% lulusan yang berasal dari jurusan matematika dan masuk melalui jalur masuk kemitraan maka menempuh studi selama kurang dari 4 tahun dengan nilai *confidence* 96,9%.
2. Pembangkitan *2-itemset* pada algoritma *hash-based* menggunakan tabel *hash* sedangkan pada algoritma *apriori* menggunakan *minimum support*. Namun jumlah *2-itemset* yang dihasilkan tidak memiliki perbedaan yaitu sebanyak 49 kombinasi *2-itemset*.

#### 5. Daftar Pustaka

- [1] Han, J. dan M. Kamber. 2006. "*Data Mining Concepts and Techniques Second Edition*". Morgan Kauffman. San Francisco.
- [2] Turban, E, *et al.* 2005. *Decicion Support Systems and Intelligent Systems*. Andi. Yogyakarta.
- [3] Ulmer, D. 2002. *Mining an Online Auctions Data Warehouse*. Di dalam: *The Mid-Atlantic Student Workshop on Programming Languages and Systems. Proceedings of MASPLAS'02*; Pace University, 19 April 2002. <http://csis.pace.edu/csis/masplas/p8.pdf>