

PENERAPAN ASSOCIATION RULE MINING-FREQUENT ITEMSET DENGAN ALGORITME FREQUENT PATTERN GROWTH (FP- GROWTH) PADA DATASET KELULUSAN MAHASISWA S1

Fajhria Budi Ramadhanti¹⁾, Dewi Retno Sari Saputro²⁾, Purnami Widyaningsih³⁾

^{1,2,3)}Program Studi Matematika, FMIPA, Universitas Sebelas Maret Surakarta

fajhriabudi@student.uns.ac.id, dewiretnoss@staff.uns.ac.id, purnami_w@staff.uns.ac.id

Abstrak

Data mining adalah suatu proses mencari suatu informasi yang bermanfaat dalam sebuah dataset dengan metode tertentu yang salah satunya adalah metode association rule (aturan asosiasi). Algoritme FP-Growth merupakan algoritme yang dapat digunakan untuk menentukan himpunan data yang sering muncul (frequent itemset) dalam sebuah dataset. Aturan asosiasi dapat diterapkan pada data kelulusan mahasiswa menggunakan algoritme FP-Growth. Oleh karena itu, pada artikel ini diterapkan metode aturan asosiasi dengan algoritme FP-Growth pada data kelulusan mahasiswa S1 FMIPA UNS tahun 2010-2019 berdasarkan enam atribut. Keenam atribut tersebut adalah jenis kelamin, program studi, asal provinsi, IPK, jumlah SKS, dan lama studi. Adapun langkah yang digunakan pada penelitian ini yaitu cleaning data, penentuan atribut dan penerapan algoritme FP-Growth pada data kelulusan mahasiswa S1. Data tersebut diolah dengan aturan asosiasi dan menghasilkan sebuah aturan yang merupakan kumpulan dari frequent itemsets yang diurutkan dengan nilai confidence tertinggi. Hasil penelitian dan pembahasan diperoleh 32 rules dengan rentang confidence 90-97%.

Kata Kunci: Association rule; FP-Growth; frequent itemset

1. PENDAHULUAN

Perkembangan teknologi informasi digunakan hampir pada semua bidang kehidupan, salah satunya pada bidang pendidikan. Pendidikan yang semakin unggul memerlukan informasi dari data yang pernah ada, seperti halnya pada perguruan tinggi yang setiap tahunnya memperoleh *dataset* namun belum semua data tersebut diolah dan dianalisis. Dengan adanya teknik data *mining*, data tersebut akan menghasilkan informasi yang berguna untuk membuat keputusan dalam menentukan strategi pendidikan yang lebih baik. Salah satu data pada perguruan tinggi adalah data kelulusan. Data tersebut dapat digunakan untuk menemukan informasi yang sebelumnya belum ada, misal keterkaitan antar atribut yang dapat memengaruhi tingkat kelulusan.

Data *mining* didefinisikan sebagai proses penentuan sebuah hubungan, pola, atau kecenderungan dengan memeriksa dalam *dataset* yang tersimpan dalam penyimpanan dengan menggunakan teknik statistik dan matematik [6]. Salah satu metode yang digunakan dalam data *mining* adalah aturan asosiasi yang didefinisikan sebagai metode untuk menentukan aturan asosiatif antara suatu kombinasi item. Dalam aturan asosiasi terdapat beberapa algoritme, salah satunya adalah algoritme *FP-Growth*. Konsep dari algoritme *FP-Growth*

adalah konstruksi sebuah *tree* yang disebut *FP-Tree*. Penggunaan *FP-Tree* yang digunakan pada algoritme *FP-Growth* bertujuan untuk memperoleh *frequent itemset* dari sebuah *dataset*.

Berdasarkan data kelulusan mahasiswa S1 FMIPA UNS dengan tujuh program studi yang ada, rata-rata lama studi yang ditempuh oleh mahasiswa adalah 4,6 tahun. Hal ini menunjukkan bahwa terdapat mahasiswa S1 FMIPA UNS menempuh masa studi lebih dari 4 tahun yang disebabkan oleh faktor-faktor dari atribut yang memengaruhi. Oleh karena itu, pada artikel ini diterapkan algoritme *FP-Growth* pada data kelulusan S1 FMIPA UNS.

2. METODE PENELITIAN

Metode penelitian dibagi menjadi dua bagian, yang pertama yaitu bagian data penelitian dan bagian kedua adalah langkah penelitian.

A. Data Penelitian

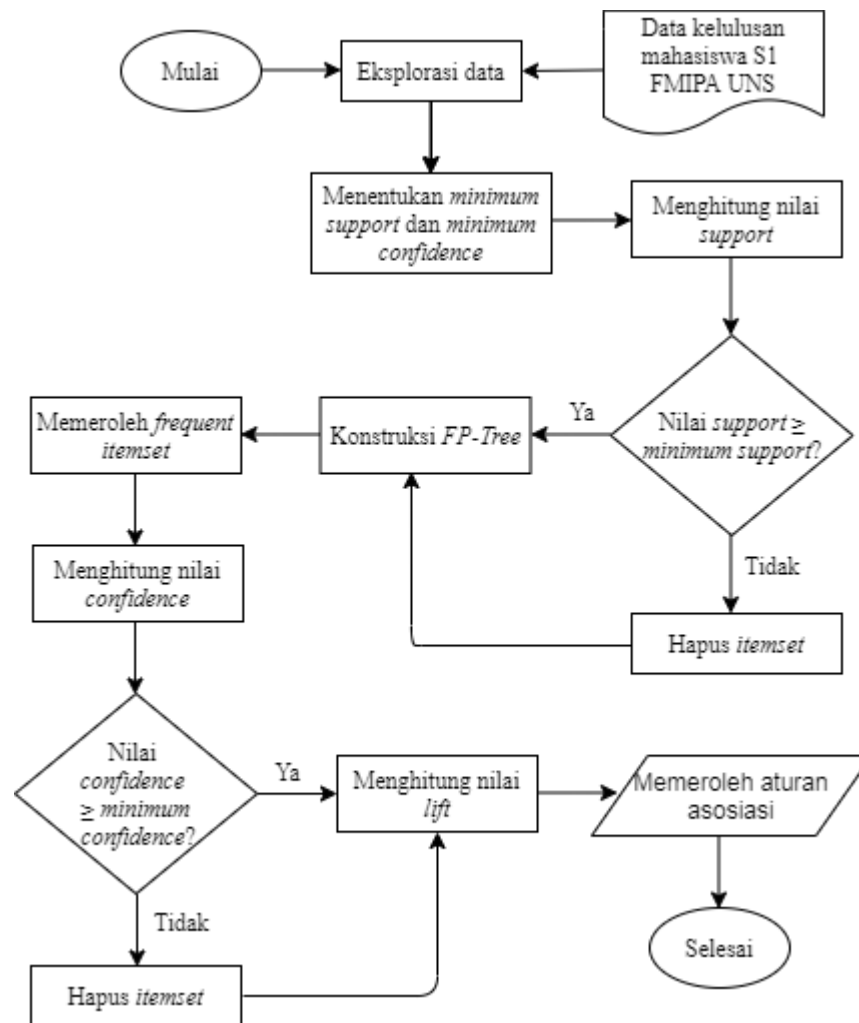
Penelitian ini menggunakan data sekunder yaitu data kelulusan S1 FMIPA UNS dari tahun 2010 hingga tahun 2019 yang diperoleh dari Biro Kemahasiswaan dan Alumni UNS. Data tersebut terdiri atas enam atribut dengan jumlah 835 data. Berikut 10 sampel data pertama yang ditunjukkan pada Tabel 1.

Tabel 1. 10 sampel data kelulusan mahasiswa S1 FMIPA UNS

No	Jenis Kelamin	Program Studi	Asal Provinsi	IPK	Jumlah SKS	Lama Studi (tahun)
1	Perempuan	Matematika	Jawa Tengah	2.62	147	7,7
2	Laki-laki	Matematika	Jawa Tengah	3.01	149	6
3	Perempuan	Matematika	Jawa Tengah	2.81	144	5,9
4	Perempuan	Matematika	Jawa Tengah	2.74	149	6
5	Perempuan	Matematika	Jawa Tengah	3.15	146	4,7
6	Perempuan	Matematika	Jawa Tengah	2.88	144	4,9
7	Perempuan	Matematika	Jawa Tengah	3.08	144	4
8	Perempuan	Matematika	Jawa Tengah	3.68	144	3,7
9	Perempuan	Fisika	Riau	2.71	144	4,9
10	Laki-laki	Fisika	Jawa Tengah	2.67	146	4,9

B. Langkah Penelitian

Berikut adalah langkah-langkah penelitian yang ditunjukkan oleh *flowchart* pada Gambar 1.



Gambar 1. Flowchart langkah penelitian

3. HASIL PENELITIAN DAN PEMBAHASAN

3.1. Data Mining

Data mining merupakan analisis peninjauan dataset untuk menemukan hubungan yang tak terduga dan meringkas dataset tersebut dengan cara yang berbeda sehingga dapat dipahami dan bermanfaat bagi pemilik data [7]. Tahapan data mining dibagi menjadi enam bagian [4], yaitu

1. *cleaning* data untuk membuang data yang tidak lengkap,
2. integrasi data atau penggabungan data dari beberapa sumber,
3. transformasi data atau data diubah menjadi bentuk yang sesuai untuk diproses pada *software*,
4. aplikasi teknik data mining menggunakan manual atau *software*,

5. evaluasi pola yang ditemukan untuk mendapatkan informasi yang berguna, dan
6. presentasi hasil dengan teknik visualisasi.

3.2. Aturan Asosiasi

Aturan asosiasi merupakan teknik data *mining* untuk menemukan aturan asosiatif antara suatu kombinasi item [4]. Konsep dari aturan asosiasi adalah mencari semua kemungkinan hubungan “jika \rightarrow maka” antar item dan memilih item/itemset yang paling sering muncul (*most-likely*) [9]. Menurut Agrawal dan Srikant [1], terdapat beberapa pernyataan tentang aturan asosiasi, yaitu misalkan $I = \{i_1, i_2, \dots, i_n\}$ merupakan himpunan dari *item* dalam transaksi, dan D merupakan himpunan transaksi dimana setiap T adalah satu himpunan *item* sehingga $T \subseteq I$. Berikut adalah definisi aturan asosiasi menurut Agrawal dan Srikant [1], dengan memisalkan X dan Y adalah himpunan *item* pada I dimana $X \neq Y$.

Definisi 3.1. Aturan Asosiasi dinyatakan dalam bentuk $X \rightarrow Y$, dengan $X \subset I, Y \subset I$, dan $X \cap Y = \emptyset$

Aturan asosiasi menggunakan dua parameter, yaitu *support* dan *confidence*. *Support* (nilai penunjang) adalah persentase kombinasi item dalam sebuah *dataset*. Sementara, *confidence* (nilai kepastian) adalah presentase kuatnya hubungan antar-*item* dalam aturan asosiasi.

3.3. Algoritme FP-Growth

Frequent Pattern Growth (FP-Growth) adalah algoritme alternatif yang digunakan untuk menentukan himpunan data yang paling sering muncul (*frequent itemset*) dalam sebuah *dataset* [8]. Algoritme *FP-Growth* menggunakan konsep konstruksi pohon yang disebut *FP-Tree* untuk memperoleh *frequent itemset* [5]. Berikut langkah untuk menentukan aturan asosiasi menggunakan algoritme *FP-Growth*.

1. Menentukan *minimum support* dan *minimum confidence*.
2. Menghitung nilai *support* pada setiap *item* menggunakan rumus berikut.

$$\text{Support}(A) = \frac{\text{Jumlah data yang memuat atribut } X}{\text{Jumlah seluruh data}}$$

Hapus itemset yang nilai *support* \leq *minimum support*

3. Megkonstruksi *FP-Tree* yang berisi akar, *label item*, *support count*, dan panah penghubung.
4. Menentukan *frequent itemset* melalui *FP-Tree* yang telah terbentuk yang merupakan lintasan tunggal.
5. Menghitung nilai *confidence* pada *frequent itemset* yang dihasilkan menggunakan rumus berikut.

$$Confidence(X \rightarrow Y) = \frac{Support(X, Y)}{Support(X)}$$

6. Menghitung nilai *lift* pada masing-masing *itemset* menggunakan rumus berikut.

$$Lift(X \rightarrow Y) = \frac{Confidence(X \rightarrow Y)}{Support(Y)}$$

Berry dan Linoff [3] menyebutkan bahwa nilai *confidence* yang terlalu tinggi dapat mengabaikan nilai *support* dari *itemset* yang muncul dalam aturan asosiasi yang telah terbentuk. Hal tersebut dapat diatasi dengan menghitung nilai *lift* pada setiap *itemset* aturan asosiasi. Jika nilai *lift* ≥ 1 maka menunjukkan terdapat manfaat pada aturan asosiasi yang terbentuk [9].

3.4. Penerapan Algoritme FP-Growth

Data kelulusan S1 FMIPA UNS dapat diterapkan dengan algoritme *FP-Growth*. Parameter yang digunakan pada algoritme *FP-Growth* yaitu *support*, *confidence*, dan *lift*. Pada penelitian ini digunakan *minimum support* sebesar 10% dan *minimum confidence* sebesar 90%. Berikut kode serta nilai *support* dalam bentuk persen yang telah memenuhi *minimum support* pada masing-masing atribut yang ditunjukkan pada Tabel 2.

Tabel 2. Kode atribut dengan nilai *support*

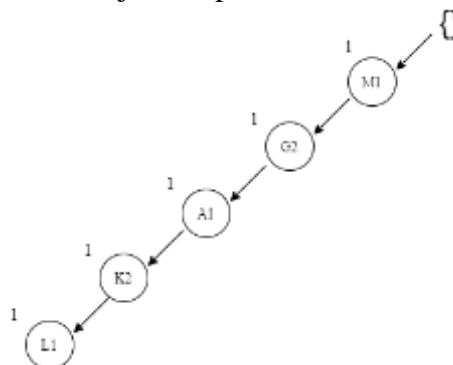
Kode Atribut	Keterangan	Support (%)
M1	Program Studi Matematika	24,7
M2	Program Studi Fisika	20,1
M3	Program Studi Kimia	18,4
M4	Program Studi Biologi	20,5
M5	Program Studi Informatika	10,7
G1	Jenis Kelamin Laki-laki	30,5
G2	Jenis Kelamin Perempuan	69,6
A1	Asal Provinsi Jawa Tengah	72,2
A2	Asal Provinsi Jawa Timur	10,4
I2	Lulus dengan Sangat Memuaskan	82,2
I3	Lulus dengan pujian (cumlaude)	12,9
K1	Mengambil SKS = 144	25,1
K2	Mengambil SKS > 144	74,9
L1	Lama Studi ≤ 4 Tahun	81,4
L2	Lama Studi > 4 Tahun	18,6

Algoritme *FP-Growth* menggunakan konsep konstruksi *FP-Tree* untuk menghasilkan kombinasi *frequent itemset*. Sebelum mengkonstruksi *FP-Tree*, disusun *itemset* yang memenuhi *minimum support* pada 835 data kelulusan mahasiswa S1 FMIPA UNS. Pada sampel data Tabel 1, data ke-1 hingga data ke-10 diberi kode sesuai pada Tabel 2 dan menghapus atribut-atribut yang kurang dari *minimum support*. Berikut Tabel 3 dengan urutan *itemset* pada 10 data pertama.

Tabel 3. 10 *Itemset*

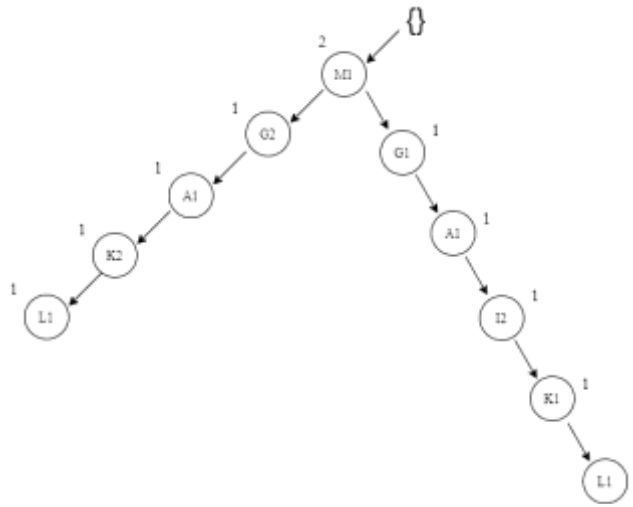
Data ke-	<i>Itemset</i>
1	M1,G2,A1,K2,L1
2	M1,G1,A1,I2,K2,L1
3	M1,G2,A1,I2,K1,L1
4	M1,G2,A1,K2,L1
5	M1,G2,A1,I2,K2,L1
6	M1,G2,A1,I2,K1,L1
7	M1,G2,A1,I2,K1,L1
8	M1,G2,A1,I3,K1,L2
9	M2,G2,K1,L1
10	M2,G1,A1,K2,L1

Setelah menyusun *itemset* pada masing-masing data, dilakukan konstruksi *FP-Tree*. Pada data ke-1 membentuk simpul M1,G2,A1,K2,L1 dengan *support count* awal bernilai satu. Konstruksi *FP-Tree* pada data ke-1 ditunjukkan pada Gambar 2.



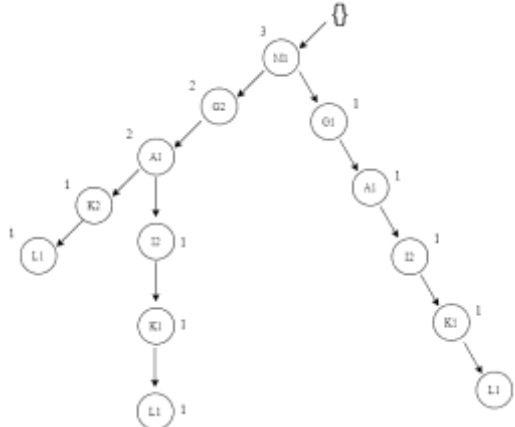
Gambar 2. Konstruksi *FP-Tree* ke-1

Selanjutnya mengkonstruksi *FP-Tree* pada data ke-2 yang membentuk simpul M1,G1,A1,I2,K2,L1. Atribut M1 diberi nilai *support count* sebesar dua karena dilewati sebanyak dua kali dan atribut lain diberi *support count* awal sebesar satu. Konstruksi *FP-Tree* pada data ke-2 ditunjukkan pada Gambar 3.



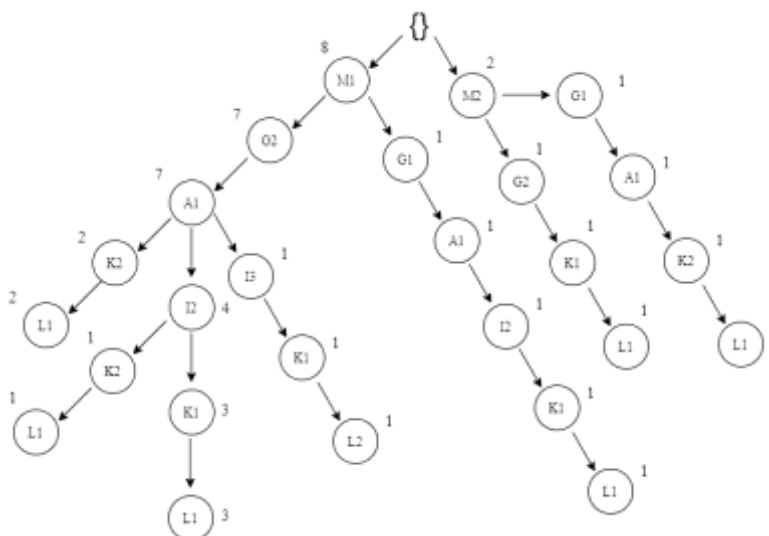
Gambar 3. Konstruksi *FP-Tree* data ke-2

Langkah selanjutnya melakukan konstruksi *FP-Tree* pada data ke-3 yang membentuk simpul M1,G2,A1,I2,K1,L1. Karena memiliki tiga *prefix* yang sama dengan lintasan data ke-1, yaitu M1,G2, dan A1, sehingga lintasan data ke-3 dapat di padatkan ke lintasan data ke-1. Selain itu, *support count* M1 ditambahkan menjadi tiga karena dilewati sebanyak tiga kali dan *support count* G2 dan A1 ditambahkan menjadi dua karena telah dilewati sebanyak dua kali. Sedangkan I2,K1, dan L1 diberi *support count* awal satu. Berikut Konstruksi *FP-Tree* pada data ke-3 yang ditunjukkan pada Gambar 4.



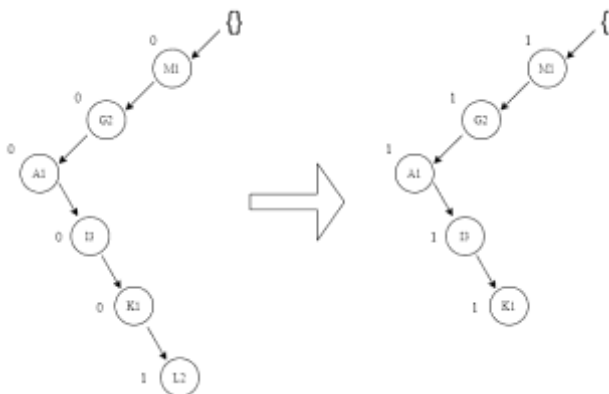
Gambar 4. Konstruksi *FP-Tree* data ke-3

Proses selanjutnya melakukan proses konstruksi *FP-Tree* untuk data ke-4 hingga data ke-10. Berikut konstruksi *FP-Tree* pada data ke-10 yang ditunjukkan pada Gambar 5.



Gambar 5. Konstruksi *FP-Tree* data ke-10

Setelah semua data di konstruksikan pada *FP-Tree*, selanjutnya mencari *frequent itemset*. Membangkitkan *FP-Tree* dilakukan dengan dari *item* yang jumlah frekuensi kemunculannya paling kecil. Diambil contoh kondisi *FP-Tree* untuk *item* L2. Proses pertama yang dilakukan yaitu mengekstrak semua lintasan yang berakhir item L2 dan selain item L2 merubah *support count* menjadi nol. Setelah itu, hapus item L2 dan memberi nilai *support count* sebesar satu. Hasilnya terdapat satu *frequent itemset* yang terbentuk yaitu M1,G2,A1,I3,K1 yang ditunjukkan pada Gambar 6.



Gambar 6. *Frequent itemset* pada kondisi *FP-Tree* item L2

Dari 835 data diperoleh *frequent itemset* sebanyak 151. Selanjutnya dari *frequent itemset* yang telah diperoleh, ditentukan nilai *confidence* dan *lift* untuk memperoleh aturan asosiasi. Aturan asosiasi yang terbentuk merupakan bentuk dari jika \rightarrow maka dengan nilai *confidence* \geq *minimum*

confidence serta nilai *lift* ≥ 1 . Setelah melakukan perhitungan *confidence* dan *lift* diperoleh 32 aturan asosiasi. Berikut 10 aturan asosiasi pertama algoritme *FP-Growth* dengan nilai *support*, *confidence*, dan *lift* yang ditunjukkan pada Tabel 4.

Tabel 4. 10 Aturan asosiasi dengan nilai parameternya

No.	Aturan asosiasi	Frekuensi	<i>Support</i> (%)	<i>Confidence</i> (%)	<i>Lift</i>
1	M3,G2,I2 \rightarrow L1	96	11,5	96,9	1,19
2	M5 \rightarrow L1	89	10,7	96,6	1,19
3	M3,I2 \rightarrow L1	131	15,7	96,2	1,18
4	M3,A1,I2 \rightarrow L1	100	11,9	96	1,18
5	M3,K2,I2 \rightarrow L1	116	13,9	95,7	1,18
6	M3,G2 \rightarrow L1	112	13,4	95,5	1,17
7	M3,A1,K2,I2 \rightarrow L1	89	10,7	95,5	1,17
8	M3 \rightarrow L1	154	18,4	95,5	1,17
9	M3,G2,K2 \rightarrow L1	101	12,1	95	1,17
10	M3,A1 \rightarrow L1	117	14,0	94,9	1,16

Pada Tabel 4 dapat diketahui bahwa aturan pertama memiliki nilai *confidence* sebesar 96,9% yang berarti *frequent itemset* tersebut merupakan kombinasi atribut yang sering muncul. Aturan pertama berbunyi “jika mahasiswa tersebut berasal dari program studi Kimia, berjenis kelamin perempuan, dan lulus dengan predikat sangat memuaskan maka mahasiswa tersebut lulus ≤ 4 tahun”. Aturan-aturan yang dihasilkan memiliki manfaat pada aturan asosiasi yang terbentuk karena memiliki nilai *Lift* ≥ 1 .

4. SIMPULAN

Berdasarkan penelitian yang telah dilakukan, dapat disimpulkan bahwa penerapan algoritme *FP-Growth* pada data kelulusan mahasiswa S1 FMIPA UNS tahun 2010 hingga tahun 2019 menghasilkan 32 aturan asosiasi dengan rentang *confidence* 90%-97%.

5. DAFTAR PUSTAKA

- Agrawal, R., Imielinski T. & Swami A. (1993). *Mining Association Rules Between Sets of Items in Large database*. Proceeding of the 1993 ACM International Conference on Management of Data. 207-216.
- BANPT. (2007). *Akreditasi Institusi Perguruan Tinggi*. Departemen Pendidikan Nasional, Jakarta.
- Berry, M.J.A. & Linoff, G.S. (2004). *Data Mining Techniques for Marketing, Sales, and Customer Relationship Management Second Edition*. Wiley Publishing Inc., Hoboken.

- Demir, N. (2004). *Association Rules Mining in Plastic Pipe Sector*. International Journal of Scientific Engineering and Applied Science, **1**(4).
- Han J. & Kamber, M. (2006). *Data Mining : Concepts and Techniques Third Edition*. Morgan Kaufmann, USA.
- Han J., Pei J., & Yin Y. (2001). *Mining Frequent Itemset Without Candidate Generation*. International Conference on Management of Data. **29**, 1-12.
- Kusrini, & Luthfi, E.T. (2009). *Algoritma Data Mining*. Penerbit Andi, Yogyakarta.
- Larose, D.T. (2005). *Discovery Knowledge in Data: An Introduction to Data Mining*. John Wiley and Sons, Canada.
- Samuel, D. (2008). *Penerapan Struktur FP-Tree dan Algoritma FP-Growth dalam Opimasi Penentuan Frequent Itemset*. Institut Teknologi Bandung, Bandung.
- Santosa, B. (2007). *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Graha Ilmu, Yogyakarta.
- Syafiqoh, N.N. & Candra F. (2019). *Penerapan Algoritma FP-Growth Untuk Mengetahui Variabel yang Memengaruhi Tingkat Kelulusan Mahasiswa FT UR*. Computation Technology and its Application. **1**(1).