

IMPLEMENTASI ALGORITMA CT-PRO UNTUK MENEMUKAN POLA PADA DATA SISWA SMA (STUDI KASUS: MADRASAH ALIYAH NEGERI (MAN) KARANGANON KLATEN)

Rohmania Putri Nurlaili¹⁾, Sari Widya Sihwi²⁾ Meiyanto Eko Sulisty³⁾

^{1,2,3}Jurusan Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Sebelas Maret

Jalan Ir. Sutami No 36 A, Surakarta, 57126

Telp : (0271) 646994, Fax : (0271) 646655

E-mail : mant.manzz@gmail.com¹⁾, sari.widya.sihwi@gmail.com²⁾, mekosulistyo@uns.ac.id³⁾

Abstrak

Data yang dimiliki oleh setiap sekolah tentu terus bertambah setiap tahunnya seiring dengan penambahan siswa. Data-data tersebut berupa data pribadi siswa yang meliputi jenis kelamin, nilai UN SMP, jarak rumah, penghasilan orang tua, pekerjaan orang tua siswa dan nilai SMA. Ternyata dari data berlimpah tersebut dapat dimanfaatkan untuk dilihat keterkaitannya pada masing-masing data sehingga dapat dimanfaatkan oleh suatu sekolah dalam pengambilan keputusan salah satunya dalam meningkatkan nilai ujian nasional. Untuk melihat keterkaitan tersebut salah satunya digunakan teknik data mining dengan metode asosiasi. Dalam penelitian ini mencari keterkaitan antara data pribadi dengan nilai UN SMA dengan studi kasus di MAN Karanganyar Klaten dengan menggunakan algoritma CT-Pro. Hasil dari penelitian ini adalah siswa laki-laki cenderung memberikan nilai UN SMA antara 60-70, siswa perempuan dengan penghasilan orang tuanya kurang dari 500.000 memberikan nilai UN SMA kurang dari 50, tetapi siswa perempuan dengan pekerjaan orang tuanya sebagai tani/buruh memberikan nilai UN SMA lebih tinggi yaitu antara 70-80, dan nilai UN SMP tidak mempengaruhi nilai UN SMA.

Kata kunci: Data mining, Asosiasi, CT-Pro.

Abstract

School data of certain school would continue to grow each year as the number of students. Those data are in the form of students' personal data which include gender, Junior high school examination score, the distance of house, parental income, occupation of parents and high school score. It turns out that the abundant data can be used for viewing relation on each of the data thus it can be used by certain school as decision making on increasing the score of national exam. To see these linkages one of the way is the use of data mining techniques through association method. This study is looking for a link between personal data with National Examination score through case study in the MAN Karanganyar Klaten by using the CT-Pro algorithm. The results of this study are male students tend to give high school national examination score between 60-70, female students with parents earning less than 500,000 got less than 50 score of this examination. On the other hand female students whose parents work as farmers/ laborers give a higher score on the final examination between 70-80, and score of Junior High School National Examination does not affect the score of Senior High School Final Examination.

Keywords: Data mining, Association, CT-Pro.

1. PENDAHULUAN

Setiap sekolah memiliki data siswa baik tentang profilnya maupun hasil pembelajaran di sekolah. Contoh dari data-data tersebut antara lain meliputi jenis kelamin, alamat, jarak rumah, penghasilan orang tua, pekerjaan orang tua, nilai ujian nasional SMP (Sekolah Menengah Pertama) dan data dari nilai-nilai yang diperoleh siswa saat mengikuti proses pembelajaran di sekolah meliputi nilai ulangan harian, nilai raport dan nilai ujian nasional. Data-data tersebut setiap tahunnya bertambah dan terus berlimpah dengan semakin banyaknya siswa. Data yang sangat banyak hanya akan menjadi sia-sia jika tidak dimanfaatkan. Ternyata dari data yang berlimpah itu dapat dimanfaatkan untuk dilihat keterkaitannya pada masing-masing data. Misalnya data siswa antara jarak rumah dengan nilai ujian nasional SMA (Sekolah Menengah Atas), kita dapat mengetahui seberapa besar keterkaitan hubungan mereka. Dengan kita mengetahui keterkaitan ini dapat dimanfaatkan oleh suatu sekolah dalam pengambilan keputusan, salah satunya dalam meningkatkan nilai ujian nasional. Untuk melihat keterkaitan tersebut, salah satunya digunakan teknik data mining. Data mining merupakan analisis dari peninjauan kumpulan

data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan cara yang berbeda dengan sebelumnya, yang dapat dipahami dan bermanfaat bagi pemilik data. Dalam data mining, teknik yang digunakan untuk melihat keterkaitan antar data menggunakan asosiasi. Asosiasi digunakan untuk mengenali kelakuan dari kejadian-kejadian khusus atau proses dimana link asosiasi muncul pada setiap kejadian. Penting tidaknya suatu aturan asosiatif dapat diketahui dengan dua parameter yaitu *support* dan *confidence*. *Support* (nilai penunjang) adalah presentase kombinasi item tersebut dalam database, sedangkan *confidence* (nilai kepastian) adalah kuatnya hubungan antar item dalam aturan asosiasi [1]. Ada beberapa algoritma untuk aturan asosiasi, diantaranya adalah algoritma Apriori, algoritma Frequent Pattern-Growth (FP-Growth) dan algoritma Compact Tree-Apriori (CT-PRO). Pada penelitian ini, algoritma yang digunakan adalah algoritma CT-PRO, karena CT-PRO menyajikan kompleksitas waktu dari waktu yang terbaik dan waktu yang terburuk, perbandingan CT-PRO dengan algoritma yang lainnya seperti Apriori, FP-Growth dan Oportune Project (OP) menunjukkan bahwa algoritma CT-PRO melebihi algoritma lain pada semua level *support* pada dataset yang sering muncul dan juga lebih baik pada dataset yang jarang [2].

2. LANDASAN TEORI

2.1 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 [3]. Tujuan data mining adalah mendapatkan hubungan atau pola yang mungkin memberikan indikasi yang bermanfaat [1].

Data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu [4]:

1. Deskripsi, digunakan untuk menggambarkan pola dan kecenderungan yang terdapat dalam data.
2. Klasifikasi, contoh kasusnya adalah mendiagnosis penyakit seorang pasien untuk mendapatkan termasuk kategori penyakit apa.
3. Estimasi, hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik daripada ke arah kategori.
4. Prediksi, hampir sama dengan klasifikasi dan estimasi kecuali bahwa dalam prediksi nilai dari hasil akan ada di masa mendatang.
5. Pengklusteran, merupakan pengelompokan record, pengamatan atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan.
6. Asosiasi, yaitu menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja. Contoh asosiasi dalam penelitian adalah menemukan barang dalam supermarket yang dibeli secara bersamaan dan barang yang tidak pernah dibeli secara bersamaan.

2.2 Asosiasi

Analisis asosiasi adalah teknik *data mining* untuk menemukan aturan asosiatif antara suatu kombinasi item [1]. Analisis asosiasi berguna untuk menemukan hubungan penting yang tersembunyi diantara set data yang sangat besar [5]. Kekuatan asosiasi dapat diukur dengan *support* dan *confidence*. *Support* (nilai penunjang) yaitu prosentase kombinasi item tersebut dalam database dan *confidence* (nilai kepastian) yaitu kuatnya hubungan antar *item* dalam aturan asosiatif.

Metodologi asosiasi terbagi menjadi dua tahap [1], yaitu :

1. Analisis pola frekuensi tinggi

Tahap ini mencari kombinasi item yang memenuhi syarat minimum dari nilai *support* dalam database berdasarkan persamaan (1)

$$support(A) = \frac{jml \text{ transaksi mengandung } ngA}{total \text{ transaksi}} \quad (1)$$

2. Pembentukan aturan asosiasi

Setelah semua frekuensi tinggi ditemukan, dicari aturan asosiasi yang memenuhi syarat minimum untuk *confidence* dengan menghitung *confidence* aturan asosiatif $A \rightarrow B$ menggunakan persamaan (2)

$$confidence = P(B|A) = \frac{\sum \text{transaksi mengandung } ngA \& B}{\sum \text{transaksi mengandung } ngA} \quad (2)$$

2.3 Algoritma CT-PRO

Algoritma ini merupakan pengembangan dari algoritma FP-GROWTH dengan melakukan modifikasi pada *tree* yang digunakan. Algoritma ini menggunakan struktur *Compressed FP-Tree* (CFP-Tree) dimana informasi dari sebuah FP-Tree diringkas dengan struktur yang lebih kecil atau ringan, sehingga baik pembentukan *tree* maupun *frequent itemset mining* yang dilakukan menjadi lebih cepat.

Langkah-langkah algoritma CT-PRO adalah sebagai berikut:

1. Menemukan *item-item* yang *frequent*
 - Data-data yang telah dikumpulkan, diseleksi dan pilih data yang relevan (data yang lengkap).
 - Data-data yang ada, kemudian dilakukan transformasi data.
 - Kemudian masing masing data diseleksi berdasarkan minimum support yang telah ditentukan, kemudian didapat *Item Frequent Table*.
 - Masing-masing *item* dihitung frekuensi kemunculannya sehingga dihasilkan *global item table*.
2. Membuat CFP-Tree
 - Setelah ditemukan *item-item* yang *frequent* kemudian dilakukan pembangunan CFP-Tree. *Frequent item* yang ada diurutkan sesuai *global item* dari nilai yang terbesar ke terkecil.
3. Melakukan penggalian *frequent patterns*
 - *Global item table* diurutkan dari item terkecil ke terbesar, karena algoritma CT-PRO bekerja dengan melakukan *bottom-up mining*.
 - Masing-masing *global item table* di lakukan pencarian *node* yang berkaitan dengan *item* tersebut, dan kemudian diseleksi berdasarkan *minimum support*nya maka diperoleh *Local Item Table*.
 - Dari *local item table* dibuat CFP-Tree berdasarkan *minimum support* yang telah ditentukan.
 - Kemudian *item* yang memenuhi *minimal support* dilakukan *mining* dengan rumus *confidence*.

3. METODOLOGI

3.1 Pengumpulan Data

Tahap ini ditempuh guna mendapatkan informasi data yang berkaitan dengan obyek yang dikaji. Data yang digunakan dalam penelitian ini menggunakan data dari siswa MAN Karangnom Klaten sebanyak 700 data mulai dari siswa angkatan 2005 sampai angkatan 2012, dan masing-masing data yang diambil meliputi atribut nama siswa, jenis kelamin, jarak rumah, penghasilan orang tua, pekerjaan orang tua, nilai ujian nasional SMP dan nilai ujian nasional SMA.

3.2 Implementasi (Penerapan Metode)

Kode program dalam pembuatan aplikasi merupakan penerapan dari metode yang digunakan yaitu asosiasi dengan algoritma CT-PRO. Bahasa pemrograman yang digunakan adalah PHP (*Hypertext Preprocessor*). Media penyimpanan data menggunakan *MySQL database*. Berikut ini alur proses yang dilakukan:

- 1) *Preprocessing*
 - *Data cleaning*: proses menghilangkan *noise* dan data yang tidak konsisten atau tidak relevan.
 - *Transformasi data*: data-data yang telah dibersihkan kemudian diubah menjadi satu format yang sama untuk mempermudah proses *mining* dan analisis pola data. *Transformasi* data dilakukan dengan mengubah parameter data-data pribadi menjadi sebuah kode *item*.
- 2) Menemukan *item-item* yang *frequent*

Membuat *item frequent table* yaitu dengan menghitung kemunculan tiap *item*. Kemudian dilakukan seleksi data dengan menentukan *minimum support* terlebih dahulu. Kemudian data-data yang memenuhi *minimum support* diurutkan berdasarkan jumlah kemunculan masing-masing *item* dengan memberikan *index* berdasarkan count dari terbesar ke terkecil, sehingga diperoleh *global item table*.
- 3) Mapping dan Membuat CFP-Tree

Setelah ditemukan *item-item* yang *frequent* kemudian dilakukan pembangunan CFP-Tree. Untuk membangun CFP-Tree sebelumnya, data hasil *transformasi* dimapping berdasarkan *index* pada *global item table*. Selanjutnya adalah pembangunan CFP-Tree. Langkah yang pertama yaitu pembacaan transaksi ID yang pertama akan membuat sebuah simpul sehingga terbentuk lintasan transaksi. *Support count* dari setiap simpul bernilai satu. Lalu pembacaan transaksi kedua, jika terjadi prefik transaksi sama dengan transaksi pertama, maka lintasan transaksi kedua dapat ditimpakan pada transaksi yang sama sambil menambah support count dari transaksi yang sama tersebut dan selanjutnya membuat lintasan baru sesuai dengan transaksi kedua, dan seterusnya untuk transaksi berikutnya. Proses ini dilanjutkan sampai CFP-Tree berhasil dibangun berdasarkan tabel data transaksi yang diberikan.
- 4) Tahap *Mining*

Algoritma CT-PRO bekerja dengan *bottom-up mining* sehingga *global item table* diurutkan mulai dari *item* berfrekuensi terkecil hingga terbesar. Masing-masing *global item table* di lakukan pencarian *node* yang berkaitan dengan *item* tersebut maka diperoleh *local item table*.

Selanjutnya dari *local item table* kemudian dibangun local CFP-Tree berdasarkan jumlah *minimum support* yang telah ditentukan. Dari local CFP-Tree, kemudian dilakukan *mining* dengan rumus penghitungan *Confidence*.

3.3 Analisis hasil

Pada tahap ini dilakukan analisis terhadap hasil penelitian dengan mengamati hasil/output dari perhitungan yang dilakukan sistem dengan menggunakan *minimum support count* 3% dan *minimum confidence* 40%.

4. PEMBAHASAN

Dalam penelitian ini, ada tahapan utama yang harus dilewati adalah transformasi data. Untuk mempermudah proses *mining*, data harus di transformasi menjadi satu format yang sama ke dalam kode-kode yang sesuai pada tabel 1 berikut.

Setiap transaksi mengandung 5 atribut yaitu jenis kelamin, nilai SMP (Sekolah Menengah Pertama), penghasilan orang tua, jarak rumah dan nilai SMA (Sekolah Menengah Atas). Masing-masing atribut data dibagi lagi menjadi beberapa jenis *item*. Untuk mempermudah penghitungan, maka dilakukan kodifikasi data seperti pada tabel 1.

Tabel 1. Tabel Kodifikasi

ITEM	KODE	ITEM	KODE
Jenis Kelamin Laki-laki	1	Pekerjaan orang tua swasta	14
Jenis Kelamin Perempuan	2	Pekerjaan orang tua pedagang	15
Nilai ujian nasional SMP $\geq 9,00 - 10,00$	3	Pekerjaan orang tua PNS	16
Nilai ujian nasional SMP $\geq 8,00$ sampai $< 9,00$	4	Pekerjaan orang tua lainnya	17
Nilai ujian nasional SMP $\geq 7,00$ sampai $< 8,00$	5	Jarak rumah < 7 KM	18
Nilai ujian nasional SMP $\geq 6,00$ sampai $< 7,00$	6	Jarak rumah ≥ 7 KM sampai < 15 KM	19
Nilai ujian nasional SMP $\geq 5,00$ sampai $< 6,00$	7	Jarak rumah > 15 KM	20
Nilai ujian nasional SMP $< 5,00$	8	Nilai ujian nasional SMA $\geq 9,00 - 10,00$	21
Pendapatan orang tua < 500.000	9	Nilai ujian nasional SMA $\geq 8,00$ sampai $< 9,00$	22
Pendapatan orang tua antara 500.000-1.000.000	10	Nilai ujian nasional SMA $\geq 7,00$ sampai $< 8,00$	23
Pendapatan orang tua antara 1.000.000-3.000.000	11	Nilai ujian nasional SMA $\geq 6,00$ sampai $< 7,00$	24
Pendapatan orang tua $> 3.000.000$	12	Nilai ujian nasional SMA $\geq 5,00$ sampai $< 6,00$	25
Pekerjaan orang tua tani/buruh	13	Nilai ujian nasional SMA $< 5,00$	26

Berikut diberikan contoh sebagai kasus menggunakan data survey 5 transaksi dengan *minimum support count* 20% dan *minimum confidence* 50%.

Tabel 2. Data Transaksi

Id siswa	Jenis kelamin	Nilai SMP	Penghasilan ortu	Pekerjaan ortu	Jarak rumah (KM)	Nilai SMA
1	P	9,92	< 500.000	Tani/buruh	8,5	6,72
2	L	8,82	< 500.000	Tani/buruh	8	7,22
3	P	8,22	1.000.000 – 3.000.000	swasta	9	6,94
4	P	8,06	< 500.000	Tani/buruh	10	6,8
5	L	7,89	< 500.000	Tani/buruh	5	7,16

Kemudian data-data pada tabel 2 dilakukan transformasi sesuai dengan kodifikasi pada tabel 1, dan berikut tabel 3 adalah hasil transformasi data.

Tabel 3. Data hasil Transformasi

No id	A	B	C	D	E	F
1	2	3	9	13	19	24
2	1	4	9	13	19	23
3	2	4	11	14	19	24
4	2	4	9	13	19	24
5	1	5	9	13	18	23

Keterangan :

A : Jenis Kelamin

B : Nilai Ujian SMP

C : Penghasilan orang tua

D : Pekerjaan orang tua

E : Jarak Rumah

F : Nilai Ujian SMA

Setelah data ditransformasi, selanjutnya tahap pengolahan data sesuai dengan algoritma CT-Pro :

a) menghitung frekuensi kemunculan tiap *item* sebagaimana tersaji pada tabel 4.

Tabel 4. Frequent Item Table

Kode Item	Frekuensi item	Kode Item	Frekuensi item
1	2	13	4
2	3	14	1
3	1	18	1
4	3	19	4
5	1	23	2
9	4	24	3
11	1		

b) membangun *global item table*

Masing-masing *item* diseleksi berdasarkan *minimum support count* yang telah ditentukan yaitu 20%. Kemudian data-data yang *frequent* diurutkan dari yang memiliki frekuensi terbesar ke terkecil sehingga terbentuk *global item table*. Tabel 5 adalah tabel data yang telah lolos seleksi *minimum support count* dan selanjutnya id pada tabel 5 disebut dengan id global item.

Tabel 5. berikut tabel *item yang frequent*.

Id global item	Kode Item	<i>Frequent Item</i>	Id global item	Kode Item	<i>Frequent Item</i>
1	9	4	5	4	3
2	13	4	6	24	3
3	19	4	7	1	2
4	2	3	8	23	2

c) *Mapping data*

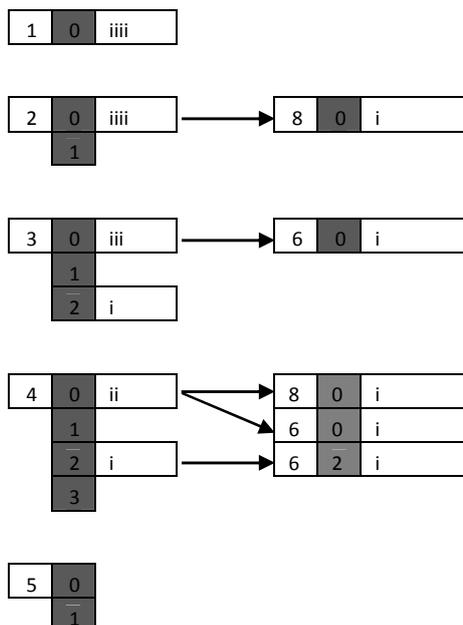
Mapping yaitu memetakan data transaksi dari tabel 3 terhadap id global *item* pada tabel 5. Kode *item* yang tidak lolos seleksi *minimum support* dihilangkan. Pada tabel 6 berikut adalah hasil *mapping data* transaksi.

Tabel 6. Tabel Mapping

A	B	C	D	E	F
3	-	1	2	3	6
6	4	1	2	3	8
3	4	-		3	6
3	4	1	2	3	6
6	-	1	2		8

d) Membuat CFP-Tree

Gambar berikut merupakan hasil pembentuka CFP-Tree dari data survey setelah melalui tahap *mapping*.



2
3
4

Gambar 1. Hasil pembentukan CFP-Tree setelah pembacaan sampai transaksi ke-5

Pada gambar 1, kolom pertama adalah kode id *global item*, kolom kedua adalah index, kolom ketiga adalah frekuensi kemunculan. Gambar 1 menunjukkan hasil CFP-Tree dari kelima transaksi. Setiap simpul pada *Tree* mengandung nama sebuah *item* dan frekuensi kemunculannya dalam tiap lintasan transaksi. CFP-Tree dibentuk dengan cara berikut :

- Pembacaan transaksi pertama pada tabel 5, sehingga diperoleh lintasan 1 → 2 → 3 → 6
- Pembacaan transaksi kedua yaitu 1 → 2 → 3 → 4 → 8, karena memiliki awalan yang sama maka transaksi kedua dapat ditimpakan di index ke 0 pada node 1, 2 dan 3, karena 1 merupakan root. Sambil menambahkan count untuk setiap node dan selanjutnya untuk id 8 membuat lintasan baru sesuai dengan pembacaan transaksi kedua.
- Pembacaan transaksi terus dilanjutkan sesuai dengan ketentuan sebelumnya. Apabila transaksi diawali dengan node baru yang bukan root, maka node berada pada index baru. Tahap ini terus berlangsung sampai pembacaan semua transaksi dan CFP-Tree berhasil dibangun.

e) Penerapan *Mining*

Tujuan dalam penelitian ini adalah menemukan pola asosiasi antara data nilai ujian nasional SMA dengan atribut data siswa yang lainnya. Oleh karena itu, nilai confidence yang dihitung hanya pada kombinasi node-node yang mengndung *item* nilai SMA.

5. ANALISA HASIL

Setelah dilakukan perhitungan terhadap data-data siswa MAN Karangnom Klaten, dengan jumlah data survey 700 data dan nilai *minimum support countrnya* 3% dan *minimum confidencenya* 40%, maka diperoleh hasil yang tersaji pada tabel 7 berikut.

Tabel 7. Hasil pengujian

No.	Kombinasi Item	Nilai Confidence (%)	No.	Kombinasi Item	Nilai Confidence (%)
1	19 23	52,73	16	19 13 23	60,00
2	19 24	44,21	17	19 13 24	53,65
3	19 25	44,05	18	19 13 25	56,83
4	19 26	49,15	19	19 13 26	51,69
5	18 24	45,06	20	2 23	71,82
6	18 25	45,37	21	2 24	59,66
7	18 26	41,53	22	2 25	63,44
8	9 23	62,73	23	2 26	72,03
9	9 24	63,95	24	7 23	40,91
10	9 25	61,67	25	7 26	61,86
11	9 26	55,93	26	2 13 23	47,27
12	13 23	69,09	27	2 9 25	40,53
13	13 24	58,37	28	2 7 26	48,31
14	13 25	59,03	29	1 24	40,34
15	13 26	53,39	30	6 25	44,93

Tujuan pada penelitian ini adalah menemukan pola asosiasi antara data nilai ujian nasional SMA dengan atribut data siswa yang lainnya. Dari hasil penelitian pada tabel 7 diatas, nilai ujian nasional SMA ada pada kode item 23, 24, 25 dan 26. Sehingga diperoleh hasil kombinasi dari kode-kode tersebut sebagai berikut :

- Dari tabel 7, terlihat item dengan kode 19, 9, 13, 2 dan 9|13 mendominasi pada semua range nilai SMA, begitu juga dengan kombinasi dengan 18 hampir pada semua range nilai SMA, artinya item-item dengan kode 19, 9, 13, 2 dan 18 tidak mempengaruhi terhadap besar kecilnya nilai ujian SMA siswa.
- Kombinasi 2 | 13 | 23 artinya siswa yang berjenis kelamin perempuan dan pekerjaan orang tuanya sebagai petani/buruh, akan menghasilkan nilai SMA antara 70-80.
- Kombinasi 1 | 24 artinya siswa laki-laki akan menghasilkan nilai SMA antara 60-70 terlihat pada tabel 7 dengan nilai confidence sebesar 40,34%.

- Kombinasi 2 | 9 | 26 artinya siswa perempuan dan orang tuanya yang berpenghasilan <500.000 akan menghasilkan nilai SMA kurang dari 50.
- Nilai-nilai ujian nasional SMA yang dihasilkan cenderung lebih rendah dari nilai ujian nasional SMP, terlihat pada tabel 7 nomer 25 dan 30. Siswa yang memiliki nilai UN SMP antara 60-70 hasil UN SMA antara 50-60 begitu juga siswa dengan nilai UN SMP antara 50-60 hasil UN SMA kurang dari 50.

6. SIMPULAN DAN SARAN

6.1 Simpulan

- Nilai ujian nasional SMA yang rendah, salah satunya disebabkan oleh penghasilan dalam keluarga yang rendah pula. Namun, tidak berlaku pada siswa yang pekerjaan orang tuanya sebagai petani/buruh, justru malah menghasilkan nilai yang lebih bagus.
- Siswa laki-laki akan lebih memberikan nilai akhir kisaran 60-70.
- Nilai ujian nasional SMP yang bagus belum tentu akan memberikan nilai ujian nasional SMA yang bagus, artinya nilai ujian nasional SMP tidak memberikan pengaruh terhadap nilai ujian nasional SMA.

6.2 Saran

Untuk penelitian selanjutnya digunakan data yang lebih banyak lagi agar mendapatkan hasil yang lebih akurat dan kesimpulan yang lebih memuaskan. Penelitian sebaiknya dilakukan tidak hanya pada satu sekolah, namun ke berbagai sekolah dengan harapan aplikasi yang dihasilkan dapat digunakan diberbagai sekolah secara umum.

7. DAFTAR RUJUKAN

- [1] Kusriani, E. T. (2009). *Algoritma Data mining*. Yogyakarta: Andi.
- [2] Suchahyo, Y. G., & Gopalan, R. P. (2004). *CT-PRO: A Bottom-Up Non Recursive Frequent Itemset Mining Algorithm Using Compressed FP-Tree Data Structure*. Department of Computing, Curtin University of Technology.
- [3] Turban, E. (2005). *Decision Support System and Intelligent Systems*. Yogyakarta: Andi Offset.
- [4] Larose, D. T. (2005). *Discovering Knowledge in Data : An Introduction to Data mining*. John Willey and Sons, Inc.