

SISTEM REKOMENDASI PEMBELAJARAN PADA *E-LEARNING* MENGGUNAKAN ALGORITMA *CT-PRO*

Khoirul Ummah¹

Program Studi Teknik Informatika, Universitas Trunojoyo Madura

Masuk: 4 September 2014, revisi masuk : 11 Januari 2015, diterima: 26 Januari 2015

ABSTRACT

The progress of information technology in education, especially the use of e-learning in education institutions developed rapidly, SMA 2 Pamekasan was an example of schools that could use elearning. However, the teachers only used to take the value of the test, assignment by the teacher and to access course materials. Therefore, in this research use data that are considered essential to assist teachers measured students's progress and help students answer the difficult questions. Subjects that used in this e-learning was chemical. Chemical was one subject that was considered difficult by most students. So, this application would help students solve a problems. The algorithm that used in this application was a ct-pro, it was algorithm of association rules to find combinations of data's relations. The data were the wrong answers of 150 students. Questions consist of 400 numbers and 20 learning materials. In this research, the data was divided into 6, the data from 5 students, 30 students, 60 studenst, 90 students, 120 students and 150 students. The results of this reseacrh was if there were more data, the combination formed itemset also be more and more, and the processing time will also be longer. However, if the greater the minimum support was entered, then the combination was formed will be less.

Keywords: Association Rules, CT-Pro, minimum support.

INTISARI

Perkembangan teknologi informasi dalam dunia pendidikan khususnya penggunaan *e-learning* di lembaga pendidikan berkembang cukup pesat. SMAN 2 Pamekasan adalah contoh sekolah yang menggunakan *elearning*. Namun, sampai saat ini penggunaan *e-learning* hanya digunakan untuk melakukan ujian online, pemberian nilai oleh guru dan mengakses materi pelajaran. Oleh karenanya, dalam penelitian ini menggunakan data-data yang telah ada untuk membantu para pengajar dalam mengukur perkembangan dan membantu siswa dalam mengerjakan soal yang dianggap sulit. Mata pelajaran yang digunakan dalam *e-learning* ini adalah Kimia. Kimia merupakan salah satu mata pelajaran yang dianggap sulit oleh sebagian siswa, karenanya dengan aplikasi ini akan sedikit membantu siswa dalam mengerjakan soal dan membantu guru dalam melihat perkembangan siswa. Algoritma yang digunakan dalam aplikasi ini adalah *CT-PRO*, yaitu algoritma dalam *association rules* untuk mencari hubungan kombinasi dari setiap data. Data yang digunakan adalah jawaban salah dari 150 siswa. Soal terdiri dari 400 nomor dari 20 materi yang ada. Dalam penelitian ini, data dipecah menjadi 6, yaitu data dari 5 siswa, 30 siswa, 60, 90, 120, dan 150 siswa. Hasil penelitian dari pembagian data tersebut, diketahui bahwa semakin banyak data, maka kombinasi *itemset* yang terbentuk juga akan semakin banyak, dan waktu proses juga akan semakin lama. Namun, semakin besar *minimum support* yang diinputkan, maka kombinasi yang terbentuk akan semakin sedikit.

Kata Kunci: Association Rules, CT-Pro, minimum support

PENDAHULUAN

Kemajuan ilmu pengetahuan dan teknologi, khususnya teknologi informasi banyak membawa dampak positif. Salah satunya bagi kemajuan dunia pendidikan. Pendidikan formal, informal, dan non formal dapat menikmati fasilitas berupa proses belajar mengajar berbasis teknologi informasi yang memberikan banyak kemudahan dan manfaat bagi tiap-tiap individu. Konsep yang kemudian terkenal dengan sebutan *e-learning* ini membawa pengaruh terjadinya proses transformasi pendidikan konvensional ke dalam bentuk digital, baik secara isi (*contents*) dan sistemnya. Adanya kemudahan dan manfaat dari sistem pembelajaran ini, mendorong berbagai lembaga pendidikan memanfaatkan sistem *e-learning* untuk meningkatkan efektivitas dan fleksibilitas pembelajaran. Pengajar dapat memberikan ujian atau latihan secara *online* kepada pengguna sistem yang dalam hal ini adalah pelajar atau siswa. Ujian secara *online* dapat mempermudah penilaian yang dilakukan pengajar terhadap siswa dan melihat sejauh mana perkembangan peserta didik.

SMA Negeri 2 Pamekasan merupakan salah satu lembaga pendidikan yang ada di Kabupaten Pamekasan, SMA Negeri ini sudah mulai menggunakan *e-learning* sejak beberapa tahun lalu. Namun, sampai saat ini penggunaan *e-learning* hanya digunakan untuk melakukan ujian *online*, pemberian nilai oleh guru dan mengakses materi pelajaran. Padahal, seiring penggunaan *e-learning* tersebut akan mendapatkan informasi lebih tentang perkembangan para siswa apabila memanfaatkan data-data yang telah ada dan memberikan sedikit bantuan bagi para siswa dalam pengerjaan soal ujian.

Pada penelitian ini dibangun aplikasi untuk mencari informasi tentang perkembangan siswa dengan melihat jawaban-jawaban siswa yang salah dalam suatu ujian. Dengan data tersebut akan diperoleh suatu kombinasi keterkaitan materi yang dirasa menyulitkan para siswa, dan akan dijadikan sarana untuk membantu para siswa dalam mengerjakan soal yang

akan berupa *Hints* atau rekomendasi pembelajaran. Aplikasi ini menggunakan algoritma *ct-pro*, yakni sebuah algoritma dalam *Association Rule* untuk menemukan relasi antar data.

Dalam penelitian yang lain, Dominguez bersama dengan Kalina dan James C. Curran telah mencoba membangun *hints* serupa dalam sebuah ujian. Proses *Association Rules* yang dipakai menggunakan Algoritma Apriori (Dominguez, 2010). Penelitian selanjutnya yang dilakukan Sucahyo dan Raj juga pernah membandingkan kinerja Algoritma *FP-Growth* dan *ct-pro*. Hasil penelitian menjelaskan bahwa bahwa algoritma *ct-pro* merupakan pengembangan dari algoritma *fp-growth*. *Ct-pro* menggunakan *compressed fp-tree (CFP TREE)* yang jumlah nodenya bisa setengah kali dari *fp-tree* dan menjelaskan bahwa *ct-pro* memiliki performa yang lebih baik jika dibandingkan dengan algoritma *fp-growth* dan apriori (Giri, 2005). Pada penelitian lainnya juga telah diimplementasikan algoritma *Association rules* yaitu Apriori, *FP-Growth*, *ct-pro*, dan apriori Cristian Borgelt sebagai bagian dari pengembangan *data mining workbench*. Dari hasil pengujian pada *dataset chess*, *CT-Pro* paling cepat dibandingkan dengan algoritma yang lain. Keunggulan *Ct-pro* adalah dari penggunaan memori yang lebih hemat dan digunakannya struktur data *CFP-Tree* yang memungkinkan proses pencarian *frequent itemset* menjadi lebih cepat (Ruldeviyani, 2008).

Berdasarkan ketiga penelitian tersebut, dapat dilihat bahwa penelitian serupa telah pernah dilakukan sebelumnya. Namun, juga telah diketahui telah ada algoritma yang lebih baik lagi. Oleh karena itu dalam penelitian ini akan menggunakan algoritma *CT-Pro* yang lebih baik dari algoritma Apriori dan *FP-Growth* untuk mencari kombinasi berdasarkan pola jawaban yang salah. Hasil akhir dari penelitian ini diharapkan dapat memperoleh informasi berdasarkan kombinasi yang terbentuk. *Rule* yang dihasilkan juga akan digunakan sebagai fitur *hints* atau rekomendasi pembelajaran untuk

membantu siswa yang kesulitan dalam memahami soal, sekaligus sebagai informasi bagi pengajar tentang materi-materi yang sulit dikuasai oleh para peserta didik.

Sekolah Menengah Atas Negeri 2 Pamekasan merupakan lembaga pendidikan negeri yang berada di Jl. Jokotole No.234 Pamekasan.

Association rules adalah salah satu teknik yang terkenal dalam data mining. *Association rules* merupakan teknik data mining untuk mencari hubungan antar item dalam suatu *data set* (Hand, 2001). Konsep ini sendiri diturunkan dari terminologi *market basket analysis*, yaitu pencarian hubungan dari beberapa produk di dalam transaksi pembelian (Witten, 2005). *Association rule* meliputi dua tahap, yaitu mencari kombinasi yang paling sering terjadi dari suatu itemset. Dan mendefinisikan *Condition* dan *Result* (untuk *conditional association rule*) (Ulmer, D. 2002)

Association Rules juga bermanfaat untuk pemakaian data web yang berdasarkan personalitas. Pendekatan ini diadopsi dari hubungan dengan *collaborative filtering* (Samuel, 2008). Dalam menentukan suatu *association rules*, terdapat suatu *interestingness measure* (ukuran kepercayaan) yang didapatkan dari hasil pengolahan data dengan perhitungan tertentu (Wahyuni, 2009) Umumnya ada dua ukuran, Support dan Confidence.

Support :suatu ukuran yang menunjukkan seberapa besar tingkat dominasi suatu item/itemset dari keseluruhan transaksi. Ukuran ini menentukan apakah suatu item/itemset layak untuk dicari *confidence*-nya (misal, dari keseluruhan transaksi yang ada, seberapa besar tingkat dominasi yang menunjukkan bahwa item A dan B dibeli bersamaan).

$$Support = P(A \cup B)$$

$$= \frac{\text{jumlah transaksi yang mengandung A dan B}}{\text{total transaksi}} \dots (i)$$

Confidence: suatu ukuran yang menunjukkan hubungan antar 2 item secara conditional (misal, seberapa

sering item B dibeli jika orang membeli item A)

$$Confidence = P(A \cap B)$$

$$= \frac{\text{jumlah transaksi yang mengandung A dan B}}{\text{total transaksi yang mengandung A}} \dots (ii)$$

Misalkan ada sebuah pernyataan suatu *association rule*; *If A then B* [*support*=20%, *confidence*=60%], dimana A dan B adalah kumpulan *item* yang dibeli oleh konsumen perusahaan X. Dapat diartikan bahwa item A dan B dibeli bersamaan sebesar 20% dari keseluruhan data transaksi yang dianalisis dan 60% dari semua konsumen yang membeli *item* A juga membeli *item* B. Kedua ukuran ini nantinya berguna dalam menentukan *interesting association rules*, yaitu untuk dibandingkan dengan batasan (*threshold*) yang ditentukan oleh *user*. Batasan tersebut umumnya terdiri dari *min_support* dan *min_confidence*

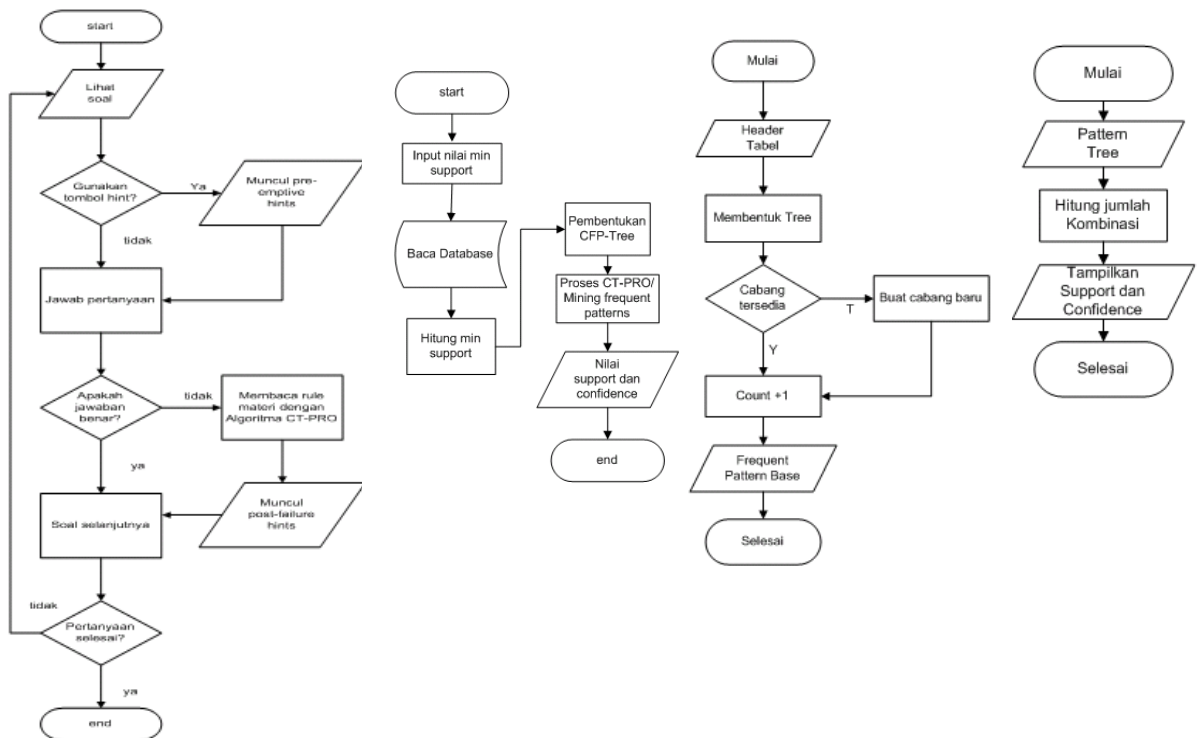
METODE

Algoritma ini berakar dari *FP-Growth* dimana modifikasi yang dilakukan adalah pada struktur data yang digunakan. Struktur data yang digunakan adalah *Compressed FP-Tree (CFP-Tree)* dimana informasi dari sebuah *FP-Tree* diringkas dengan struktur yang lebih kecil, sehingga baik pembentukan tree maupun *frequent itemset mining* dapat dilakukan lebih cepat. Pada tahap *mining* algoritma *CT-PRO* juga menggunakan pendekatan *bottom-up* dimana *item* pada *item tabel* dan *CFP-Tree* dilakukan *scan* dari jumlah terkecil hingga terbesar. Algoritma *CT-PRO* memiliki tiga tahap yaitu (Kurnia, 2013) : menemukan *item-item* yang *frequent*, membuat struktur data *CFP-Tree* dan terakhir melakukan *mining frequent patterns*. Algoritma *CT-PRO* dapat dilihat pada Gambar 1.

Langkah-langkah kerja algoritma *ct-pro*: Langkah 1). Mencari *frequent item*, pada tahap ini terjadi proses-proses sebagai berikut: a). Dari *dataset* yang ada, dilakukan seleksi berdasarkan *minimum support* yang ditentukan sehingga menghasilkan *frequent item*. b). Dari *frequent item* yang telah terbentuk,

dihitung frekuensi kemunculan setiap item sehingga menghasilkan *global item* tabel. Langkah 2). Membangun *CFP-Tree*, pada tahap ini terjadi proses-proses sebagai berikut: a). *Frequent item* yang telah didapatkan, diurutkan berdasarkan *global item* tabel yang ada secara menurun (diurutkan mulai dari *item* berfrekuensi terbesar hingga terkecil). b). Dengan *frequent item* yang telah terurut ini dibentuk *global CFP-Tree*, aturan pembentukan *global CFP-Tree* sebagai berikut: 1). *CFP-Tree* terdiri dari *tree* yang memiliki *root* yang mewakili *index* dari *item* dengan tingkat kemunculan tertinggi dan kumpulan *subtree* sebagai anak dari *root*. 2). Jika $I = \{i_1, i_2, \dots, i_k\}$ adalah kumpulan dari

frequent item dalam transaksi, item dalam transaksi akan dimasukkan ke dalam *CFP-Tree* dimulai dari *root subtree* yang merupakan i_1 dalam *item* tabel. 3). *Root* dari *CFP-Tree* merupakan *level-0* dari *tree*. 4). Setiap *node* dalam *CFP-Tree* memiliki empat *field* utama yakni *item-id*, *parent-id*, *count* yang merupakan jumlah *item* pada *node* tersebut, dan *level* yang menunjukkan struktur data *tree* pada *node* tersebut dimulai dari *item* yang terdapat pada *item* tabel dengan *level* yang terdapat pada *CFP-Tree*.



Gambar 2. Flowchart sistem dan CT-PRO

Langkah 3). *Mining*, pada tahap ini terjadi proses-proses sebagai berikut: a). Pada tahap *mining* ini, algoritma *ct-pro* bekerja dengan melakukan *bottom-up mining* sehingga *Global Item* tabel diurutkan mulai dari *item* berfrekuensi terkecil hingga terbesar. b). Untuk setiap *item* yang terdaftar pada *global item* tabel

yang telah diurutkan, dilakukan pencarian *node* yang berkaitan dengan *item* tersebut pada *global CFP-Tree*. Dari semua *node* yang ditemukan untuk setiap *item* inilah yang disebut dengan *local frequent item* dan digunakan untuk membuat *local item* tabel. c). Pada pembuatan *local item* tabel ini juga

dilakukan berdasarkan jumlah *minimum support* yang telah ditentukan. d). Setelah itu, dibuat *local CFP-Tree* berdasarkan *local item* tabel yang terbentuk.

Aturan pembentukan *local CFP-Tree* sama dengan pembentukan *global CFP-Tree*, yang membedakan adalah pada *global CFP-Tree* yang digunakan dalam pembentukan *tree*-nya adalah *global item* tabel yang terbentuk dari *global item* tabel data. Sedangkan pada *local CFP-Tree* yang digunakan dalam pembentukan *tree*-nya adalah *local item* tabel yang terbentuk dari *local frequent item*. c). Dari *local CFP-Tree* dibentuk *frequent pattern* sesuai dengan *item* yang *dimining*. Dari *frequent pattern* dihitung masing-masing *item* yang memenuhi dihitung *confidencenya*. Apabila memenuhi *minimum confidence* maka masing-masing *item* yang bersangkutan dijadikan sebagai *knowledge*. Rancangan sistem aplikasi dapat dilihat pada Gambar 1

Perhitungan dalam aplikasi ini dimulai dengan memasukkan *minimum support* yang selanjutnya dilakukan proses oleh sistem berupa pemangkasan data yang tidak memenuhi *minimum support*, sehingga diperoleh *header table*. Dari *header tabel* inilah yang kemudian dilakukan pembentukan *tree*, pencarian kombinasi dan kemudian didapatkan nilai *support* dan *confidence* berdasarkan jumlah kombinasi yang terbentuk.

PEMBAHASAN

Data yang digunakan dalam tugas akhir ini merupakan data dari jawaban salah siswa kelas XII yang diperoleh dari SMAN 2 Pamekasan. Data didapatkan dari siswa sebanyak 150 orang yang telah diberikan soal sebelumnya. Dalam implementasinya, jawaban-jawaban salah tersebut yang akan dijadikan sebagai data mentah/transaksi untuk mencari kombinasi keterkaitan dari seluruhnya. Selanjutnya, data akan dipecah menjadi 5 dengan jumlah siswa yang berbeda. Hal ini untuk melihat perbedaan kinerja algoritma CT-PRO dan perbedaan kombinasi yang muncul untuk data yang

jumlahnya berbeda dan *minimum support* yang berbeda pula.

Dalam hal ini, akan digunakan 5 data awal sebagai data ujicoba untuk mencoba sistem, seperti pada tabel dibawah. *Minimum support* yang diinputkan adalah 50% dari seluruh transaksi data.

Tabel 1. Data Percobaan.

Siswa	Materi yang tidak dikuasai
1	M01,M16,M13,M09,M02,M06,M04,M03
2	M19,M15,M13,M06,M07,M03,M02,M01
3	120,M16,M10,M08,M06,M02
4	M18,M17,M16,M11,M03,M02,M09
5	M02,M13,M12,M06,M05,M03,M01,M16

Selanjutnya adalah proses CFP-Tree untuk mencari frekuensi tiap materi berdasarkan *minimum support* yang diminta, materi yang tidak memenuhi akan dihapus dari data karena tidak berpengaruh pada perhitungan selanjutnya. Dan materi yang tersisa, dalam hal ini adalah M02, M03, M06, M16, M01, M13 akan diurutkan berdasarkan frekuensi kemunculan.

Tabel 2. Data Terurut.

Siswa	Materi yang sulit
1	M02,M03,M06,M16,M01,M13
2	M02,M03,M06,M01,M13
3	M02,M06,M16
4	M02,M03,M16
5	M02,M03,M06,M16,M01,M13

Proses terakhir pada aplikasi ini adalah perhitungan *support* dan *confidence*. Setelah sebelumnya ditemukan kombinasi-kombinasi *itemset* yang memenuhi *minimum support*. Sekarang hanya perlu menggunakan rumus *support* dan *confidence*.

Tabel 3. Tabel *support* dan *confidence*

Kombinasi Itemset	Support	Confidence
{M02,M13}	$\frac{3}{5} \times 100\% = 60\%$	$\frac{3}{5} \times 100\% = 60\%$
{M03,M13}	$\frac{3}{5} \times 100\% = 60\%$	$\frac{3}{4} \times 100\% = 75\%$
{M06,M13}	$\frac{3}{5} \times 100\% = 60\%$	$\frac{3}{4} \times 100\% = 75\%$
{M02,M01}	$\frac{3}{5} \times 100\% = 60\%$	$\frac{3}{5} \times 100\% = 60\%$
{M03,M01}	$\frac{3}{5} \times 100\% = 60\%$	$\frac{3}{4} \times 100\% = 75\%$
{M06,M01}	$\frac{3}{5} \times 100\% = 60\%$	$\frac{3}{4} \times 100\% = 75\%$
{M02,M16}	$\frac{4}{5} \times 100\% = 80\%$	$\frac{4}{5} \times 100\% = 80\%$
{M03,M16}	$\frac{3}{5} \times 100\% = 60\%$	$\frac{3}{4} \times 100\% = 75\%$

{M06,M16}	$\frac{3}{5} \times 100\% = 60\%$	$\frac{3}{4} \times 100\% = 75\%$
{M02,M06}	$\frac{4}{5} \times 100\% = 80\%$	$\frac{4}{5} \times 100\% = 80\%$
{M03,M06}	$\frac{3}{5} \times 100\% = 60\%$	$\frac{3}{4} \times 100\% = 75\%$
{M02,M03}	$\frac{4}{5} \times 100\% = 80\%$	$\frac{4}{5} \times 100\% = 80\%$
13	Kombinasi	

Setelah proses perhitungan *support* dan *confidence* dari semua kombinasi berhasil didapatkan. Selanjutnya adalah proses perbandingan untuk jumlah data yang berbeda dan *minimum support* yang berbeda pula.

Tabel 4. Tabel Perbandingan

KESIMPULAN

No	Jumlah Siswa	Minimum Support	Jumlah Kombinasi
1	5	25%	13
		50%	3
		75%	3
2	30	25%	15
		50%	2
		75%	0
3	60	25%	17
		50%	3
		75%	0
4	90	25%	20
		50%	3
		75%	0
5	120	25%	22
		50%	4
		75%	0
6	150	25%	23
		50%	4
		75%	0

Beberapa kesimpulan yang dapat diambil dari hasil penelitian ini adalah Jumlah kombinasi yang terbentuk berbanding lurus dengan jumlah data, untuk nilai *minimum support* yang sama. Semakin besar nilai *minimum support*, semakin sedikit kombinasi yang terbentuk.. Pada aplikasi ini Jumlah soal pada aplikasi ini masih bisa ditambah, jumlah siswa dan materi bisa ditambah sehingga *rule* dan kombinasi yang didapat semakin banyak. Aplikasi ini masih bisa dikembangkan dengan metode lain

DAFTAR PUSTAKA

- Dominguez, A.K., Yacef, K., Curran, J.R. Data Mining for *Individualised Hints in eLearning*. *Educational Data Mining Conference Proceedings*. 1: 91-100. 2010.
- Giri, Sucahyo Yudho. And Gopalan, raj P. "CT-Pro : A Bottom-Up Non Recursive Frequent Itemset Mining Algorithm Using Compressed FP-Tree Data Structure " *Departement of computing, Curtin Univercity of technology kent st,bentley western australia 6102*. 2005.
- Ruldeviyani, yova and Fahrian, Muhammad, "implementasi algoritma-algoritma association rules sebagai bagian dari pengembangan data mining algorithms collection": November 2008.
- Hand, D., Mannila. H., Smyth, P. *Principles of Data Mining*. Cambridge: A Bradford Book. The MIT Press. 2001
- Witten, I. H. and Frank, E. *Data Mining: Practical Machine Learning Tools and Techniques 2nd Edition*. Morgan Kaufmann Publisher. 2005
- Ulmer, D. 2002. *Mining an Online Auctions Data Warehouse*. The Mid-Atlantic Student Workshop on Programming Languages and Systems. New York: Pace University.
- Samuel, D. *Penerapan Stuktur FP-Tree dan Algoritma FP-Growth dalam Optimasi Penentuan Frequent Itemset*. Institut Teknologi Bandung. 2008.
- Wahyuni, F. S., Siahaan, D. O., Faticah. C. *Penggunaan Cluster-Based Sampling Untuk Penggalan Kaidah Asosiasi Multi Obyektif*. KURSOR. 5: 59-67. 2009.
- Kurnia, Parama Fadli. 2010. "innovation of information technology : algoritma ct-pro" . URL: http://paramafadlikurnia.blogspot.sg/2010/12/algoritma-ct-pro_5.html, diakses 09 maret 2013.