

PENENTUAN REKOMENDASI PRODUK DENGAN METODE DATA MINING ASOSIASI *GENERALIZED SEQUENCE PATTERN* (GSP)

Elly Muningsih

Manajemen Informatika, AMIK BSI Yogyakarta
Ambarketawang, Jl. Ring Road Barat Gamping, Sleman Yogyakarta
e-mail : elly.emh@bsi.ac.id

ABSTRAK

Perusahaan yang mampu menjawab kebutuhan pasar akan menjadi lebih unggul dibanding yang lain. Salah satunya adalah mengetahui produk-produk yang diminati atau disukai konsumennya. Produk yang diminati konsumen dalam satu wilayah akan berbeda dengan dengan wilayah lainnya. Tujuan dari penelitian ini adalah untuk mengetahui produk yang paling diminati oleh konsumen sehingga perusahaan bisa merekomendasikan produk tersebut sehingga diharapkan dapat meningkatkan penjualan. Wilayah Geografis yang diteliti adalah wilayah Jawa, Sumatera, Kalimantan dan Sulawesi. Penelitian ini menggunakan metode Generalized Sequence Pattern (GSP) yang fungsi utamanya adalah menemukan sequensial atau urutan. Untuk pengukuran penelitian menggunakan metode Precision, Recall dan F1. Data penelitian diambil dari data transaksi penjualan sebuah online shop di Jogja yang telah melakukan pengiriman ke seluruh wilayah Indonesia. Hasil penelitian menunjukkan bahwa produk yang diminati konsumen satu wilayah berbeda dengan lainnya. Sedangkan pengukuran menghasilkan akurasi yang tinggi, dimana akurasi tertinggi adalah wilayah Kalimantan dan akurasi terkecil adalah Sumatera.

Kata Kunci : produk, metode Generalized Sequence Pattern (GSP), transaksi penjualan

ABSTRACT

Companies that are able to answer the needs of the market will be superior than others. One is to know the products of interest or preferred customers. Products that consumers demand in the region will be different from the other regions. The aim of this study was to determine the products most in demand by consumers so that the company can recommend the products that are expected to boost sales. Geographic regions studied were Java, Sumatera, Kalimantan and Sulawesi. This study uses the Generalized Sequence Pattern (GSP) methods whose main function is to find sequential or sequence. For the measurement of research using methods Precision, Recall and F1. Research source were taken from a sales transaction data online shop in Yogyakarta which has been sending to the entire territory of Indonesia. The results showed that the products that consumers demand a different region to another. While measurement produce high accuracy, where the highest accuracy is the Kalimantan and the smallest accuracy is Sumatera.

Keywords : product, Generalized Sequence Pattern (GSP) method, sales transaction

I. PENDAHULUAN

Suatu perusahaan yang ingin sukses dan berhasil dalam dunia bisnis dituntut untuk selalu bisa menerapkan strategi dan langkah yang tepat agar tidak tertinggal oleh *trend* dan perkembangan yang ada. Dapat menjawab kebutuhan pasar merupakan salah satu faktor penting yang menjadikan suatu perusahaan dapat unggul dibanding yang lain. Teknologi informasi memegang peranan yang penting untuk ini, karenanya pemanfaatan teknologi yang tepat akan menghasilkan informasi yang terkini. Teknologi internet saat ini berkembang sangat pesat terutama dalam dunia bisnis, hal ini dapat dilihat dengan munculnya *electronic commerce* (*e-commerce*) [1]. Untuk kepentingan perancangan strategi pemasaran, pemahanan akan kebutuhan dan keinginan konsumen menjadi pedoman yang utama. Modus tindakan pembelian yang dilakukan konsumen adalah mencapai suatu kepuasan dimana permintaan akan produk menjadi bervariasi disamping pula pola konsumsi konsumen yang berbeda-beda. Variasi yang dimaksud akhirnya mendorong pembagian atau dikenal dengan segmentasi pasar. Segmentasi pasar merujuk pada proses pembagian pasar [2].

Segmentasi pasar dalam dunia pemasaran dibedakan menjadi 4 jenis yaitu Segmentasi Geografis, Segmentasi Demografi, Segmentasi Psikografi dan Segmentasi Tingkah Laku [3]. Perbedaan segmentasi pasar ini menentukan sasaran promosi produk yang berbeda-beda pada konsumen, dimana diketahui bahwa ada kesamaan pola perilaku belanja konsumen di segmentasi yang sama. Dari pola belanja inilah ditemukan *related product* atau asosiasi antar item produk yang dibeli pada tiap-tiap segmen. Selanjutnya *related product* inilah yang kemudian akan dijadikan sebagai acuan atau

rekomendasi produk untuk kepentingan promosi penjualan dalam rangka penerapan strategi pemasaran diferensiasi [4].

Namun demikian penentuan rekomendasi produk untuk promosi penjualan pada masing-masing segmentasi geografis kurang akurat karena harus berdasarkan pengetahuan dari jumlah data transaksi penjualan yang besar [4]. Karena hal itu untuk mendapatkan pengetahuan tersebut maka diperlukan suatu proses pengolahan data historis transaksi besar diperlukan suatu teknik *data mining*. *Data Mining* didefinisikan sebagai suatu proses untuk menemukan pola-pola dalam data, dimana proses dilakukan secara otomatis atau semi-otomatis. Pola yang ditemukan harus memiliki makna yang menghasilkan manfaat atau keuntungan, dan data yang digunakan adalah dalam jumlah besar [5]. Teknik data mining yang akan digunakan pada penelitian ini metode *Generalized Sequential Pattern (GSP)*.

A. Asosiasi Generalized Sequence Pattern (GSP)

Sebuah record transaksi penjualan biasanya berisi tanggal transaksi dan item yang terjual. Namun ada juga data transaksi yang berisi identitas dari pelanggan misalnya *IDCustomer*, dan dengan *IDCustomer* dapat dilakukan *Mining Sequential Pattern* [6]. *Mining Sequential Pattern* menganalisa pola pembelian konsumen yang menemukan pola sering atau berkali-kali pada sebuah urutan database yang dipandang penting pada banyak masalah *data mining* [7]. Secara umum, algoritma *mining sequential pattern* dibedakan menjadi dua metode utama [7] [8] yaitu :

- 1) *Apriori-based*, terdiri dari algoritma GSP (*Generalized Sequential Pattern Mining*) yang merupakan metode *mining sequential pattern* dengan format horizontal dan algoritma SPADE (*Sequential Pattern Discovery using Equivalent Class*) yang mengadopsi format vertical pada *mining sequential pattern*
- 2) *Projection-based*, terdiri atas algoritma *Freespan* dan *Prefixspan*, yang menerapkan sebuah pola pembagian dan rangkaian strategi untuk efisiensi *mining sequential pattern*.

Algoritma GSP secara umum dipandang sebagai algoritma traversal pertama yang menemukan semua urutan yang sering muncul dengan cara melewati beberapa data [8]. Menurut J. Zaki (1997) dalam [6] menyebutkan algoritma GSP atau dengan nama lain *apriori all* adalah suatu algoritma yang dapat memproses dan menemukan semua pola sekuensial dan non sekuensial yang ada. Berdasarkan pada atribut penutupan suatu pola sekuensial, GSP mengadopsi banyak cara pada calon generasi dan pendekatan uji pada *mining sequential pattern* [7]. Algoritma GSP digunakan pada *mining sequence* dan baik untuk memecahkan masalah *mining sequence* yang banyak didasarkan pada sebuah algoritma *Apriori* [8]. Fungsi utama dari algoritma GSP yaitu menemukan pola sekuensial atau urutan, seperti ditampilkan pada Tabel 1 berikut :

TABEL I.
TABEL POLA SEQUENCE DATABASE

IDCustomer	Sequance
10	<a(abc)(ac)d(cf)>
20	<(ad)c(bc)(ac)>
30	<(ef)ab)(df)cb)>
40	<eg(af)cbc>

Pada *IDCustomer* ke-10 memiliki sequence <a(abc)(ac)d(cf)>. Ini berarti bahwa *subsequence* dari *IDCustomer* ke-10 yaitu a, (abc), (ac), d, dan (cf). Maka pada transaksi pertama membeli a, transaksi kedua membeli a, b dan c. Kemudian pada transaksi ketiga membeli a dan c. pada transaksi selanjutnya membeli d dan pada transaksi terakhir membeli c dan f. Setiap transaksi disebut juga *subsequence*.

Algoritma GSP menciptakan banyak cara dalam database [8]. Salah satu caranya, setiap item atau produk tunggal (urutan pertama) dihitung. Dari item yang sering keluar atau muncul, tercipta sebuah kumpulan kandidat urutan kedua, dan cara lain dibuat untuk melengkapi urutan tersebut. Urutan kedua yang sering keluar digunakan untuk menghasilkan kandidat urutan ketiga. Ulangi proses ini sampai tidak ada urutan yang sering keluar ditemukan lagi.

B. Precision, Recall dan F1

Precision didefinisikan sebagai suatu rasio item relevan yang dipilih terhadap semua item yang terpilih. Hal ini dapat diartikan juga sebagai kecocokan antara permintaan informasi dengan jawaban atas permintaan tersebut. *Precision* dapat dihitung dengan jumlah produk relevan yang dipilih konsumen dibagi dengan semua produk yang terpilih baik relevan maupun tidak. Sedangkan *Recall* didefinisikan sebagai rasio dari item relevan yang dipilih terhadap total jumlah item relevan yang tersedia. *Recall* dihitung dengan jumlah rekomendasi produk yang relevan yang dipilih konsumen dibagi dengan jumlah

semua rekomendasi yang relevan baik dipilih maupun rekomendasi yang tidak terpilih. *Precision* dan *Recall* dirumuskan :

$$Precision = \frac{\Sigma \text{item produk relevan yang dipilih konsumen}}{\Sigma \text{semua item produk yang terpilih}} \quad (1)$$

$$Recall = \frac{\Sigma \text{item produk relevan yang dipilih konsumen}}{\Sigma \text{semua item produk yang relevan}} \quad (2)$$

Sedangkan untuk representasi dari penggabungan antara *Precision* dan *Recall* digunakan rumus *F1*, yaitu :

$$F1 = 2PR / (P+R) \quad (3)$$

Nilai *F1* merupakan tingkat akurasi terhadap sistem dalam memberikan rekomendasi produk yang diinginkan. Sistem akan dianggap baik jika memiliki tingkat akurasi (*F1*) yang tinggi.

II. METODE PENELITIAN

Ada enam tahapan atau langkah CRISP-DM (*Cross-Industry Standard Process for Data Mining*) dalam *data mining* [9]:

1. *Business Understanding*

Tahap pertama adalah memahami tujuan dan kebutuhan dari sudut pandang bisnis, kemudian menterjemahkan pengetahuan ini ke dalam pendefinisian masalah dalam data mining. Dan selanjutnya akan ditentukan rencana dan strategi untuk mencapai tujuan tersebut. Pada tahap ini dilakukan pengamatan dan diketahui bahwa *online shop* tidak pernah melakukan rekomendasi produk, transaksi hanya berdasarkan pembelian konsumen. Ini menjadi permasalahan *online shop* dimana penjualan produk juga keuntungan tidak bisa meningkat dan merupakan imbas dari analisis yang kurang akurat. Tujuan dari penelitian ini adalah untuk mengetahui produk yang paling diminati oleh konsumen sehingga *online shop* bisa merekomendasikan produk tersebut sehingga diharapkan dapat meningkatkan penjualan.

2. *Data Understanding*

Tahap ini dimulai dengan pengumpulan data yang kemudian akan dilanjutkan dengan proses untuk mendapatkan pemahaman yang mendalam tentang data, mengidentifikasi masalah kualitas data.

a. Pengumpulan data awal

Data penelitian diambil dari data transaksi penjualan *online shop* Ragam Jogja dari bulan Januari 2011 – Mei 2012.

b. Mendeskripsikan data

Data transaksi penjualan yang digunakan berjumlah 235 transaksi.

c. Pemilihan atribut

Atribut yang digunakan adalah berdasarkan produk atau jenis produ yang dijual. Dalam penelitian ini kode produk yang digunakan berjumlah 31.

3. *Data Preparation*

Tahap ini dapat diulang beberapa kali. Pada tahap ini juga mencakup pemilihan tabel, *record*, dan atribut-atribut data, termasuk proses pembersihan dan transformasi data untuk kemudian dijadikan masukan dalam tahap pemodelan (*modeling*). Proses preparation ini mencakup tiga hal utama yaitu:

a. *Data Selection*

Data Selection pada data pada tahap ini adalah mengambil 80 data dari data utama dengan metode pemilihan sample *purposive sampling*.

b. *Data Preprocessing*

Preprocessing pada tahap ini adalah membagi data yang digunakan untuk sample menjadi 4 kelompok berdasarkan wilayah Jawa, Sumatera, Kalimantan dan Sulawesi dengan masing-masing wilayah berisi 20 data. Data *cleansing* bekerja untuk membersihkan nilai yang kosong, tidak konsisten atau mungkin tupel yang kosong (*missing value* dan *noisy*) sehingga data yang tidak valid tidak akan digunakan dalam proses berikutnya.

c. *Data Transformation*

Proses transformasi data dengan cara merubah kode produk yang terjual sebagai atribut lama dengan kode produk baru untuk memudahkan pemrosesan data dan beberapa atribut lain yang tidak digunakan dihilangkan, seperti ditampilkan pada tabel 2 berikut :

TABEL II.
TAHAP SELECTION, PREPROCESSING DAN TRANSFORMATION

No	Atribut Lama	Nama Produk	Atribut Baru
1.	AB_WJ	Batik ABG Warna Panjang	p1
2.	AB_WD	Batik ABG Warna Pendek	p2
3.	AB_MJ	Batik ABG Melati Panjang	p3
4.	AB_MD	Batik ABG Melati Pendek	p4
5.	AB_DD	Dres ABG Batik	p5
6.	BD_SG	Blous Batik Dewasa Sogan	p6
7.	BD_K1	Blous Batik Dewasa Kualitas 1	p7
8.	BD_K2	Blous Batik Dewasa Kualitas 2	p8
9.	DD_K1	Dres Dewasa Batik Kualitas 1	p9
10.	DD_K2	Dres Dewasa Batik Kualitas 2	p10
11.

4. Modelling

Pada tahap ini dilakukan pemilihan tipe dan algoritma yang digunakan juga penerapannya. Dan aturan asosiasi *Generalized Sequential Pattern* (GSP) untuk menentukan urutan produk yang paling banyak diminati pada tiap-tiap segmentasi konsumen berdasarkan wilayah geografi yaitu wilayah Jawa, Sumatera, Kalimantan dan Sulawesi. Penerapan aturan asosiasi dengan metode *Generalized Sequential Pattern* (GSP) untuk menentukan urutan produk yang paling diminati pada tiap-tiap segmen konsumen berdasarkan wilayah geografi.

5. Evaluation

Pada tahap ini dilakukan evaluasi terhadap sampel data yang dihasilkan. Dari hasil pengolahan data tersebut kemudian dihitung perbandingan antara jumlah produk yang dianggap relevan dan tidak relevan dengan jumlah produk yang dipilih dan tidak dipilih dengan menggunakan metode *Precision* dan *Recall*. Untuk selanjutnya proses analisis data, akan dihitung tingkat akurasi dengan menggunakan rumus F1, sehingga akan diketahui jumlah akurasi dari sistem tersebut.

6. Deployment

Pada tahap ini diketahui hasil dari penerapan metode *Generalized Sequential Pattern* (GSP) dalam penentuan rekomendasi produk *online shop* berdasarkan segmentasi geografis berbeda antara wilayah satu dengan lainnya.

III. HASIL

A. Penerapan Metode Generalized Sequential Pattern (GSP)

Transaksi Penjualan berisi produk terjual untuk wilayah Jawa ditampilkan pada Tabel 3 dibawah ini :

TABEL III.
DAFTAR TRANSAKSI WILAYAH JAWA

No	Daftar Transaksi
1.	p2p5p12p18
2.	p12p14p22p23p26p27
3.	p8p26p31
4.	p7p8p9p15
5.	p14p18p20
6.	p1p18
7.	p1p6p14p15p22
8.	p1p2p6p8p10p22p27p28
9.	p17p26
10.	p16p17
11.	Dst

Langkah penerapan algoritma GSP adalah sebagai berikut :

1. Hitung frekuensi atau jumlah item produk yang keluar pada transaksi, kemudian tentukan *minimum support* (dukungan terkecil). *Minimum support* adalah nilai minimal dari frekuensi yang diambil untuk batas dukungan yang diinginkan. Pada penelitian ini diambil 3. Frekuensi produk terjual wialayah Jawa ditampilkan pada Tabel 4 seperti berikut :

TABEL IV.
FREKUENSI PRODUK TERJUAL

Produk	Jumlah	Produk	Jumlah	Produk	Jumlah
p1	4	p11	1	p21	0
p2	3	p12	3	p22	4
p3	0	p13	2	p23	3
p4	0	p14	5	p24	0
p5	1	p15	2	p25	0
p6	4	p16	1	p26	4
p7	1	p17	3	p27	3
p8	3	p18	7	p28	2
p9	1	p19	0	p29	0
p10	3	p20	3	p30	0
				p31	2

2. Hapus anggota dengan nilai frekuensi dibawah 3, ditampilkan pada Tabel 5 berikut :

TABEL V.
PRODUK DENGAN FREKUENSI MINIMAL 3

Produk	Frekuensi
p1	4
p2	3
p6	4
p8	3
p10	3
p12	3
p14	5
p17	3
.....

3. Kombinasikan elemen-elemen ke dalam dua bagian dan temukan frekuensi tertinggi pada database transaksionalnya, ditampilkan pada Tabel 6 berikut :

TABEL VI.
KOMBINASI ELEMEN / PRODUK

Produk	frekuensi	produk	frekuensi
p1p2	2	p1p22	2
p1p6	3	p2p22	1
p2p6	2	p6p22	2
p1p8	1	p8p22	1
p2p8	1	p10p22	1
p6p8	1	p12p22	0
p1p10	1	p14p22	3
p2p10	1	p17p22	0
p6p10	2	p18p22	0
p8p10	1	p20p22	0
p1p12	1	p1p23	0
p2p12	2	p2p23	0
.....

4. Kemudian hapus nilai frekuensi dibawah 3 seperti ditampilkan pada Tabel 7 :

TABEL VII.
FREKUENSI 3

Produk	Frekuensi
p2p6	3
p18p20	3
p14p22	3

Dari Tabel 7. diperoleh data bahwa urutan produk yang diminati konsumen untuk wilayah Jawa adalah “p2p6p14p18p20p22”.

Dengan cara yang sama pada langkah penerapan metode GSP, didapatkan hasil untuk urutan produk yang diminati berdasarkan segmentasi geografis yaitu seperti ditampilkan pada Tabel 8 berikut

TABEL VIII.
HASIL URUTAN PRODUK DIMINATI PER WILAYAH

Wilayah	Urutan Produk yang Diminati	Nama Produk
---------	-----------------------------	-------------

Jawa	p2p6p14p18p20p22	Batik ABG Warna Pendek, Blous Batik Dewasa Sogan, Batik Eksklusif, Hem Batik Pendek Warna, Bawahan Batik Model Rok, Sarimbit Batik Sogan
Sumatera	p14p18p22p24p26p28	Batik Eksklusif, Hem Batik Pendek Warna, Sarimbit Batik Sogan, Sarimbit Batik Model Blous, Sarimbit Batik Model Gamis, Batik Anak Lelaki
Kalimantan	p7p8p14p18	Blous Batik Dewasa Kualitas 1, Blous Batik Dewasa Kualitas 2, Batik Eksklusif, Hem Batik Pendek Warna
Sulawesi	p6p17p18p22p23	Blous Batik Dewasa Sogan, Hem Batik Pendek Sogan, Hem Batik Pendek Warna, Sarimbit Batik Sogan, Sarimbit Batik PG Eksklusif

B. Pengukuran dan Evaluasi

Pengukuran hasil penelitian ini menggunakan metode *Precision*, *Recall* dan *F1*. Data yang digunakan dalam melakukan pengukuran didapatkan dari 3 data transaksi baru yang kemudian diurutkan menjadi data tunggal. Data baru adalah data transaksi setelah tanggal 31 Mei 2012. Untuk perhitungan wilayah Jawa adalah sebagai berikut :

Data : p1, p10, p14, p18, p20, p22, p25

X=5, Y=2, Z=1;

Sehingga *F1* dapat dihitung :

$$Precision = X/(X+Y) = 5/(5+2) = 0,71$$

$$Recall = X/(X+Z) = 5/(5+1) = 0,83$$

$$F1 = 2 PR / (P+R) \\ = (2 \times 0.71 \times 0.83) / (0.71 + 0.83) \\ = 1.18/1.54 = 0,76$$

Dengan nilai *F1* = 0.76, maka tingkat akurasi untuk wilayah Jawa adalah tinggi. Hasil pengukuran nilai akurasi untuk wilayah lainnya ditampilkan pada Tabel 9 :

TABEL IX.
PENGUKURAN NILAI AKURASI PER WILAYAH

Wilayah	P	R	Z	F1	Akurasi
Jawa	5	2	1	0,76	Tinggi
Sumatera	5	3	1	0,71	Tinggi
Kalimantan	4	1	0	0,89	Tinggi
Sulawesi	5	2	0	0,83	Tinggi

IV. PEMBAHASAN

Dari hasil diatas dapat diambil beberapa hal atau poin penting, yaitu :

- 1) Produk yang paling diminati di tiap-tiap wilayah yaitu wilayah Jawa produk p2, wilayah Sumatera produk p14, wilayah Kalimantan produk p7, dan wilayah Sulawesi Produk p6
- 2) Terdapat satu produk yang diminati semua wilayah yaitu produk p18
- 3) Terdapat satu produk yang diminati di tiga wilayah Jawa, Sumatera dan Sulawesi yaitu produk p22
- 4) Terdapat dua produk yang diminati dua wilayah yaitu wilayah Jawa dan Sulawesi yaitu produk p6, dan wilayah Jawa dan Kalimantan yaitu produk p14

Sedangkan dari hasil pengukuran didapatkan hasil bahwa akurasi untuk metode *Generalized Sequence Pattern* dengan metode pengukuran *precision*, *recall* dan *F1* menghasilkan akurasi yang tinggi, dimana akurasi tertinggi adalah wilayah Kalimantan dan akurasi terkecil adalah Sumatera. Dengan hasil yang sudah didapatkan maka *online shop* bisa melakukan promosi penjualan dengan cara merekomendasikan produk sesuai wilayah konsumen.

IV. SIMPULAN DAN SARAN

Penelitian ini menghasilkan urutan produk yang diminati konsumen tiap-tiap wilayah, dimana diketahui bahwa minat produk konsumen berbeda-beda antara satu wilayah dengan wilayah lainnya. Pengukuran yang dilakukan juga menunjukkan hasil yang akurat. Dengan adanya hasil ini diharapkan *online shop* bisa menerapkannya untuk rekomendasi produk berdasarkan wilayah untuk meningkatkan penjualan dan keuntungan. Saran yang ingin disampaikan untuk penelitian berikutnya adalah bisa dilakukan perbandingan atau komparasi menggunakan metode asosiasi yang lain agar diketahui hasil yang lebih baik.

REFERENSI

- [1] R. Ustadiyanto, "Framework e-Commerce". Yogyakarta: Penerbit Andi, 2001.
- [2] Widyatama, "Segmentasi Pasar". Retrieved Maret 20, 2012, from <http://dspace.widyatama.ac.id/bitstream/handle/10364/999/bab2b.pdf?sequence=10>, 2008.
- [3] R. R. Giri and N. Afiyanti, "An Analysis of Online User Forum Community Segmentation at www.Kaskus.us", *International Conference on Economics Marketing and Management*. Singapore: IACSIT Press, 2012.
- [4] F. H. Setiawan, *Penerapan Fuzzy C-Means Dan Apriori Untuk Rekomendasi Promosi Produk Berdasarkan Segmentasi Konsumen*. Semarang, 2011.
- [5] Witten, I. H., Frank, E., & Hall, M. A. "Data mining: Practical Machine Learning Tools and Techniques 3rd Edition". Elsevier, 2011.
- [6] G. S. Budhi, A. Handoyo and C. O. Wirawan, "Algoritma Generalized Sequential Pattern Untuk Menggali Data Sekuensial Sirkulasi Buku Pada Perpustakaan UK Petra". *Seminar Nasional Aplikasi teknologi Informasi (SNATI 2009)*. Yogyakarta, 2009.
- [7] J. Han, J. Pei and X. Yan, "Sequential Pattern Mining by Pattern-Growth : Principles and Extension". *StudFuzz 180* , 183-220, 2005.
- [8] S. M. Halawani, S. Shaik and E. Prasad, "Sales Promotion System in E-Commerce using Data Mining Technique". *IJCSNS International Journal of Computer Science and Network Security* , 10 (5), 103-109, 2010.
- [9] P. Chapman, J. Clinton, R. Kerber et al, "CRISP-DM 1.0 Step-by-Step Data Mining Guide", 2000.