

IMPLEMENTASI DATA MINING DALAM PENCARIAN DAERAH STRATEGIS UNTUK PENGENALAN SEKOLAH SWASTA DENGAN METODE FP-GROWTH

M. Afdal

Program Studi Sistem Informasi Fakultas Sains dan Teknologi UIN Suska Riau
Jl. HR. Soebrantas No. 155 Simpang Baru, Panam, Pekanbaru, Riau 28293
Email: m.afdal@uin-suska.ac.id

ABSTRAK

Sekolah Menengah Kejuruan (SMK) adalah salah satu institusi pendidikan resmi yang disahkan oleh pemerintah. Berdasarkan data pokok Sekolah Menengah Kejuruan untuk dinas pendidikan Kota Pekanbaru tahun 2016 terdapat sebanyak 60 Sekolah yang terdiri dari 7 SMK Negeri dan 53 SMK Swasta. Persaingan di dalam dunia bisnis, khususnya dalam bidang pendidikan pada SMK membuat pihak sekolah harus mencari pola sasaran daerah yang strategis dalam pengenalan sekolah. Dengan semakin banyaknya SMK Swasta di Kota Pekanbaru, membuat setiap sekolah berusaha mencari calon siswa baru kedaerah-daerah yang potensial. Salah satu cara yang dapat dilakukan untuk penentuan daerah strategis adalah dengan memanfaatkan teknik *Data Mining*. Dari data-data siswa yang ada disekolah dapat diolah menggunakan algoritma *FP-Growth* sehingga menghasilkan *Frequent Itemset* yang menjadi informasi baru untuk dimanfaatkan oleh sekolah dalam menentukan daerah yang strategis. Dalam penelitian ini yang menggunakan data siswa kelas X dengan nilai *minimum support* = 0.04 dan nilai *minimum confidence* = 0.68 dinyatakan bahwa siswa yang berasal dari kecamatan Payung Sekaki adalah daerah yang paling strategis dalam pengenalan sekolah dengan tingkat kepercayaan 100% dan didukung oleh 4.7% dari data keseluruhan dengan nilai *lift ratio* 1.472.

Kata Kunci: *Association Rule, Data Mining, FP-Growth, Frequent Itemset*

I. PENDAHULUAN

SMK Tunas Karya merupakan salah satu institusi pendidikan menengah kejuruan Swasta yang berada di Kota Pekanbaru. Menurut data pokok Direktorat Jenderal Pendidikan Dasar dan Menengah Kementerian Pendidikan dan Kebudayaan jumlah Sekolah Menengah Kejuruan (SMK) mencapai angka 12.942 Sekolah pada tahun ajaran 2015/2016. Sementara itu untuk daerah kota Pekanbaru berdasarkan data pokok Sekolah Menengah Kejuruan terdapat sebanyak 60 Sekolah yang terdiri dari 7 SMK Negeri dan 53 SMK Swasta.

Setiap tahun nya SMK Tunas Karya selalu mengalami grafik penerimaan siswa baru yang kurang konsisten, karna sulitnya mendatangkan calon siswa baru untuk belajar di SMK Swasta. Persaingan di dalam dunia bisnis, khususnya dalam bidang pendidikan pada Sekolah Kejuruan Swasta membuat pihak sekolah harus mencari pola sasaran daerah yang strategis dalam pengenalan Sekolah. Dengan semakin banyaknya SMK Swasta di Kota Pekanbaru, membuat setiap sekolah berusaha mencari calon siswa baru ke daerah-daerah yang potensial.

Jika pemasaran tidak ditentukan secara baik, dalam arti tidak diupayakan mencari sasaran daerah yang potensial, maka hanya akan menghabiskan banyak waktu dan biaya yang seharusnya bisa diminimalisir melalui pemilihan daerah target promosi yang baik. Salah satu cara yang dapat diterapkan adalah dengan menggunakan konsep *Data Mining*. Menurut Gunadi dan Sensuse (2012) *Data Mining* adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Salah satu contoh dalam penelitian yang menggunakan data-data pada pendekatan *database*, sehingga dapat memprediksi konsumsi energi HVAC dan kondisi lingkungan beberapa zona yang dibangun oleh algoritma *Data Mining* (Zeng dan Kusiak, 2015).

Algoritma *FP-Growth* merupakan salah satu alternatif yang dapat dimanfaatkan untuk menentukan himpunan data yang paling sering muncul dalam sebuah kumpulan data (Gunadi dan Sensuse, 2012) karena pada algoritma *FP-Growth* dibutuhkan pemanggilan *database* hanya dua kali (Kaur dan Madan, 2015). Berdasarkan permasalahan di atas, maka dimungkinkan untuk dibuat strategi dalam mendukung pengenalan sekolah. Sehingga dimasa akan datang pihak sekolah bisa lebih memfokuskan terhadap daerah yang cukup potensial.

II. METODE PENELITIAN

Adapun tahapan dalam melakukan penelitian ini dijelaskan sebagai berikut:

A. Mengumpulkan dan menyeleksi Data.

Dalam tahap ini dilakukan pengumpulan data untuk lebih mengetahui mengenai data yang diteliti. Dari data yang dikumpulkan akan dapat diketahui

mengenai sistem yang sedang berjalan disekolah. Data-data yang diperlukan dalam penelitian ini adalah data siswa SMK Tunas Karya pada T.A 2015-2016.

B. Menganalisa Data dengan menerapkan Metode *FP-Growth*.

Tahapan ini data yang telah dikumpulkan akan dianalisis. Analisis menggunakan *Association Rule* yang dimaksud dilakukan melalui mekanisme penentuan *minimum support* dan *minimum confidence* dari suatu hubungan *item*. Sebuah *association rule* dikatakan *interesting* jika nilai *support* adalah lebih besar dari *minimum support* dan juga nilai *confidence* adalah lebih besar dari *minimum confidence*.

C. Mengimplementasi Software

Dalam tahap ini dilakukan uji coba terhadap data-data penelitian tersebut dengan memanfaatkan *Rapidminer* sebagai *tools* menggunakan metode *association rule* dengan menerapkan algoritma *FP-Growth*.

D. Menguji Hasil

Proses ini merupakan proses yang terakhir dari rangkaian dalam sistem ini. Proses ini akan menampilkan tabel hasil yang berisi dari data-data siswa yang telah diinputkan sebelumnya. Pengujian dengan menggunakan *tools Data Mining Rapidminer 7.0.1*

III. HASIL DAN PEMBAHASAN

Analisa Kebutuhan Data

Data yang digunakan dalam penelitian ini adalah data siswa Sekolah Menengah Kejuruan (SMK) Tunas Karya Pekanbaru Tahun Ajaran 2015/2016 yang disimpan dalam *database Microsoft excel* oleh pihak administrasi sekolah. Selanjutnya data akan menjadi data inputan pada *tools Rapidminer 7.0.1* yang digunakan. Adapun variabel data siswa seperti pada tabel 1.

Tabel 1 Variabel Dataset Siswa

Variabel	Keterangan
NIS	Identitas diri siswa
Tgl Lahir	Tanggal lahir siswa
Nama	Nama lengkap
Tempat	Tempat lahir

Jurusan	Jurusan prodi yang dipilih siswa
Agama	Agama keyakinan siswa
Asal Sekolah	Asal sekolah waktu Sekolah Menengah Pertama
Nama Orang Tua	Nama lengkap orang tua
Alamat	Alamat lengkap siswa
Kecamatan	Kecamatan dari alamat siswa
Pekerjaan	Pekerjaan dari orang tua siswa
Penghasilan Ortu	Penghasilan dari ayah dan ibu siswa

Pada Tabel 1 terdapat 12 variabel *dataset* siswa. Dari 12 *datasets* tersebut akan dipilih 4 variabel data untuk penelitian yang sesuai dengan kebutuhan, terdapat hubungan yang dapat diambil kesimpulan dari masing-masing variabel nantinya, pola hubungan ini nantinya yang dapat dianalisa dari *association rule* yang diperoleh berdasarkan tingkat kepercayaan (*confidence*) yang tinggi dan didukung dari seluruh data (*support*), dilanjutkan dengan kevalidan data oleh *lift ratio*. Variabel yang digunakan dalam penelitian ini adalah seperti pada tabel 2.

Tabel 2 Variabel Dataset Siswa untuk Penelitian

No	Variabel	Jenis	Inisial
1	Jurusan	Teknik Komputer Jaringan	TKJ
		Teknik Sepeda Motor	TSM
		Teknik Kendaraan Ringan	TKR
2	Kecamatan	Payung Sekaki	PS
		Marpoyan Damai	MD
		Tenayan Raya	TR
		Kampar	KP
		Tampan	TN
		Rumbai	RB
		Labuh Baru Barat	LBB
		Labuh Baru Timur	LBT
		Siak Hulu	SH
		Rumbai Pesisir	RP
		Senapelan	SL
		Tambang	TG
Lima Puluh	LP		

No	Variabel	Jenis	Inisial
		Sukajadi	SK
		Bukit Raya	BR
3	Asal Sekolah	Sekolah Menengah Pertama Negeri	SN
		Sekolah Menengah Pertama Swasta	SS
		Madrasah Tsanawiyah	MT
		Pesantren	PE
4	Penghasilan Orang Tua	< Rp. 1.500.000 (Rendah)	R
		> Rp. 1.500.000 < Rp. 5.000.000 (Sedang)	S
		> Rp. 5.000.000 (Tinggi)	T

Berdasarkan variabel penelitian yang telah dijelaskan, terdapat beberapa pembagian dari setiap variabel tersebut, dan juga pemberian inisial disetiap variabel agar nantinya dalam pembuatan *FP-Tree* lebih efisien dan mudah, karna saat pembuatan *FP-Tree* inisial variabel sangat diperlukan agar tampak lebih jelas dan singkat.

Implementasi Proses *Selection Data*

Pada tahapan ini dilakukan implementasi proses seleksi data siswa SMK Tunas Karya Pekanbaru seperti analisa yang telah dilakukan sebelumnya sesuai dengan kebutuhan penelitian dengan rincian sebagai berikut:

1. Data yang digunakan adalah data siswa Kelas X SMK Tunas Karya Pekanbaru Tahun Ajaran 2015/2016 yang berjumlah 106 data.
2. Selanjutnya melakukan pemilihan terhadap variabel yang akan digunakan, yaitu jurusan, asal sekolah, alamat dan penghasilan orang tua.

Implementasi Proses *Preprocessing Data*

Pada tahapan ini dilakukan *preprocessing* data terhadap data hasil *selection* data pada tahapan sebelumnya. Pada tahapan ini dilakukan pengecekan data apakah terdapat data yang mengandung *missing value* dan *inconsistent* data. Jika ada maka dilakukan penghapusan data terhadap data-data tersebut. Daftar data yang digunakan pada penelitian ini dapat dilihat pada tabel 3.

Tabel 3. Data Penelitian

No	Jurusan	Asal Sekolah	Alamat	Penghasilan Ortu
1	TKR	SMP Tri Bhakti	Payung Sekaki	Rp. 1.500.000

2	TKR	SMP IT Aziziyyah	Payung Sekaki	Rp. 2.000.000
3	TKR	MTs. Bustanul Ulum	Tenayan Raya	Rp. 2.500.000
4	TKR	SMP Tunas Karya	Labuh Baru Timur	Rp. 2.500.000
5	TKR	SMP Santa veronika	Rumbai	Rp. 1.000.000
6	TKR	MTs. Masmur	Marpyan Damai	Rp. 3.000.000
7	TKR	SMP N 16 - Pekanbaru	Tenayan Raya	Rp. 1.500.000
8	TKR	SMP Tri Bhakti	Marpyan Damai	Rp. 800.000
9	TKR	SMP Methodist	Payung Sekaki	Rp. 2.500.000
10	TKR	SMP N 24 Pekanbaru	Rumbai	Rp. 1.000.000
....
100	TKJ	SMP Bukit Raya	Tenayan Raya	Rp. 1.000.000
101	TKJ	SMP Tuah Negeri	Payung Sekaki	Rp. 2.000.000
102	TKJ	SMP 4 Siak Hulu	Siak Hulu	Rp. 1.000.000
103	TKJ	SMP Bukit Mutiara	Tenayan Raya	Rp. 1.000.000
104	TKJ	SMP N Satap	Bukit Raya	Rp. 2.100.000
105	TKJ	SMP Muhammadiyah	Marpyan Damai	Rp. 1.500.000
106	TKJ	SMP 4 Siak Hulu	Payung Sekaki	Rp. 2.000.000

Setela ditentukan 106 data yang terpilih, langkah selanjutnya adalah melakukan inialisasi data tersebut, hasil inialisasi variabel pada data siswa SMK Tunas Karya Pekanbaru yang digunakan pada penelitian ini seperti pada tabel 4.

Tabel 4. Inialisasi Data Penelitian

No	Jurusan	Asal Sekolah	Alamat	Penghasilan Ortu
1	TKR	SS	PS	R
2	TKR	SS	PS	S
3	TKR	MT	TR	S
4	TKR	SS	LBT	S
5	TKR	SS	RB	R
6	TKR	MT	MD	S
7	TKR	SN	TR	R
8	TKR	SS	MD	R
9	TKR	SS	PS	S
10	TKR	SN	RM	R
....
100	TKJ	SS	TR	R
101	TKJ	SS	PS	S
102	TKJ	SN	SH	R
103	TKJ	SS	TR	R
104	TKJ	SS	BR	S
105	TKJ	SS	MD	R
106	TKJ	SN	PS	S

Setelah data tersebut sudah dalam bentuk inisial, kemudian data tersebut dirubah dalam format *biner* pada *database* awal yang digunakan yaitu pada *Microsoft Excel*. *Dataset* akan bernilai 1 jika pada data transaksi variabelnya terpilih dan bernilai 0 jika variabelnya tidak terpilih. Pada penelitian ini total keseluruhan variabel yang sudah diinisialkan adalah 25 variabel, yaitu TKJ, TSM, TKR, PS, MD, TR, KP, TN, RB, LBB, LBT, SH, RP, SL, TG, LP, SK, BR, R, S dan T.

Pengujian Pada Aplikasi *Rapidminer 7.0.1*

Setelah inisialisasi variabel terbentuk, kemudian rubah data tersebut dalam bentuk *biner* pada *Microsoft Excel*, selanjutnya adalah melakukan uji data pada aplikasi *Rapidminer 7.0.1*.

	A	B	C	D	E	F	G	H	I	J
1	TKR	TKJ	TSM	SS	SM	MT	PE	PS	TR	RB
2	1.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.000
3	1.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.000
4	1.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000
5	1.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000
6	1.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	1.000
7	1.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000
8	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000
9	1.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000
10	1.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.000
11	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	1.000
12	1.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000
13	1.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000
14	1.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000

Gambar 1. *Import Data*

Pada gambar 1 merupakan *import* data yang sudah disimpan dalam folder pada komputer, terlihat pada gambar tersebut bahwa data yang di *import* dalam format *biner*.

Pada pengujian 106 data dengan inputan *minimum support* 0.04 dan *minimum confidence* 0.68 maka *rule-rule* yang diperoleh adalah sebanyak 18 *rule*, dari 18 *rule* tersebut dipilih *rule* yang terkuat dalam penentuan daerah strategis, pada hasil nilai *lift ratio* > 1 maka dapat ditentukan *rule* yang *valid* seperti pada tabel 5.

Tabel 5. *Output Association Rule*

No	Jika	Maka	Confidence	Support	Lift Ratio
1	PS, TKJ	R	0.688	0.104	1.012
2	PS, SS	R	0.696	0.151	1.024
3	SS, TSM	R	0.700	0.066	1.031
4	SS, MD	R	0.700	0.066	1.031
5	TKJ, MD	R	0.714	0.047	1.052
6	SS	R	0.717	0.311	1.056
7	SN, TSM	R	0.737	0.132	1.085
8	TKJ	R	0.750	0.255	1.104
9	TSM	R	0.750	0.226	1.104
10	MT	R	0.750	0.085	1.104
11	TKJ, S	SN	0.750	0.057	1.691
12	PS, TSM	R	0.778	0.132	1.145
13	PS, SN, TSM	R	0.857	0.057	1.262
14	R, SS, TSM	PS	0.857	0.057	1.854
15	PS, SS, TKJ	R	0.875	0.066	1.288
16	SS, TKJ	R	0.882	0.142	1.299
17	SS, TSM	PS	0.900	0.085	1.947
18	PS, MT	R	1.000	0.047	1.472

Pembahasan Hasil Pengujian

Dari *rule* yang sudah dihasilkan pada implementasi pada 106 data siswa tersebut, maka *rule* yang diperoleh adalah sebanyak 18 *rule*, dari 18 *rule* tersebut dipilih *rule* yang terkuat dalam penentuan daerah strategis, Adapun hasil dari nilai *lift ratio* > 1 adalah hanya membentuk 1 *rule* yang valid dengan kepercayaan 100% . Akan tetapi untuk menentukan hasil yang sebenarnya di dalam *FP-Growth* kita bisa melihat hasil beberapa nilai dari *minimum confidence* yang paling tertinggi yang akan dijadikan acuan untuk menentukan daerah strategis yang terlihat pada tabel 6.

Tabel 6. *Strong Association Rule*

No	Jika	Maka	Confidence %	Support %	Lift Ratio
1	PS, MT	R	100%	4.7%	1.472
2	SS, TSM	PS	90%	8.5%	1.947
3	PS, SS, TKJ	R	87%	6.6%	1.288
4	R, SS, TSM	PS	85%	5.7%	1.854
5	PS, SN, TSM	R	85%	5.7%	1.262

Pada tabel 6 terdapat 5 *rule* yang memiliki tingkat kepercayaan yang paling tinggi dan didukung oleh persentase dari keseluruhan data dengan nilai *lift ratio* > 1 untuk mendapatkan *valid*-nya sebuah *rule*. Adapun penjelasan dari *strong association rule* pada tabel 6 adalah sebagai berikut:

Rule 1 : *If* dia beralamat di Payung Sekaki dan berasal dari MTs *then* penghasilan orang tuanya rendah dengan tingkat kepercayaan 100% dan didukung oleh 4.7% dari data keseluruhan dengan nilai *lift ratio* 1.472.

Rule 2 : *If* dia berasal dari SMA Swasta dan memilih jurusan TSM *then* dia beralamat di Payung Sekaki dengan tingkat kepercayaan 90% dan didukung oleh 8.5% dari data keseluruhan dengan nilai *lift ratio* 1.947.

Rule 3 : *If* dia beralamat di Payung Sekaki dan berasal dari SMA Swasta dan dia memilih jurusan TKJ *then* penghasilan orang tuanya rendah dengan tingkat kepercayaan 87% dan didukung oleh 6.6% dari data keseluruhan dengan nilai *lift ratio* 1.288.

Rule 4 : *If* penghasilan orang tuanya rendah dan berasal dari SMA Swasta dan dia memilih jurusan TSM *then* dia beralamat di Payung Sekaki dengan tingkat kepercayaan 85% dan didukung oleh 5.7% dari data keseluruhan dengan nilai *lift ratio* 1.854.

Rule 5 : *If* dia beralamat di Payung Sekaki dan berasal dari SMA Negeri dan dia memilih jurusan TSM *then* penghasilan orang tuanya rendah dengan tingkat kepercayaan 85% dan didukung oleh 5.7% dari data keseluruhan dengan nilai *lift ratio* 1.262.

IV. PENUTUP

A. Kesimpulan

Pada penelitian ini terdapat *output 18 rules*, kemudian diambil 5 *strong association rule* yang memiliki tingkat kepercayaan yang tertinggi dan didukung oleh persentase dari keseluruhan data dengan nilai *lift ratio* > 1 dari pengujian *minimum support* 0.04 dan *minimum confidence* 0.68. Dari pengujian data siswa SMK Tunas Karya Pekanbaru kelas X Tahun Ajaran 2015/2016, maka diketahui bahwa siswa yang berasal dari kecamatan Payung Sekaki adalah daerah yang paling strategis dalam pengenalan sekolah dengan tingkat kepercayaan 100% dan didukung oleh 4.7% dari data keseluruhan dengan nilai *lift ratio* 1.472.

DAFTAR PUSTAKA

- Anis Kurniawati.,(2014).“Pemetaan Pola Hubungan Program Studi Dengan Algoritma Apriori – Studi Kasus Spmu Unnes.” *Edu Komputika Journal*. 1 (1). 51-58.
- Benni R Siburian.,(2014).“Aplikasi Data Mining Untuk Menampilkan Tingkat Kelulusan Mahasiswa Dengan Algoritma Apriori.” *Pelita Informatika Budi Darma*. VII (2). 56-61.
- D. Magdalene Delighta Angeline.,(2013). “Association Rule Generation For Student Performance Analysis Using Apriori Algorithm”. *The SIJ Transactions on Computer Science Engineering & its Applications (CSEA)*. 1 (1). 12-16.
- Goldie Gunadi., dan Dana Indra Sensuse.,(2012).“ Penerapan Metode *Data Mining Market Basket Analysis* Terhadap Data Penjualan Produk Buku Dengan Menggunakan Algoritma Apriori Dan *Frequent Pattern Growth* (Fp-Growth) : Studi Kasus Percetakan PT. Gramedia.” *Jurnal TELEMATIKA MKOM*. 1 (1). 118 132.
- Jagmeet Kaur., and Neena Madan.,(2015).”Association Rule Mining: A Survey.” *International Journal of Hybrid Information Technology*. 8 (7). 239-242.
- N Punitha., and R.Amsaveni.,(2011). “Methods and Techniques to Protect the Privacy Information in Privacy Preservation Data Mining.” *Int. J. Comp. Tech. Appl.*2 (6). 2091-2097.
- Rama Novta Rinaldi., Antonius Rachmat., dan Budi Santoso.,(2014). “Implementasi Algoritma FP-GROWTH untuk Sistem Rekomendasi Buku di Perpustakaan UKDW.” *Jurnal INFORMATIKA* . 10 (1). 29-39.
- Ririanti.,(2014).”Implementasi Algoritma Fp-Growth Pada Aplikasi Prediksi Persediaan Sepeda Motor (Studi Kasus Pt. Pilar Deli Labumas). “ *Pelita Informatika Budi Darma*. VI (1). 139-144.
- Yaohui Zeng., Zijun Zhang., and Andrew Kusiak.,(2015).”Predictive modeling and optimization of a multi-zone HVAC system with data mining and firefly algorithms.” *Journal Elsevier Energy* (86). 393-402.