
**PENERAPAN DATA MINING MENGGUNAKAN ASSOCIATION RULES
UNTUK MENDUKUNG STRATEGI PEMASARAN
CALON MAHASISWA BARU
(STUDI KASUS IBI DARMAJAYA)**

Nurjoko¹, Abdi Darmawan²

^{1,2}MTI Institut Informatika dan Bisnis Darmajaya

E-mail : nurjoko.mti@gmail.com, abdi211276@gmail.com

Abstrak

. Pertumbuhan yang pesat dari akumulasi data telah menciptakan kondisi kaya akan data tapi minim informasi. Data mining merupakan penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data dalam jumlah besar yang diharapkan dapat mengatasi kondisi tersebut. Dengan memanfaatkan data calon mahasiswa dan data mahasiswa, diharapkan dapat menghasilkan informasi tentang sumber informasi yang digunakan oleh calon mahasiswa yang dapat dimanfaatkan oleh manajemen dalam melakukan usaha-usaha yang terkait dengan peningkatan jumlah mahasiswa dengan melakukan pemilihan media promosi yang tepat melalui teknik data mining.

Algoritma yang digunakan adalah algoritma apriori, informasi yang ditampilkan berupa nilai support dan confidence dari masing-masing pola yang terbentuk. Dari uji coba terhadap data calon mahasiswa dan data mahasiswa dengan model Association rules dihasilkan beberapa aturan dengan tingkat nilai maksimum confidence sebesar 81,1% Untuk sumber informasi yang berasal dari media Koran dan tingkat nilai maksimum confidence sebesar 77,8% untuk sumber informasi melalui media brosur. Saat ini media promosi yang tepat untuk digunakan oleh TIMP emasaran dalam menjaring calon mahasiswa baru adalah media Koran dan media brosur dengan tidak mengabaikan media promosi yang lain.

Kata kunci: *data mining, Association rules, algoritma apriori, confidence, support*

Abstract

The rapid growth of data accumulation has created the conditions for data-rich but information poor. Data mining is the mining or discovery of new information by looking for certain patterns or rules of a number of large amounts of data are expected to treat the condition. By utilizing the data of prospective students and student data, is expected to yield information about the resources used by the candidate that can be used by management in making efforts associated with an increase in the number of students by choosing the appropriate media campaign through data mining techniques.

The algorithm used is the a priori algorithm, the information displayed in the form of support and confidence value of each pattern is formed. Of tests on the data of prospective students and student data with the model Association rules generated some rules with a maximum value of confidence level of 81.1% for resources from newspapers and media level maximum value of 77.8% confidence sebesarb to resources through the media brosur. Saat ini right promotional media for use by TIMP emasaran in recruiting new students are media newspapers and brochures with not ignore other promotional media.

Keywords: *data mining, Association rules, a priori algorithm, confidence, support*

1. Pendahuluan

Kemampuan teknologi informasi untuk mengumpulkan dan menyimpan berbagai tipe data telah jauh meninggalkan kemampuan untuk menganalisis, meringkas dan mengekstraksi “pengetahuan” dari data. Sementara para pelaku bisnis memiliki kebutuhan-kebutuhan untuk memanfaatkan “gudang data” yang sudah dimilikinya, para peneliti melihat peluang itu untuk melahirkan sebuah teknologi baru yang menjawab kebutuhan ini, yaitu Data Mining.

Kumpulan dari data yang tersimpan selama bertahun-tahun dengan jumlah yang besar akan sangat penting dan bermanfaat serta tidak menjadi sebuah kuburan data apabila kita mampu menggali informasi-informasi yang tersembunyi yang terdapat didalamnya, hal tersebut sangat berguna dan penting sebagai informasi yang mendukung atau menjadi pedoman dan dasar dalam pengambilan sebuah keputusan.

Keputusan seringkali dibuat tidak berdasarkan pada data-data yang ada yang tersimpan didalam kumpulan atau tumpukan data tersebut tetapi keputusan hanya berdasarkan pada intuisi dari pengambil keputusan. Hal ini disebabkan belum adanya sistem atau perangkat lunak yang dapat membantu dalam pencarian informasi yang tepat, cepat, dan akurat. Pemanfaatan teknologi informasi dalam menjalankan proses bisnis di IBI darmajaya menyebabkan akumulasi atau bertambahnya data dalam jumlah yang besar setiap tahunnya.

Diantara data data yang mengalami peningkatan jumlahnya setiap tahun adalah data pendaftaran calon mahasiswa baru, data mahasiswa, data nilai, dan data alumni. Data akan terus menumpuk dan menjadi kuburan data serta tidak memiliki nilai atau informasi yang dapat dihasilkan dari data tersebut, oleh sebab itu diperlukan suatu teknik atau perangkat untuk membantu dalam mentransformasikan data dengan jumlah yang besar tersebut menjadi informasi yang berguna yaitu dengan penerapan data mining untuk strategi promosi penerimaan calon mahasiswa baru yang dapat menjadi dasar atau pedoman dalam mendukung pengambilan sebuah keputusan, terutama keputusan tentang strategi promosi dalam penerimaan calon mahasiswa baru.

Perubahan-perubahan pola pendidikan yang begitu cepat, silih berganti serta globalisasi di segala bidang termasuk bidang pendidikan, memunculkan persaingan yang sangat ketat di bidang bisnis jasa pendidikan.

IBI Darmajaya sebagai salah satu perguruan tinggi swasta yang cukup besar dan terkenal yang ada di Bandar Lampung harus selalu siap dengan perubahan-perubahan yang dikeluarkan pemerintah di bidang pendidikan, juga harus siap bersaing tidak hanya dengan perguruan tinggi sejenis tetapi juga dengan perguruan tinggi yang lain baik yang ada di Bandar Lampung atau pun diluar Propinsi Lampung. Agar dapat memenangkan persaingan khususnya dalam meningkatkan jumlah mahasiswa baru yang akan mendaftar, maka diperlukan sebuah strategi pemasaran.

IBI Darmajaya telah melakukan berbagai strategi pemasaran baik melalui media elektronik maupun dengan media cetak, seperti membagikan brosur kepada calon mahasiswa, iklan di surat kabar dan pemasangan spanduk di tempat yang dianggap strategis. Kondisi yang terjadi saat ini jumlah mahasiswa yang diterima atau mendaftar belum mampu memenuhi target sesuai dengan yang telah ditetapkan. Hal ini menyebabkan tim promosi perguruan tinggi harus bekerja keras untuk bias mendapatkan mahasiswa sesuai dengan target yang telah ditetapkan oleh perguruan tinggi.

Dengan terbatasnya data dan informasi yang dimiliki tim promosi terhadap profil mahasiswa yang sudah ada, maka tim promosi melakukan promosi secara masal, tanpa melihat segmen pasar yang harus dituju. Konsekuensinya adalah biaya promosi menjadi lebih mahal dan tingkat respon yang rendah.

Teknik data mining dengan model *Association rules* (aturan asosiasi) atau *affinity analysis* (analisis afinitas) yang digunakan dalam penelitian ini menerapkan algoritma *apriori* sebagai pemrosesan data. Ide dasar dari algoritma ini adalah dengan mengembangkan *frequent itemset*. Dengan menggunakan satu item dan secara rekursif mengembangkan *frequent itemset* dengan dua item, tiga item dan seterusnya hingga *frequent itemset* dengan semua ukuran. Untuk mengembangkan *frequent set* dengan dua item, dapat menggunakan *frequent set item*. Alasannya adalah bila set satu item tidak melebihi *isupport minimum*, maka sembarang ukuran item set yang lebih besar tidak akan melebihi *support minimum* tersebut. Secara umum, mengembangkan set dengan k -item menggunakan frequent set dengan $k - 1$ item yang dikembangkan dalam langkah sebelumnya. Setiap langkah memerlukan sekali pemeriksaan keseluruhan isi *database* (Santoso, 2007).

Algoritma *apriori* adalah algoritma paling terkenal untuk menemukan pola frekuensi tinggi. Pola frekuensi tinggi adalah pola-pola item di dalam suatu *database* yang memiliki frekuensi atau *support* di atas ambang batas tertentu yang disebut dengan istilah minimum *support* atau *threshold*. *Threshold* adalah batas minimum transaksi. Jika jumlah transaksi kurang dari *threshold* maka item atau kombinasi item tidak akan diikuti perhitungan selanjutnya. Penggunaan *threshold* dapat mempercepat perhitungan (Santoso, 2007).

2. METODOLOGI PENELITIAN

2.1 Metode Pengumpulan Data

Agar diperoleh data-data yang relevan dan lengkap, dalam penelitian ini akan digunakan metode pengumpulan data sebagai berikut :

- a. **Mengadakan interview** yaitu dengan mengadakan wawancara kepada orang-orang yang berkepentingan atau berkompeten dalam bidang akademik dan pemasaran
- b. **Dokumentasi**, dengan melakukan Studi Pustaka diharapkan dapat mengumpulkan data melalui buku-buku literatur yang berhubungan dengan penelitian yang sedang dilakukan, serta studi pustaka dapat dihasilkan dari browsing di situs-situs internet.

Tahapan Pelaksanaan Penelitian

2.2 Analisis Kebutuhan Bisnis dan Informasi

Analisis kebutuhan bisnis dan informasi perlu dilakukan guna menyelaraskan penelitian yang dilakukan dengan melakukan proses evaluasi dan perencanaan strategi pemasaran penerimaan mahasiswa baru, pimpinan IBI Darmajaya membutuhkan informasi tentang calon mahasiswa, asal daerah, asal sekolah dan data media promosi yang digunakan sebagai sumber informasi untuk menentukan strategi pemasaran mahasiswa baru yang potensial bagi IBI Darmajaya.

2.3 Analisis Kebutuhan Informasi Data Mining

Data calon mahasiswa yang akan dicari hubungannya meliputi data asal sekolah, asal daerah, program studi pilihan dan data sumber informasi. Adapun yang akan diproses *mining* adalah : Hubungan Sumber Informasi dengan program studi, asal sekolah, dan asal daerah. Dari atribut asal daerah, program studi pilihan, dan asal sekolah, akan dicari hubungan dengan sumber informasi, dengan harapan dapat mengetahui media promosi yang tepat untuk digunakan dalam promosi pemasaran bagi calon mahasiswa baru berdasarkan pilihan program studi dan asal sekolah yang berada di daerah tertentu sehingga strategi promosi pemasaran bagi calon mahasiswa baru dapat ditujukan langsung

ke sekolah-sekolah di daerah tersebut dengan media promosi yang memiliki tingkat respon yang tinggi

2.4 Sumber Data dan Data yang digunakan

2.4.1 Data calon mahasiswa

Data calon mahasiswa yang digunakan dalam penelitian ini adalah data calon mahasiswa mulai angkatan 2003 sampai dengan 2010. Atribut yang akan digunakan meliputi : atribut Nomor Test, nama calon mahasiswa, atribut asal sekolah, atribut asal daerah,

2.4.2 Data mahasiswa

Atribut yang akan digunakan meliputi : NPM atau Nomor Pokok Mahasiswa yang merupakan Primary Key yang menghubungkan data mahasiswa dengan data calon mahasiswa, atribut nama mahasiswa, atribut program studi yang dipilih, atribut status daftar ulang dan atribut angkatan

2.4.3 Penggabungan Data (*Data Integration*)

Penggabungan data atau data integrasion merupakan proses menyatukan data-data dari sumber data kedalam satu database baru. *Database* calon mahasiswa dan *database* mahasiswa serta *Database* lulusan atau alumni merupakan *data source* yang akan digabung dalam pembangunan *database* baru dengan nama *database* gabungan. melalui proses ETL ketiga *data source* tersebut di *load* kedalam *database* gabungan dengan key NPM.

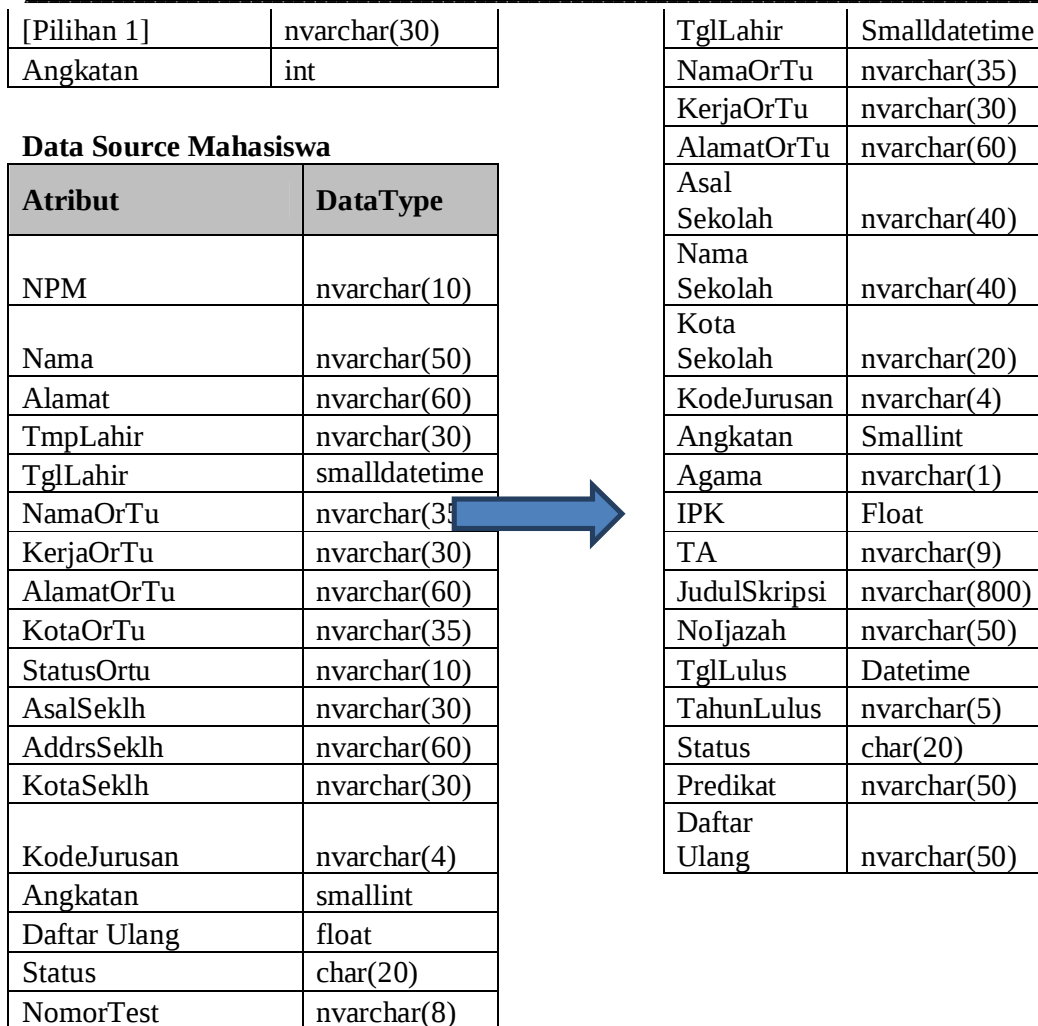
Data Gabungan dari Data calon Mahasiswa, data mahasiswa dan data alumni atau lulusan = { NPM + Nama + Alamat + Tempat Lahir + Tanggal lahir + Agama + Nama Orangtua + Pekerjaan Orangtua + alamat Orangtua + nama asal sekolah + Kota asal sekolah + asal sekolah + Kode jurusan + Angkatan + StatusDaftar ulang}. Proses *load* data dan integrasi dapat dilihat dalam gambar 3.2

Data Source Calon mahasiswa

Atribut	Data Type
NomorTest	nvarchar(8)
[Nama Siswa]	nvarchar(30)
[Jenis Kelamin]	nvarchar(15)
[Tempat Lahir]	nvarchar(20)
[Tanggal Lahir]	smalldatetime
Agama	nvarchar(10)
Status	nvarchar(15)
[Asal Sekolah]	nvarchar(40)
[Nama Sekolah]	nvarchar(40)
[Kota Sekolah]	nvarchar(20)
[Nama Ortu]	nvarchar(30)
[Alamat Ortu]	nvarchar(50)
[Kota Ortu]	nvarchar(20)
[Pekerjaan Ortu]	nvarchar(15)
KodeJurusan1	nvarchar(4)

Database Gabungan

Atribut	Data Type
NPM	nvarchar(10)
Nama	nvarchar(50)
Alamat	nvarchar(60)
TmpLahir	nvarchar(30)



Gambar tabel3.1 Tabel Integrasi Data

Semua data yang ada pada *data source* diimport ke dalam *database* gabungan, kecuali jika ada data yang sama dari dua *data source* seperti, nama, alamat orang tua dan yang lain, maka yang di *load* hanya salah satu dari *data source*, dengan asumsi memiliki nilai yang sama. Adapun atribut tabel yang terbentuk dapat dilihat pada tabel berikut :

Database Gabungan

Atribut	Data Type
NPM	nvarchar(10)
Nama	nvarchar(50)
Alamat	nvarchar(60)
Tempat Lahir	nvarchar(30)
TanggalLahir	Smalldatetime
NamaOrangTua	nvarchar(35)
Daftar Ulang	nvarchar(30)
Sumber Informasi	nvarchar(60)
Asal Sekolah	nvarchar(40)

Nama Sekolah	nvarchar(40)
Kota Sekolah	nvarchar(20)
Prodi Pilihan	nvarchar(4)
Angkatan	Smallint
Agama	nvarchar(1)
IPK	Float
Predikat Kelulusan	nvarchar(50)
Status	char(20)
JudulSkripsi	nvarchar(800)
NoIjazah	nvarchar(50)
TglLulus	Datetime
NomorTest	nvarchar(8)
NoSK	nvarchar(50)

Gambar Tabel 3.2 Tabel Database Gabungan

2.5 Proses ETL

Setelah proses *cleansing data* selesai dilakukan kemudian dilanjutkan dengan proses transformasi untuk mengubah format data operasional ke dalam format *data mining*. Transformasi data merupakan proses mengubah data atau digabung ke dalam format yang sesuai untuk diproses dalam *data mining*.

Beberapa metode *data mining* membutuhkan format data yang khusus sebelum bisa digunakan untuk proses *Data Mining*, data-data yang dikategorikan meliputi:
Atribut program studi diambil dari kode jurusan yang terdiri dari 8 (delapan) program studi berdasarkan program studi yang ada di IBI Darmajaya

Atribut	Value Atribut	Kategori
Kodejurusan	0101	S1 Teknik Informatika
	0102	D3 Teknik Komputer
	0103	S1 Manajemen Informatika
	0105	S1 Sistem Informasi
	0106	S1 Sistem Komputer
	0211	S1 Manajemen
	0212	S1 Akuntansi S1
	0213	D3 Akuntansi D3

Gambar Tabel 3.3 Tabel Pengkategorian Program Studi

Atribut Asal Sekolah di dapatkan dari *Asalsklh* dan Nama Sekolah yang di kategorikan menjadi 2 (dua) kategori yaitu Asal Sekolah dari SMU dan SMK dengan asumsi bahwa sekolah yang berasal dari pendidikan UMUM dan sekolah yang berasal dari pendidikan khusus atau kejuruan, adapun pengkategorian asal sekolah dapat dilihat pada table 3.4 Berikut ini :

Nama Atribut	Value Atribut	Kategori
AsalSklh	SMU	SMU/MAN
	SMA	
	MA	

AsalSkh	STM SMK SEMEA	SMK
---------	---------------------	-----

Gambar Tabel3.4 Tabel Pengkatagorian Asal Sekolah

Atribut sumber informasi (Brosur, Koran, Pamflet, Program SGS, Radio, Bioskop) yang di kategorikan menjadi 2 (dua) kategori yaitu True (1) dan False (0)Berikut ini :

Atribut	Value Atribut	Kategori
Sumber Informasi	True (1) atau False (0)	Brosur, Koran, Pamflet, Program SGS, Radio, Bioskop

Gambar Tabel 3.5 Tabel Pengkatagorian Sumber Informasi

PenerapanModelAssosiacion Rule denganAlgoritmaApriori

Proses *mining* untuk mengetahui tingkat keeratan hubungan antarasumber informasi dengan program studi asal sekolah, dan asal daerah,dapat diketahui dari prosentase nilai *confidence* dan *support*dari masing-masing kombinasi atribut tersebut denganrumus sebagai berikut :

Support (asal daerah, asal sekolahdan Program studidengan Sumber Informasi)

$$S = \Sigma (Ta + Tc) / \Sigma(T)$$

Keterangan :

S = Support

$\Sigma (Ta + Tc)$ = Jumlah transaksi yang mengandung *antecedent* dan *consequenc*

$\Sigma(T)$ = Jumlah transaksi

$$S = \Sigma (Ta + Tc) / \Sigma(T)$$

Support (Sumber Informasi dengan program studi, asal sekolah, asal daerah,) = **Count** {(program studi,asal sekolah, asal daerah) + (Sumber informasi)} / **jumlah transaksi**

$\Sigma (Ta + Tc)$ =

Antecedent(mewakili “ jika”) = {program studi (TI,SI,MI,SK,TK,MA,AK, AD), asal sekolah (SMU, SMK, Konversi)asal daerah (Bandar Lampung, Metro..dst)}

Consequenc (mewakili “maka”)= {Sumber Informasi (Brosur, Koran, Pamflet, Program SGS, Radio, Bioskop)}

$\Sigma(T)$ = **Jumlah transaksi** = Jumlah semua transaksi itemset

Confidence (Sumber Informasi dengan program studi asal sekolah, dan asal daerah)

$$C = \Sigma (Ta + Tc) / \Sigma(Ta)$$

Keterangan :

C = Confidence

$\Sigma (Ta + Tc)$ = Jumlah transaksi yang mengandung *antecedent* dan *consequenc*

$\Sigma(Ta)$ = Jumlah transaksi yang mengandung *antecedent*

$$C = \Sigma (Ta + Tc) / \Sigma(Ta)$$

Confidence (program studi dengan asal sekolah, asal daerah, dan pekerjaan orang tua) = **Count** (program studi,asal sekolah, asal daerah) + (Sumber informasi)}/**Count**((program studi,asal sekolah, asal daerah)

$\Sigma(Ta+Tc) =$

Antecedent(mewakili “ jika”) = {program studi (TI,SI,MI,SK,TK,MA,AK, AD), asal sekolah (SMU, SMK, Konversi) asal daerah (Bandar Lampung, Metro..dst)}

Consequencet (mewakili “maka”)= {Sumber Informasi (Brosur, Koran, Pamflet, Program SGS, Radio, Bioskop)}

$\Sigma(Ta)$ = Jumlah transaksi yang mengandung *antecedent*= Jumlah transaksi{program studi (TI,SI,MI,SK,TK,MA,AK, AD), asal sekolah (SMU, SMK, Konversi) asal daerah (Bandar Lampung, Metro..dst)}

Pemodelan Data Mining

Pemodelan *data mining* dirancang berdasarkan kebutuhan fisik dan karakteristik data yang ada. *Dataset* yang akan digunakan adalah data calon mahasiswa yang berstatus telah melakukan pembayaran daftar ulang.

No	ItemSet	Type	Input	Predict
1	Nomortest	Discrete	Primary Key	
2	Asal Daerah	Discrete	√	
3	Asal Sekolah	Discrete	√	
4	Pilihan Program studi	Discrete	√	
5	Sumber Informasi	Discrete		√

Tabel 3.6 Tabel Model Struktur

Model mining ini digunakan untuk mengetahui *structure* hubungan (*asosiasi*) antara asal daerah dan asal sekolah dengan sumber informasi. Model aturan (*rule*) yang dibentuk nantinya adalah: **if [asal daerah] dan [asal sekolah] then [sumber informasi]**

3. HASIL DAN PEMBAHASAN

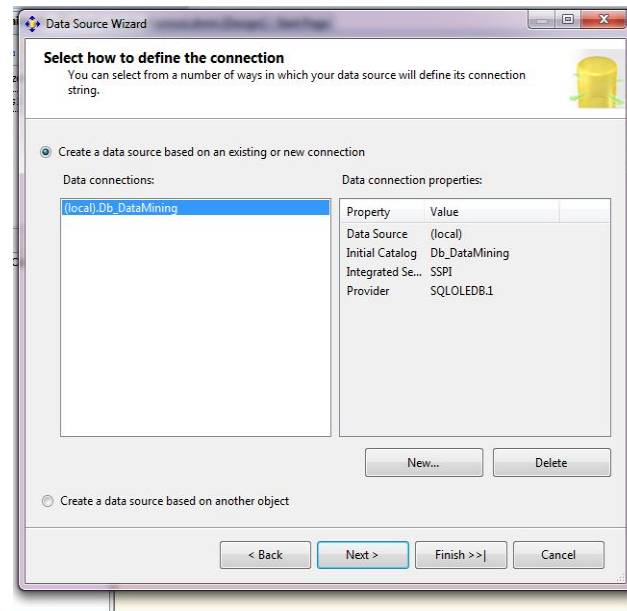
3.1 Hasil Penelitian

Hasil dari penelitian ini diperoleh sebuah project berupamodel data *mining association rule* yang membentuk pola-pola karakteristik hubungan (*asosiasi*) antara asal daerah dan asal sekolah dengan sumber informasi yang digunakan oleh calon mahasiswa yaitu media promosi yang berupa media Brosur, spanduk, iklan koran, radio, pamflet, dan program *Student Get Student* seperti yang terlihat pada gambar 4.5. Pola-pola yang terbentuk dihasilkan dari proses mining.

3.2 Proses Mining

Beberapa tahapan yang perlu dilakukan untuk membuat sebuah *Analysis project Data Mining* yaitu:

Membuat Koneksi Ke Sumber Data, koneksi digunakan untuk mengetahui sumber database yang ada di server. Informasi koneksi ini nantinya disimpan di dalam *Data Source Designer*. Koneksi yang terbentuk adalah *Data Connection : (local) Db_DataMining* seperti pada gambar 4.1 berikut ini :



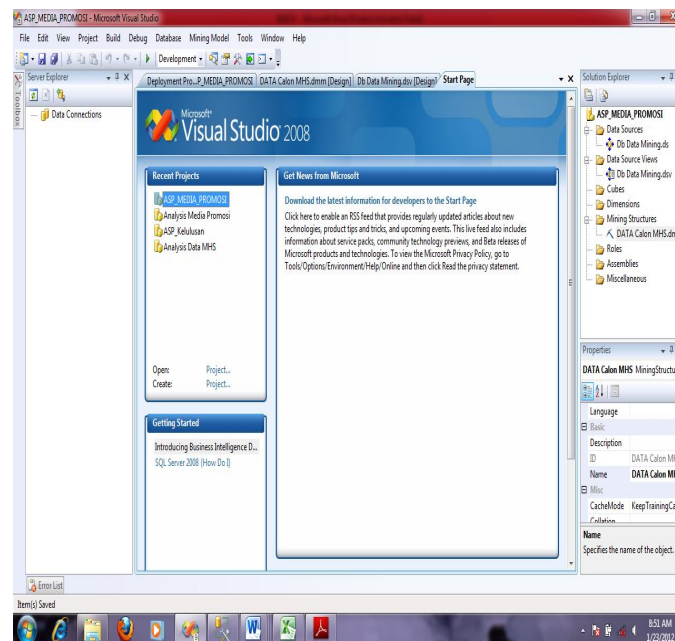
Gambar 4.1 Model Mining Data Connection

Membuat *Data Source View (DSV)*

Data Source berisi Nama Server, Nama Database dimana data berada serta beberapa hal lainnya yang dibutuhkan. Informasi dalam *Data Source View (DSV)* berisi tabel apa yang digunakan dalam proses mining. Nama server (local), Nama Database *Db_DataMining.dsv*

Membuat *Mining Structures*

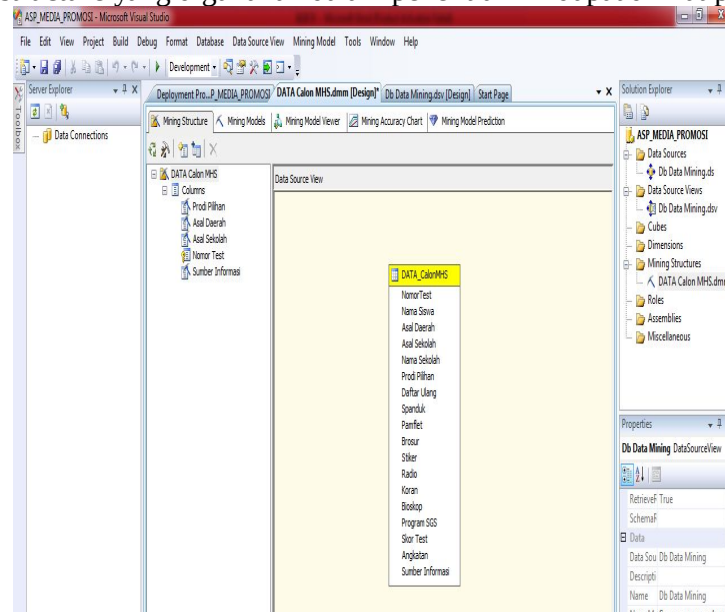
Tahapan ini berisi informasi mengenai tabel yang digunakan dalam proses mining berikut kolom-kolom input dan kolom prediksi seperti pada gambar 4.3



Gambar 4.3 Model Mining ASP_Media_Promosi

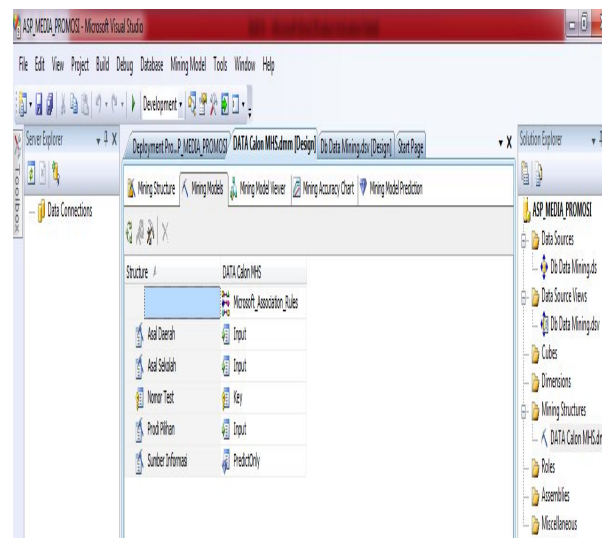
3.3 Pembahasan

Model mining structure yang digunakan dalam penelitian ini dapat dilihat pada gambar 4.4



Gambar 4.3 Model Mining Structure

Model mining structure yang dibentuk, kemudian diperoleh mining model seperti dilihat pada gambar 4.4.

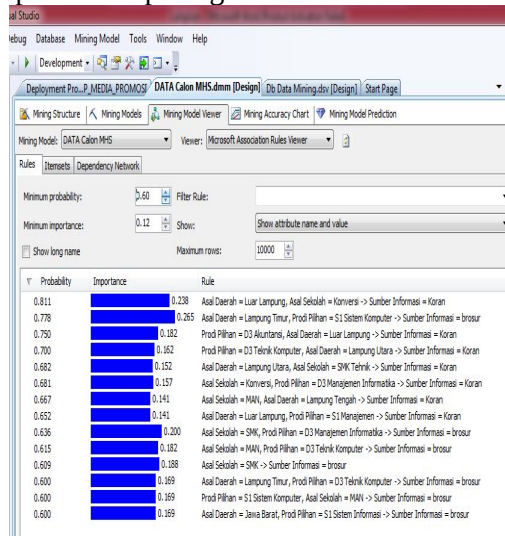


Gambar 4.4 Model Mining asosiasi rules

Dalam proses mining terdapat sebuah kolom yang berfungsi sebagai primary key. Selanjutnya ditentukan kolom mana saja yang berfungsi sebagai *input* dan *predict*. Kolom *predict* ini berfungsi sebagai tujuan (*goal*) dari algoritma *asosiasi rules*.

Uji coba yang dilakukan menggunakan *Microsoft SQL Server 2008*, *SQL Server Business Intelligence Development Studio*. Batas (*threshold*) yang digunakan untuk *minimum support* 20, *minimum probability* 50% dan *minimum importance* 10%. Jika persentase kombinasi *item* dalam *dataset* lebih dari 20 dengan *probability* lebih dari 50%

serta *importance* lebih dari 10% maka akan dibuatkan *rule* yang berkaitan dengan kombinasi *item* tersebut, tetapi jika tidak maka akan diabaikan. Jumlah data alumni yang digunakan dalam uji coba ini adalah sebanyak 9410 *record*, Hasil uji coba yang dilakukan pada penelitian ini diperoleh beberapa *rule* yang sesuai dengan kriteria yang ditentukan. Hasil uji coba tersebut dapat dilihat pada gambar 4.5



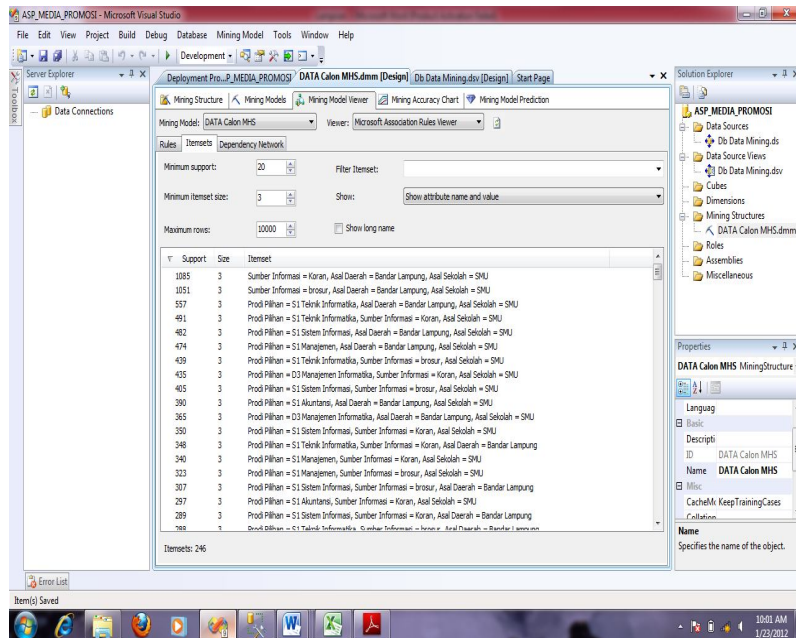
Gambar 4.5 Hasil Uji Coba Mining Sumber Informasi (Rules Viewer)

Sebagai contoh terbentuk rules :

Asal Daerah=Luar Lampung, dan asal sekolah=Konversi → Sumber Informasi=Koran

Aturan ini memberi arti bahwa kemungkinan mahasiswa yang berasal dari daerah luar lampung dan asal sekolah konversi maka sumber informasi yang digunakan pada saat mendaftar di IBI Darmajaya adalah media Koran. Hal ini diketahui dengan kemungkinan atau tingkat *confidence* yang diwakili oleh nilai *Probability* 81%. Hanya aturan-aturan yang nilai *probability* di atas 0.50 (sesuai dengan nilai *probability* yang ditentukan) yang akan ditampilkan. Demikian juga dengan nilai *Minimum Importance*, hanya rules yang memiliki nilai di atas nilai *minimum importance* yang ditampilkan. Aturan-aturan yang muncul tersebut dapat digunakan untuk mendukung penentuan media promosi yang akan digunakan sebagai strategi pemasaran.

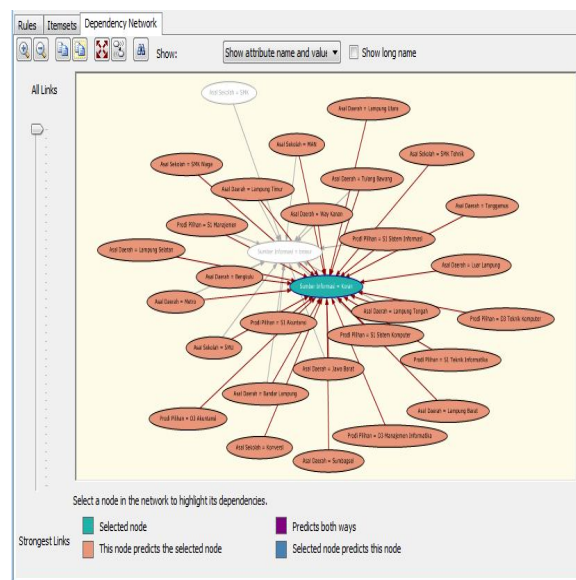
Selanjutnya untuk mengetahui seberapa besar tingkat dukungan item atau kolom-kolom yang ada dalam sebuah *frequent item*, dapat digunakan *tab Itemset* seperti dapat dilihat pada gambar 4.6. Dalam gambar ini setiap itemset memiliki size atau ukuran. Size mengindikasikan seberapa banyak item yang ada dalam sebuah *itemset*. Nilai *minimum support* digunakan untuk membatasi itemset bukan untuk membatasi jumlah *rules*. Sedangkan *minimum probability* tidak berdampak pada *itemset* melainkan berdampak pada jumlah *rules*.



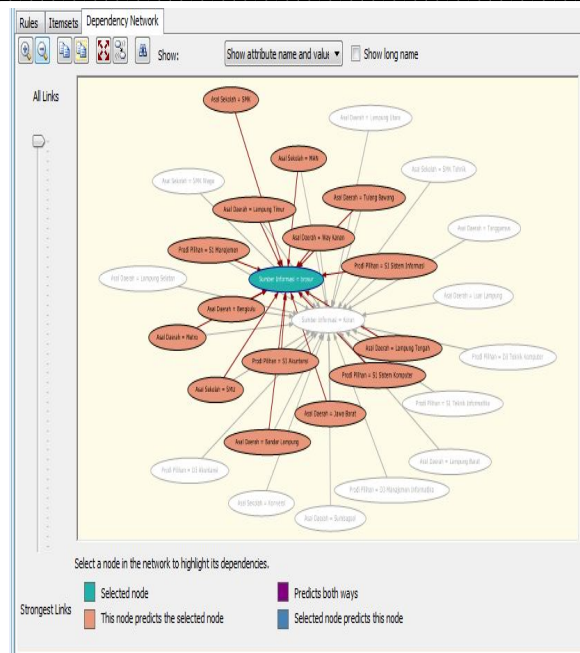
Gambar 4.6 Support Itemset

Gambar 4.6 menunjukkan bahwa jumlah data yang mengindikasikan mahasiswa yang berasal dari: asal sekolah=SMU dan dengan asal daerah Bandar lampung menggunakan sumber informasi=Koran memiliki dukungan data sebanyak 1085 dari total data calon mahasiswa yang ada. Nilai support menandakan seberapa banyak item data tersebut berada dalam keseluruhan total transaksi.

Kemudian untuk mengetahui hubungan atribut apa saja yang ada dalam *antecedent* berikut nilainya yang akan menghasilkan *consequent*, dapat dilihat dalam *Dependence Network* seperti pada gambar 4.7.



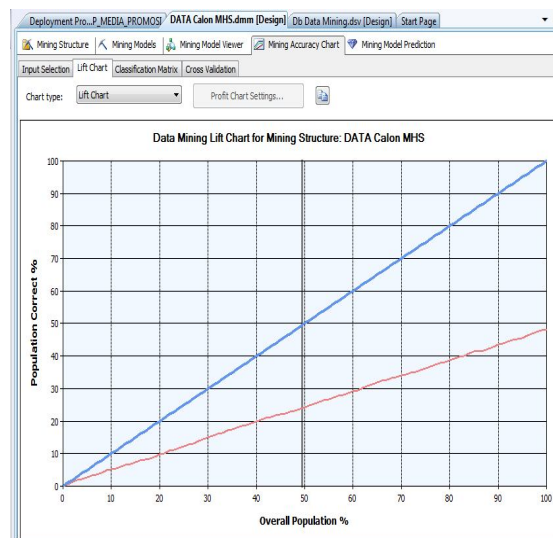
Gambar 4.7 Dependency Network Sumber Informasi=Koran



Gambar 4.8 *Dependence Network* Sumber Informasi=Brosur

Node tujuan (*goal*) akan berisi sejumlah nilai pada atribut tujuan yang telah ditentukan. Misalnya dalam atribut predict ada 1 kolom dan berisi 3 buah data, maka jumlah *node goal* akan berisi 3 buah. Masing-masing *node* akan terhubung dengan atribut yang ditentukan sebagai kolom input pada *mining structure*. Masing-masing jalur mewakili pasangan asosiasi dari sebuah aturan. Semakin rendah nilai minimum *probability* maka jumlah *node* dan jalur akan semakin banyak.

Cara untuk mengetahui seberapa besar tingkat akurasi data jika aturan-aturan yang telah terbentuk yang digunakan untuk melakukan prediksi dan mewakili seluruh populasi dapat dilihat pada *tab lift chart* seperti terlihat pada gambar 4.9



Gambar 4.9 *Lift Chart*

Gambar 4.9 menunjukkan bahwa garis lurus (garis diagonal) menunjukkan skenario yang ideal di mana prediksi selalu benar. Dalam penelitian ini yang digunakan sebagai

prediksi adalah sumber informasi apakah yang digunakan oleh calon mahasiswa pada saat mendaftar di IBI Darmajaya yang diwakili oleh sumbu y. Garis merah yang berada dibawah garis ideal menunjukkan model data *association rules*. Sumbu x adalah jumlah populasi atau data yang digunakan sebanyak 100 persen populasi, dan sumbu y adalah data model *association rules* sebesar 48 persen benar.

Model *association rule* kebanyakan diterapkan pada data transaksi dengan atribut biner yang disebut item-item (*itemset*). Namun dalam kebanyakan dunia bisnis mempunyai tipe atribut yang lebih beragam. Atribut tersebut biasa berbentuk kuantitatif atau kategori (seperti: jenis kelamin). Dengan nilai *support* dan *confidence* yang telah ditentukan, akan dihasilkan rule dengan nilai *confidence* yang bervariasi. *Support* dari suatu *itemset* merupakan prosentase dari himpunan semua data.

Rule yang terbentuk dari item yang ada mempunyai nilai *confidence* yang berbeda. *Confidence* merupakan tingkat kepercayaan atau tingkat kebenaran dari rule yang terbentuk. Rule yang dicari adalah yang mempunyai kemungkinan benar yang besar, yaitu rule yang membuat prediksi yang benar (atau sangat sering benar). Dari tabel diatas dapat diketahui, bahwa jika dengan menggunakan nilai maximum *support* lebih besar, maka nilai *confidence* yang diperoleh juga lebih besar.

Berdasarkan hasil uji coba proses *mining* yang dilakukan dihasilkan sebanyak 153 aturan dengan minimum *probability* sebesar 40% dan maksimum *probability* sebesar 81%. Adapun dari 153 aturan yang terbentuk diperoleh 14 aturan yang terbaik dengan menerapkan minimum *probability* sebesar 60%. Dari aturan yang didapat diketahui karakteristik sumber informasi yang digunakan oleh calon mahasiswa untuk mendaftar di IBI Darmajaya adalah sebagai berikut :

No	Probability	Rules
1	0.811	Asal Daerah = Luar Lampung, Asal Sekolah = Konversi -> Sumber Informasi = Koran
2	0.778	Asal Daerah = Lampung Timur, Prodi Pilihan = S1 Sistem Komputer -> Sumber Informasi = brosur
3	0.75	Prodi Pilihan = D3 Akuntansi, Asal Daerah = Luar Lampung -> Sumber Informasi = Koran
4	0.7	Prodi Pilihan = D3 Teknik Komputer, Asal Daerah = Lampung Utara -> Sumber Informasi = Koran
5	0.682	Asal Daerah = Lampung Utara, Asal Sekolah = SMK Teknik -> Sumber Informasi = Koran
6	0.681	Asal Sekolah = Konversi, Prodi Pilihan = D3 Manajemen Informatika -> Sumber Informasi = Koran
7	0.667	Asal Sekolah = MAN, Asal Daerah = Lampung Tengah -> Sumber Informasi = Koran
8	0.652	Asal Daerah = Luar Lampung, Prodi Pilihan = S1 Manajemen -> Sumber Informasi = Koran
9	0.636	Asal Sekolah = SMK, Prodi Pilihan = D3 Manajemen Informatika -> Sumber Informasi = brosur
10	0.615	Asal Sekolah = MAN, Prodi Pilihan = D3 Teknik Komputer -> Sumber Informasi = brosur
11	0.609	Asal Sekolah = SMK -> Sumber Informasi = brosur

12	0.6	Asal Daerah = Lampung Timur, Prodi Pilihan = D3 Teknik Komputer -> Sumber Informasi = brosur
13	0.6	Prodi Pilihan = S1 Sistem Komputer, Asal Sekolah = MAN -> Sumber Informasi = brosur
14	0.6	Asal Daerah = Jawa Barat, Prodi Pilihan = S1 Sistem Informasi -> Sumber Informasi = brosur

Tabel 4.1 Tabel Tingkat *Probability* model *association rule*

Sumber informasi yang paling banyak digunakan oleh calon mahasiswa adalah media Koran dengan karakteristik mahasiswa yang berasal dari daerah luar Lampung dan asal sekolah konversi dengan tingkat *Probability* sebesar 81%. Sedangkan sumber informasi brosur dengan karakteristik mahasiswa yang berasal dari Daerah = Lampung Timur, dan Program studi Pilihannya adalah S1 Sistem Komputer dengan tingkat *Probability* sebesar 77%,. Hal ini membuktikan bahwa aturan asosiasi yang diperoleh mewakili sumber informasi yang digunakan oleh mahasiswa untuk mendaftar di IBI Darmajaya adalah media Koran dan brosur, sedangkan pamlet, stiker, radio dan program SGS tidak terbentuk. Hal tersebut dikarenakan jumlah data calon mahasiswa yang menggunakan sumber informasi tersebut tidak mencukupi dari batas minimum (*threshold*) *Probability* yang telah ditentukan.

KESIMPULAN DAN SARAN

4. Kesimpulan dan Saran

4.1. Kesimpulan

Penerapan *data mining* dapat dimanfaatkan untuk mencari pola karakteristik sumber informasi yang digunakan oleh mahasiswa yang mendaftar di IBI Darmajaya. Pola karakteristik yang ditampilkan berupa nilai *support* dan *confidence* hubungan antara sumber informasi dengan data mahasiswa. Semakin tinggi nilai *confidence* dan *support* maka semakin kuat nilai hubungan antar atribut. Data mahasiswa yang diproses mining meliputi data asal daerah, asal sekolah, dan program studi yang dipilih oleh mahasiswa.

Hasil dari proses data mining ini dapat digunakan sebagai pertimbangan dalam mengambil keputusan lebih lanjut tentang sumber informasi sebagai media promosi yang tepat untuk mendukung strategi pemasaran di IBI Darmajaya. Untuk saat ini media promosi yang tepat untuk digunakan oleh tim pemasaran dalam menjaring calon mahasiswa baru adalah media Koran dan media brosur dan tidak mengabaikan media promosi yang lain.

Saran

Berdasarkan kesimpulan dari penelitian yang telah peneliti lakukan disarankan bahwa media promosi Koran dan brosur untuk lebih intensif digunakan dalam mendukung strategi pemasaran dengan membuat iklan yang menarik dan diterbitkan oleh media cetak yang menjangkau semua daerah (provinsi Lampung). Selain itu perlu diadakan penelitian lebih lanjut tentang data yang digunakan dalam prediksi sumber informasi yang digunakan oleh calon mahasiswa.

Parameter yang digunakan lebih spesifik, jadi data atau informasi yang dihasilkan lebih dapat digunakan sebagai alat pendukung bagi pihak yang bersangkutan untuk mengambil kebijakan dan sebagai sumber informasi yang lebih akurat untuk menemukan ide-ide baru yang lebih baik.

Referensi

- [1] Bounds, Gregory. 1994. *Management: A Total Quality Perspective*, South Western College Publishing, Ohio.
- [2] McLeod, Raymon. 1998. *Management Information Systems*. 7th Edition, New Jersey : Prentice Hall, Inc.
- [3] Respati, Bagus. 2006. *Sistem Pendukung Keputusan dengan Expert Choice*. Diklat Tidak Terpublikasi. Yogyakarta: Universitas Gadjah Mada.
- [4] Turban, Efraim; Aronson, Jay; Liang Peng Ting. 2005. *Decision Support Systems and Intellegent Systems*. New Jersey : Pearson Education, Inc.
- [5] Turban; McLean; Wetherbe. 1999. *Information Technology For Management*. John Wiley & Sons, Inc. USA.