

ANALISIS ASOSIASI UNTUK MENENTUKAN STRATEGI PROMOSI PERGURUAN TINGGI DENGAN ALGORITMA APRIORI

Erna Hudianti Pujiarini

Jurusan Teknik Informatika, STMIK AKAKOM Yogyakarta
Jl.Raya Janti 143 Karangjambe Yogyakarta.
e-mail: ernahudi@akakom.ac.id

ABSTRAK

Setiap organisasi perlu untuk menentukan strategi promosi untuk mengenalkan perusahaan atau organisasinya. Penentuan strategi promosi yang tepat akan dapat mengurangi biaya promosi dan mencapai sasaran promosi yang lebih tepat. Salah satu cara yang dapat dilakukan untuk menggali pengetahuan yang dapat dipergunakan untuk rencana promosi adalah dengan menggunakan analisis asosiasi. Analisis Apriori merupakan teknik data mining untuk menemukan aturan asosiatif kombinasi item. Algoritma Apriori merupakan salah satu teknik dalam analisis asosiasi. Penelitian dilakukan dengan mengamati beberapa variabel yang dipertimbangkan oleh perguruan tinggi dalam menentukan rencana promosinya yaitu daerah asal mahasiswa, jurusan mahasiswa, dan jenis sekolah menengah atas. Hasil dari penelitian ini digunakan sebagai salah satu dasar pengambilan keputusan untuk menentukan strategi promosi berdasarkan nilai support dan confidence untuk pemilihan jurusan mahasiswa berdasarkan daerah asal mahasiswa dan jenis sekolah menengah atas. Itemset yang mempunyai nilai support lebih dari 2% ada 13 itemset. Teknik Informatika banyak dipilih calon mahasiswa yang berasal dari Sleman, Bantul, Yogya, Klaten dengan nilai support 6% sampai 13%. Diperoleh 16 aturan asosiatif untuk menunjukkan asosiasi pemilihan calon mahasiswa terhadap jurusan dengan daerah asal dan jenis SMA dengan nilai confidence tertinggi 24%.

Kata Kunci: apriori, asosiasi, strategi, perguruan tinggi, promosi

ABSTRACT

Every company or organization needs to determine a promotional strategy to introduce the company or its organization. Determination of the right promotion strategy will be able to reduce promotion costs and achieve more appropriate promotional goals. One way that can be done to explore the knowledge that can be used for the promotion plan is to use association analysis. Apriori analysis is a data mining technique to find associative rules for item combinations. The Apriori algorithm is one technique in association analysis. The research was carried out by observing several variables considered by universities in determining the promotion plan, namely the area of origin of students, student majors, and the type of high school. The results of this study are used as a basis for decision making to determine promotion strategies based on the value of support and confidence for the selection of student majors based on the area of origin of students and types of high school. Items that have a support value of more than 2% have 13 itemset. Informatics Engineering is widely chosen by prospective students from Sleman, Bantul, Yogya, Klaten with a support value of 6% to 13%. Obtained 16 associative rules to show the association of prospective student selection of majors with the area of origin and type of high school with the highest confidence value of 24%.

Keywords: apriori, association, college, strategy, promotion

I. PENDAHULUAN

D ata dari situs resmi Direktorat Jenderal Pendidikan Tinggi Kementerian Pendidikan dan Kebudayaan Republik Indonesia, saat ini sudah ada sebanyak 541 universitas, 2.424 sekolah tinggi, 1107 akademi, 242 politeknik dan 131 Institut baik Perguruan Tinggi Negeri maupun Swasta di Indonesia yang tersebar dari Aceh Sampai Papua. Ditengah tingginya persaingan antar perguruan tinggi untuk mendapatkan mahasiswa, masing-masing Perguruan Tinggi berupaya menerapkan strategi promosi dengan harapan Perguruan Tinggi mereka menjadi pilihan lulusan SMA atau sederajat untuk studi lanjutnya.

STMIK AKAKOM merupakan salah satu perguruan Tinggi yang berada di Yogyakarta. STMIK AKAKOM melakukan beberapa kegiatan promosi untuk mengenalkan institusinya. Kegiatan promosi yang dilakukan diantaranya kunjungan ke beberapa mitra sekolah menengah atas, penyebaran informasi dengan beberapa media. Bagian Admisi dan Kerjasama diberikan tugas untuk merencanakan dan melakukan kegiatan promosi tersebut,

Dalam menjalankan kegiatan promosi tersebut membutuhkan beberapa sumber daya diantaranya biaya, tenaga dan waktu. Strategi promosi yang tepat diharapkan akan memberikan dampak efisiensi penggunaan sumber daya yang ada.

Data mining adalah salah satu cara untuk mencari pengetahuan dan pola dari data dalam jumlah besar. Data mining) adalah suatu proses ekstraksi atau penggalian data yang belum diketahui sebelumnya, namun dapat dipahami dan berguna dari *database* yang besar serta digunakan untuk membuat suatu keputusan bisnis yang sangat penting. Teknik – teknik data mining telah digunakan untuk menemukan pola yang tersembunyi dan memprediksi tren masa depan.[1].

Algoritma Apriori adalah salah satu metode dalam data mining yang digunakan untuk menentukan hubungan asosiasi suatu kombinasi item dengan aturan asosiatif (*Association rule*) [2]. Aturan asosiatif ditunjukkan dari nilai *support* dan *confidence* dari suatu hubungan item. Bagian pemasaran dan Admisi membutuhkan informasi mengenai hubungan asosiasi antara variabel jenis SMA, asal daerah terhadap pemilihan jurusan dengan harapan pola hubungan asosiasi tersebut dapat membantu dalam penentuan pola kunjungan promosi menurut daerah asal dan jenis SMA.

Penelitian [3] menerapkan data mining untuk mencari pola karakteristik sumber informasi yang digunakan oleh mahasiswa yang mendaftar di UNP Kediri. karakteristik yang ditampilkan berupa nilai support dan confidence hubungan antara jurusan dengan daerah asal mahasiswa Semakin tinggi nilai confidence dan support maka semakin kuat nilai hubungan antar atribut. Probabilitas tertinggi dengan nilai 72% adalah Teknik Informatika – Kediri. Sehingga dapat disimpulkan bahwa untuk promosi di daerah Kediri, pihak UNP dapat menonjolkan prodi teknik Informatika karena banyak diminati di daerah tersebut.

Penelitian [4] menerapkan data mining untuk menentukan daerah yang strategis untuk mengenalkan Sekolah Tinggi Keguruan dan Ilmu Pendidikan (STKIP) ADZKIA Dari data-data mahasiswa yang ada disekolah dapat diolah menggunakan algoritma Apriori dan FP-Growth yang menjadi informasi baru untuk dimanfaatkan oleh dalam menentukan daerah yang strategis. Dalam penelitian ini membandingkan hasil dari algoritma Apriori dan FP-Growth yang menggunakan data mahasiswa angkatan 2015/2016 dengan nilai $\text{minsupport} = 0.05\%$ dan nilai $\text{minconfidence} = 0.7\%$ telah diperoleh 19 Association Rule dan 2 rule.

II. METODE

Rancangan tahapan penelitian mengacu pada enam tahap *Cross-Industry Standard Process for Data Mining* (CRISP-DM) atau siklus hidup pengembangan data mining [5] sebagai berikut :

1. Pemahaman Bisnis (*Business Understanding*)

Pemahaman bisnis (*business understanding*), tahap pertama dalam proses CRISP-DM yang juga dapat disebut sebagai tahap pemahaman bisnis (penelitian). Tujuan pengolahan data dapat dijadikan sebagai salah satu dasar pengambilan keputusan untuk menentukan wilayah promosi yang tepat sasaran. Strategi awal dalam menerapkan tujuan dilakukannya *data mining* adalah melakukan permintaan data mahasiswa terlebih dahulu ke bagian Sistem Informasi dan Jaringan STMIK AKAKOM. *Dataset* yang diambil adalah data mahasiswa STMIK AKAKOM pada tahun 2012, 2013 dan 2014

2. Pemahaman Data (*Data Understanding*)

Dataset mahasiswa yang didapatkan dari bagian Sistem Informasi dan Jaringan STMIK AKAKOM berupa dokumen *excel* sejumlah 11997 *record*. Adapun sumber data utama yang digunakan dalam penelitian ini adalah *dataset* mahasiswa STMIK AKAKOM pada tahun 2012, 2013 dan 2014. *Dataset* mahasiswa terdiri dari atribut Angkatan, Nama Lengkap, Program Studi, Jenis Kelamin, Status Perkawinan, Kota Lahir, Tanggal Lahir, Agama, Alamat, Alamat Terakhir, Kota, Kode Pos, Negara, No. Telp, Jumlah Saudara, Tahun Tamat SMTA, Tahun Lulus SMTA, jenis SMA, Alamat SMTA. asal sekolah, daerah asal, jurusan. Jumlah data yang ada pada atribut berjumlah 3088 *record*. Hasil evaluasi terhadap kualitas data yaitu menemukan banyak nilai kosong / *null* yang disebut dengan *missing value* atau ada data yang tidak diisi pada atribut dalam *dataset* mahasiswa. Untuk data yang tidak ada nilainya maka data itu dibuang dari *dataset*.

3. Persiapan Data (*Data Preparation*)

Persiapan data mencakup semua kegiatan untuk membangun *dataset* mahasiswa yang akan diterapkan ke dalam alat pemodelan, dari data mentah awal berupa *dataset* mahasiswa dan selanjutnya akan melakukan proses *data mining*. Atribut yang digunakan adalah jenis SMA, alamat, program studi. Pada tahap ini merupakan tahap untuk memastikan data mahasiswa yang dipilih telah layak untuk dilakukan proses pengolahan.

Tidak semua atribut dimasukkan dalam set data yang digunakan dalam proses data mining karena hanya berperan sebagai referensi identifikasi, seperti jenis SMA, Kota, Program studi. Dari dataset sejumlah 3088 dilakukan preproseesing dengan menghapus missing data didapatkan 7909.

4. Pemodelan (*Modeling*)

Pemodelan dilakukan dengan mencari aturan asosiasi antara jenis SMA, kota dengan program studi. beserta nilai *support* dan *confidencenya*. Aturan asosiasi yang dibangun dengan menggunakan algoritma Apriori.

Algoritma Apriori adalah algoritma yang sudah sangat dikenal dalam melakukan pencarian frequent itemset dengan menggunakan teknik association rule. Pada algoritma apriori untuk menentukan kandidat-kandidat yang mungkin muncul dengan cara memperhatikan minimum support.

Ada dua proses utama yang dilakukan dalam algoritma Apriori [6], yaitu :

- a. Penggabungan (*Join*) Pada proses ini setiap item dikombinasikan dengan item yang lainnya sampai tidak terbentuk kombinasi lagi.
- b. Pemangkasan (*Prune*) Pada proses ini, hasil dari item yang telah dikombinasikan tadi lalu dipangkas dengan menggunakan minimum support yang telah ditentukan oleh user.

Langkah-langkah dalam algoritma Apriori sebagai berikut:

- a. Tentukan *minimum support*. Nilai *minimum support* yang dimasukkan adalah antara 0 – 100%.
- b. Iterasi 1 : hitung item-item dari support dari *database* untuk 1-itemset. Nilai *Support A, B* adalah proporsi dari transaksi dalam *database* yang mengandung A dan B. Setelah 1-itemset didapatkan, dari 1-itemset apakah diatas minimum support, apabila telah memenuhi minimum support, 1-itemset tersebut akan menjadi pola frequent tinggi.
- c. Iterasi 2 : untuk mendapatkan 2-itemset, harus dilakukan kombinasi dari k-itemset sebelumnya, kemudian scan database lagi untuk hitung item-item yang memuat support. itemset yang memenuhi minimum support akan dipilih sebagai pola frequent tinggi dari kandidat.
- d. Tetapkan nilai k-itemset dari support yang telah memenuhi minimum support dari k-itemset.
- e. Lakukan proses untuk iterasi selanjutnya hingga tidak ada lagi k-itemset yang memenuhi minimum support.

Setelah semua pola frekuensi tinggi ditemukan, barulah dicari aturan asosiasi yang memenuhi syarat minimum *confidence* dengan menghitung nilai confidence aturan asosiasi IF A THEN B. Nilai confidence aturan asosiasi IF A THEN B adalah perbandingan jumlah transaksi A dan B dengan jumlah tansaksi A dalam *database*.

5. Evaluasi (*Evaluation*)

Pada tahap ini, model sudah terbentuk dan diharapkan memiliki kualitas baik jika dilihat dari sudut pandang analisa data. Pada tahap ini akan dilakukan evaluasi terhadap keefektifan dan kualitas model sebelum digunakan dan menentukan apakah model dapat mencapai tujuan yang ditetapkan pada fase awal (*Business Understanding*).

6. *Deployment*

Pada tahap ini, pengetahuan atau informasi yang telah diperoleh akan diatur dan dipresentasikan dalam bentuk khusus sehingga dapat digunakan oleh manajemen di STMIK AKAKOM. Tahap *deployment* dapat berupa pembuatan laporan sederhana atau mengimplementasikan proses *data mining* yang berulang dengan menggunakan data yang baru.

III. HASIL

Pada awal pembentukkan kandidat itemset C1, dari dataset diperoleh 255 itemset yang terdiri dari 247 kota asal mahasiswa, 5 itemset program studi mahasiswa, 3 itemset jenis SMA mahasiswa. Itemset yang nilai supportnya kurang dari 2% ada 242 dari 255 kandidat itemset yang terbentuk sedangkan itemset yang mempunyai nilai support lebih dari 2% ada 13 itemset yaitu itemset Bantul, Gunung Kidul, Klaten, Kulon Progo, Purworejo, Yogyakarta, SMU, SMK, MA, Teknik Informatika, Sistem Informasi, Manajemen Informatika, Teknik Komputer, Komputerisasi Akutansi. Ini berarti sebagian besar asal mahasiswa STMIK AKAKOM dari Bantul, Gunung Kidul, Klaten, Kulon Progo, Purworejo, Yogyakarta. L1 itemset yang tersisa setelah di pangkas terlihat pada tabel I.

TABEL I. LARGE KE-1 ITEMSET

ITEMSET	SUPPORT (%)
Bantul	12
Gunung Kidul	4
Klaten	6
Kulon Progo	2
Purworejo	2
Yogyakarta	9
Teknik Informatika	49
Teknik Komputer	5
Manajemen Informatika	14
Komputerisasi Akuntansi	2
Sistem Informasi	29
SMK	40
SMU	55
MA	6

Itemset yang kurang memberikan sumbangan dalam dataset, yang nilai supportnya kurang dari 2% dipangkas dan diperoleh itemset Large ke-1 yaitu Bantul, Gunung Kidul, Klaten, Kulon Progo, Purworejo, Yogyakarta, SMU, SMK, MA, Teknik Informatika, Sistem Informasi, Manajemen Informatika, Teknik Komputer, Komputerisasi Akuntansi.

Itemset yang tidak termasuk dalam *large itemset* tidak diikuti dalam iterasi selanjutnya. Pada iterasi kedua akan untuk membentuk kandidat *itemset* kedua atau C2 menggunakan hasil *large itemset* pada iterasi pertama (L1) . Ditentukan kombinasi itemset A dan kombinasi itemset B dari L1 kemudian dihitung nilai supportnya. Kandidat C2 akan dipangkas dengan mengambil nilai minimum support 2%, diperoleh hasil pemangkas berupa Large ke-2 itemset atau L2 seperti tampak pada tabel II.

TABEL II. LARGE KE-2 ITEMSET

ITEMSET A	ITEMSET B	SUPPORT(%)
Teknik Informatika	Bantul	13
Teknik Informatika	Gunung Kidul	5
Teknik Informatika	Klaten	6
Teknik Informatika	Kulon Progo	2
Teknik Informatika	Purworejo	2
Teknik Informatika	Sleman	13
Teknik Informatika	Yogyakarta	10
Teknik Informatika	SMK	27
Teknik Informatika	SMU	22
Teknik Komputer	Klaten	2
Teknik Komputer	Sleman	2
Teknik Komputer	SMK	4
Teknik Komputer	SMU	2
Manajemen	Bantul	5
Manajemen	Gunung Kidul	2
Manajemen	Klaten	3
Manajemen	Sleman	5
Manajemen	Yogyakarta	3
Manajemen	SMK	7
Manajemen	SMU	12
Komputerisasi	SMU	2
Sistem Informasi	Bantul	6
Sistem Informasi	Gunung Kidul	2
Sistem Informasi	Klaten	2
Sistem Informasi	Sleman	7
Sistem Informasi	Yogyakarta	4
Sistem Informasi	SMK	6
Sistem Informasi	SMU	16
Bantul	MA	1
Purworejo	MA	0
Sleman	MA	1

Yogyakarta	MA	0
Bantul	SMK	12
Gunung Kidul	SMK	6
Klaten	SMK	6
Kulon Progo	SMK	2
Purworejo	SMK	1
Sleman	SMK	10
Yogyakarta	SMK	7
Bantul	SMU	14
Gunung Kidul	SMU	3
Klaten	SMU	6
Kulon Progo	SMU	1
Purworejo	SMU	2
Sleman	SMU	16
Yogyakarta	SMU	10

Dari L2 langkah berikutnya dibentuk aturan asosiasi (*association rule*) berupa aturan *IF* Itemset (A,B) *THEN* Itemset C dan dihitung nilai confidentnya.

IV. PEMBAHASAN

Pada awal pembentukkan kandidat itemset C1, dari dataset diperoleh 255 itemset yang terdiri dari 247 kota asal mahasiswa, 5 itemset program studi mahasiswa, 3 itemset jenis SMA mahasiswa. Itemset yang nilai supportnya kurang dari 2% ada 242 dari 255 kandidat itemset yang terbentuk sedangkan itemset yang mempunyai nilai support lebih dari 2% ada 13 itemset yaitu itemset Bantul, Gunung Kidul, Klaten, Kulon Progo, Purworejo, Yogyakarta, SMU, SMK, MA, Teknik Informatika, Sistem Informasi, Manajemen Informatika, Teknik Komputer, Komputerisasi Akutansi. Ini berarti sebagian besar asal mahasiswa STMIK AKAKOM dari Bantul, Gunung Kidul, Klaten, Kulon Progo, Purworejo, Yogyakarta.

Itemset yang kurang memberikan sumbangan dalam dataset, yang nilai supportnya kurang dari 2% dipangkas dan diperoleh itemset Large ke-1 yaitu Bantul, Gunung Kidul, Klaten, Kulon Progo, Purworejo, Yogyakarta, SMU, SMK, MA, Teknik Informatika, Sistem Informasi, Manajemen Informatika, Teknik Komputer, Komputerisasi Akutansi.

Iterasi berikutnya ditentukan kandidat itemset ke-2 atau C2 yang diperoleh dari kombinasi itemset L-1. Kemudian dihitung nilai support dan nilai confidence dari kombinasi itemset L-1. Nilai support dan nilai confidence yang diperoleh dapat diartikan sebagai berikut:

1. Untuk jenis SMA

- a) Mahasiswa yang berasal dari SMU mempunyai nilai support lebih dari 6% pada daerah Bantul, Sleman, Yogyakarta, Klaten dan Gunung Kidul.
- b) Mahasiswa yang berasal dari SMK mempunyai nilai support lebih dari 6% pada daerah Bantul, Sleman, Yogyakarta.
- c) Mahasiswa yang berasal dari MA mempunyai nilai support kurang dari 1% pada daerah semua daerah.

2. Untuk Program Studi

- a) Mahasiswa program studi Teknik Informatika yang berasal dari Sleman, Bantul, Yogya, Klaten mempunyai nilai support 6% sampai 13%.
- b) Mahasiswa program studi Teknik Komputer yang berasal dari Sleman, Klaten mempunyai nilai support 2%
- c) Mahasiswa program studi Manajemen Informatika yang berasal dari Sleman, Bantul mempunyai nilai support 5% .
- d) Mahasiswa program studi Sistem Informasi yang berasal dari Sleman, Bantul mempunyai nilai support 6% sampai 7%.
- e) Mahasiswa program studi Komputerisasi Akutansi yang berasal dari Sleman, Bantul, Yogya, Klaten mempunyai nilai support kurang dari 1%.

Pasangan aturan asosiasi dengan nilai *confidence* tinggi pada semua program studi adalah sebagai berikut:

- a) Mahasiswa yang berasal dari daerah Bantul dan SMK kemungkinan 15% akan memilih program studi Teknik Komputer.
- b) Mahasiswa yang berasal dari daerah Klaten dan SMK kemungkinan 14% akan memilih program studi Teknik Komputer.
- c) Mahasiswa yang berasal dari daerah Sleman dan SMK kemungkinan 16% akan memilih program studi Teknik Komputer.
- d) Mahasiswa yang berasal dari daerah Yogya dan SMU kemungkinan 10% akan memilih program studi Teknik Komputer.
- e) Mahasiswa yang berasal dari daerah Bantul dan SMK kemungkinan 14% akan memilih program studi Teknik Informatika.
- f) Mahasiswa yang berasal dari daerah Sleman dan SMK kemungkinan 11% akan memilih program studi Teknik Informatika.
- g) Mahasiswa yang berasal dari daerah Yogya dan SMK kemungkinan 10% akan memilih program studi Teknik Informatika.
- h) Mahasiswa yang berasal dari daerah Bantul dan SMU kemungkinan 13% akan memilih program studi Teknik Informatika.
- i) Mahasiswa yang berasal dari daerah Sleman dan SMU kemungkinan 14% akan memilih program studi Teknik Informatika.
- j) Mahasiswa yang berasal dari daerah Bantul dan SMU kemungkinan 18% akan memilih program studi Sistem Informasi.
- k) Mahasiswa yang berasal dari daerah Sleman dan SMU kemungkinan 21% akan memilih program studi Sistem Informasi.
- l) Mahasiswa yang berasal dari daerah Yogyakarta dan SMU kemungkinan 13% akan memilih program studi Sistem Informasi.
- m) Mahasiswa yang berasal dari daerah Bantul dan SMU kemungkinan 10% akan memilih program studi Manajemen Informatika.
- n) Mahasiswa yang berasal dari daerah Sleman dan SMU kemungkinan 10% akan memilih program studi Manajemen Informatika.
- o) Mahasiswa yang berasal dari daerah Bantul dan SMU kemungkinan 24% akan memilih program studi Komputerisasi Akuntansi.
- p) Mahasiswa yang berasal dari daerah Sleman dan SMU kemungkinan 24% akan memilih program studi Komputerisasi Akuntansi.

V. SIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang telah dilakukan dapat disimpulkan:

1. Dari dataset diperoleh 255 itemset yang terdiri dari 247 kota asal mahasiswa, 5 itemset program studi mahasiswa, 3 itemset jenis SMA mahasiswa.
2. Itemset yang mempunyai nilai support lebih dari 2% ada 13 itemset yaitu itemset Bantul, Gunung Kidul, Klaten, Kulon Progo, Purworejo, Yogyakarta, SMU, SMK, MA, Teknik Informatika, Sistem Informasi, Manajemen Informatika, Teknik Komputer, Komputerisasi Akuntansi.
3. Jenis SMA yang disarankan untuk dikunjungi berasal dari daerah Bantul, Sleman, Yogyakarta, Klaten dan Gunung Kidul dengan nilai support 6%.
4. Program studi Teknik Informatika banyak dipilih calon mahasiswa yang berasal dari Sleman, Bantul, Yogya, Klaten dengan nilai support 6% sampai 13%.
5. Diperoleh 16 aturan asosiatif untuk menunjukkan asosiatif pemilihan calon mahasiswa terhadap jurusan dengan daerah asal dan jenis SMA dengan nilai *confidence* tertinggi 24%.

REFERENSI

- [1] Thomas M. Connolly, and Carolyn E. Begg, “Database Systems A Practical Approach to Design, Implementation, and Management,” Sixth Edition, University of the west, Scotland, 2015.
- [2] Kusriani, Luthfi, E. T., 2009, Algoritma Data Mining, Andi Offset, Yogyakarta.
- [3] Fajar Rohman Hariri , Risky Aswi Ramadhani, “Penerapan Data Mining menggunakan Association Rules untuk Mendukung Strategi Promosi Universitas Nusantara PGRI Kediri,” SNATIKA, Volume 04, hal 138-142, 2017
- [4] Domi Sepri , M. Afdal, “Analisa Dan Perbandingan Metode Algoritma Apriori Dan Fp-Growth Untuk Mencari Pola Daerah Strategis Pengenalan Kampus Studi Kasus Di Stkip Adzkie Padang”, Jurnal Sistem Informasi Kaputama (JSIK), Vol 1 No 1, Januari 2017.
- [5] Pete Chapman , Julian Clinton, Randy Kerber, Thomas Khabaza, Thomas Reinartz, Colin Shearer and Rüdiger Wirth “CRISP-DM 1.0 Step-by-step data mining guide,” The CRISP-DM consortium, Agustus 2000.
- [6] Jiawei Han, Micheline Kamber, Jian Pei Simon, “ Data Mining Concepts and Techniques,” Third Edition, Morgan Kaufmann Publishers, USA, 2012.