

Metode Microsoft Clustering untuk Meningkatkan Potensi Penerimaan Mahasiswa Baru

Suharyadi¹, Eko Sedyono²

Abstract – *The competition among private universities in Indonesia for students admission is getting tough. This phenomenon is triggered by the fact that the main source of income for those universities is from their students. In order to win the competition is the targeting the promotion to the right area. Clustering method in data mining could be utilized get the information on the segment of prospective students, By segmenting these prospective students, a good promotion could be carried out in a more organized way. It also helps to achieve the cost effective and the target that the universities have. One of the advantages of this kind of promotion is minimizing the promotion costs and the increasing responses from prospective students. This model is built using the ready to use tool in the market, Microsoft Excel and SQL Server 2008 Data Mining Add in. By implementing thus segmenting process to the applicant, it could classify which areas contributing to the number of the prospective students and which major they are interested in.*

Keywords : *Clustering, Data Mining, Promotion Strategy Analysis.*

I. PENDAHULUAN

A. Latar Belakang

Saat ini jumlah Perguruan Tinggi Swasta (PTS) di Indonesia meningkat. Menurut pangkalan data DIKTI yang ada di <http://forlap.dikti.go.id/>, saat ini mencapai 4.077 institusi, mulai dari akademi, sekolah tinggi, institut, hingga universitas [13].

Meningkatnya jumlah PTS belum diiringi kualitas yang memadai [1]. Tingginya tingkat persaingan antar institusi pendidikan mengakibatkan setiap institusi harus dapat mengelola institusinya secara profesional, seperti halnya sebuah perusahaan, sejak dari bagaimana sebuah institusi pendidikan dapat memiliki mutu yang baik, sampai dengan daya saing yang tinggi (sesuai standar Kemenristek Dikti). Perkembangan teknologi informasi dan komunikasi menjadi bagian dari proses-proses

tersebut.

Seperti layaknya sebuah organisasi, PTS membutuhkan sumber pendapatan untuk membiayai semua kegiatan dan program. Mahasiswa adalah sumber pendapatan utama bagi keberlangsungan PTS. Semua strategi akan dilakukan oleh sebuah PTS untuk mendapatkan mahasiswa sebanyak mungkin sesuai dengan kapasitas. Jumlah lulusan Sekolah Menengah Umum (SMU) tidak semuanya akan melanjutkan ke PT. Hal ini menjadi faktor lain terbatasnya jumlah calon mahasiswa yang mendaftar ke PTS. Sementara itu Perguruan Tinggi Negeri (PTN) masih menjadi tujuan utama bagi calon mahasiswa. PTS harus menerapkan strategi promosi yang baik, agar dapat memenangkan persaingan dalam mendapatkan calon mahasiswa yang jumlahnya terbatas. Promosi yang dilakukan oleh PTS berusaha menjaring semua lulusan sekolah menengah atas tanpa memperhatikan potensi daerah yang dituju. Hal ini selain menimbulkan pemborosan tenaga juga pemborosan biaya, karena hasil yang didapat tidak sebanding dengan biaya yang dikeluarkan. Dengan mengetahui potensi suatu daerah sebagai tujuan atau sasaran promosi akan sangat membantu dalam memaksimalkan jumlah mahasiswa yang berpotensi mendaftar ke PTS, sehingga efektifitas biaya dan tenaga akan tercapai.

B. Rumusan Masalah

Bagaimana menentukan daerah yang berpotensi memberikan sumbangan jumlah calon mahasiswa baru dengan menggunakan metode klustering terhadap data pendaftar.

C. Batasan Masalah

Batasan masalah dalam penelitian ini adalah :

1. Menggunakan metode clustering dalam mengolah data pendaftar calon mahasiswa, dengan terlebih dahulu melakukan tahap-tahap data mining.
2. Menggunakan data pendaftar tahun 2009 sampai dengan tahun 2011 yang berjumlah 15.000 record.
3. Menggunakan tools Microsoft SQL Server Analysis Service, Microsoft Office Excel 2010 yang sudah dilengkapi dengan Data Mining Add-In for Microsoft Excel 2007.

B. Tinjauan Pustaka

¹ Program Studi Komputerisasi Akuntansi, Universitas Kristen Satya Wacana

email: haryadi@staff.uksw.edu

² Program Studi Magister Sistem Informasi,

Universitas Kristen Satya Wacana

email: eko@staff.uksw.edu

Penelitian tentang pemanfaatan teknologi *data warehouse* di Institusi Pendidikan Tinggi Indonesia, menunjukkan bahwa ternyata Indonesia jauh tertinggal dengan institusi pendidikan tinggi di beberapa negara maju seperti Amerika[10]. Masalah utama yang menghambat adalah institusi pendidikan tinggi di Indonesia belum banyak yang memanfaatkan sistem informasi yang berbasis teknologi informasi dan komunikasi (TIK) untuk mendukung pengelolaan institusinya. Tanpa adanya sistem informasi yang digunakan untuk membantu pengelolaan institusi, maka teknologi *Data Warehouse* tidak mungkin dapat dikembangkan. Perkembangan pemanfaatan sistem informasi untuk mendukung pengelolaan institusi pendidikan tinggi di Indonesia perlu mendapatkan perhatian khusus, walaupun Direktorat Jenderal Pendidikan Tinggi telah memperkenalkan dan membuat rancang bangun SIM-PT sejak tahun 1990.

Penelitian lain tentang penerapan *Data Warehouse* dan *Data Mining* adalah untuk pengolahan data akademik. Dalam penelitian tersebut dipaparkan tentang kecenderungan mahasiswa dalam mengambil matakuliah tertentu dan matakuliah lainnya, dimana pola-pola yang dihasilkan sangat terlihat jelas menggambarkan tentang tingkat kemampuan mahasiswa. Penelitian tersebut adalah penelitian tentang peminatan mahasiswa terhadap matakuliah dan matakuliah yang lainnya berdasarkan nilai matakuliah yang diperoleh [1].

Penelitian tentang pemanfaatan *Data Mining* untuk melakukan strategi promosi dengan *Direct Marketing*, membahas tentang perbandingan efektifitas pemasaran dengan metode *mass mailing* dengan *direct mailing*. Hasil penelitian tersebut membuktikan bahwa pemasaran dengan cara *direct mailing* dinilai lebih efektif dalam menjangkau pembeli yang aktif, meskipun hasil penjualannya lebih sedikit namun prosentase dari biaya promosi juga sedikit sehingga laba yang dihasilkan lebih besar. Dengan *direct marketing*, teknologi *Data Mining* digunakan untuk calon *customer* yang potensial. Berdasarkan penelitian tersebut dihasilkan perbandingan seperti pada Tabel 1.

Tabel 1. Perbandingan *Mass Mailing* dan *Direct Mailing* [4]

| | Mass Mailing | Direct mailing |
|--|---------------------|-----------------------|
| Number of customers mailed | 600,000 | (20%) 120,000 |
| Cost of printing, mailing (\$ 0.71 each) | \$426,000 | \$85,200 |
| Cost of data mining | \$0 | \$40,000 |

| | | |
|------------------------------|-----------|-----------|
| Total promotion cost | \$426,000 | \$125,200 |
| Response rate | 1,0% | 3,0% |
| Number of Sales | 6,000 | 3,600 |
| Profit from sale 9\$70 each) | \$420,000 | \$252,000 |
| Net profit from promotion | -86,000 | \$126,800 |

Tabel 1 menjelaskan perbandingan hasil antara promosi yang dilakukan dengan mengirimkan surat ke seluruh calon dan mengirimkan surat hanya kepada calon yang potensial. Data membuktikan bahwa promosi dengan *direct marketing* menghasilkan profit yang lebih besar dibandingkan dengan promosi yang konvensional atau *mass mailing* [4].

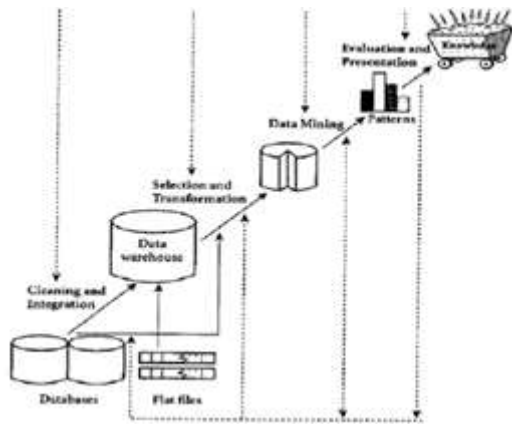
II. LANDASAN TEORI

A. Data Mining

Data Mining memiliki hakikat sebagai disiplin ilmu yang tujuan utamanya adalah untuk menemukan, menggali, atau menambang pengetahuan dari data atau informasi yang dimiliki [10]. *Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar

B. Tahapan Data Mining

Tahapan dalam *Data Mining*, adalah sebagai berikut : Pembersihan data (*data cleaning*); Integrasi data (*data integration*); Seleksi Data (*Data Selection*) ; Transformasi data (*Data Transformation*); Proses *mining*; Evaluasi pola (*pattern evaluation*) [2]. Pada umumnya proses *Data Mining* berjalan interaktif karena tidak jarang hasil *Data Mining* pada awalnya tidak sesuai dengan harapan analisisnya, sehingga perlu dilakukan desain ulang prosesnya. Tahap-tahap dalam *Data Mining* ditunjukkan pada Gambar 1.



Gambar 1. Tahap-tahap Data Mining [2]

Algoritma K-Mean Cluster Analysis [7]

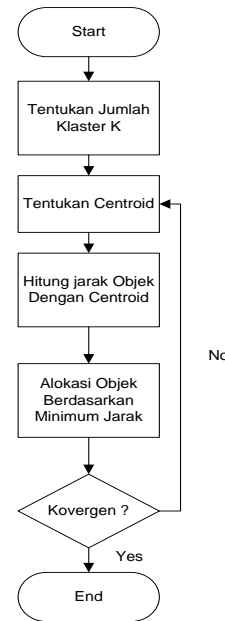
Jika diberikan sekumpulan data $X=(x_1,x_2,\dots,x_n)$, maka algoritma *k-means cluster analysis* akan mempartisi X dalam *k* buah *cluster*, setiap *cluster* memiliki *centroid* (titik tengah) atau *mean* dari data dalam *cluster* tersebut.

Pada tahap awal algoritma *k-means cluster analysis* akan memilih secara acak *k* buah data sebagai *centroid* (titik tengah), kemudian jarak antara data dengan *centroid* dihitung dengan menggunakan *Euclidean distance*. Nilai *k* menentukan banyaknya anggota kluster, oleh karena itu nilai *k* dapat ditentukan (*supervised clustering*) atau tidak ditentukan sebelumnya (*unsupervised clustering*). Pada proses selanjutnya, data akan ditempatkan dalam kluster yang terdekat, yang dihitung dari titik tengah kluster. *Centroid* baru akan ditetapkan, jika semua data sudah ditempatkan dalam *cluster* terdekat.

Proses penentuan *centroid* dan penempatan data dalam *cluster* diulangi sampai nilai *centroid* konvergen (*centroid* dari semua *cluster* tidak berubah lagi). Secara umum *K-Means Cluster analysis* menggunakan algoritma sebagai berikut :

1. Tentukan *k* sebagai jumlah *cluster* yang akan dibentuk
2. Bangkitkan *k Centroid* (titik pusat *cluster*) awal secara *random*
3. Hitung jarak setiap data ke masing-masing *centroid* dari masing-masing *cluster*
4. Alokasikan masing-masing data ke dalam *centroid* yang paling terdekat
5. Lakukan iterasi, kemudian tentukan posisi *centroid* baru dengan cara menghitung rata-rata dari data-data yang berada pada *centroid* yang sama
6. Ulangi langkah 3, jika posisi *centroid* baru dan *centroid* lama tidak sama.

Algoritma *k-means cluster analysis* dalam bentuk *flowchart* ditunjukkan pada Gambar 2.



Gambar 2. Flowchart Algoritma K-Mean Cluster Analysis

C. Menentukan Centroid

Penentuan *centroid* awal dilakukan secara *random/acak* dari data/objek yang tersedia sebanyak jumlah *cluster k*, kemudian untuk menghitung *centroid cluster* berikutnya ke *i*, *v_i* digunakan rumus pada Persamaan 2.1.

$$V_k = \frac{\sum_{i=1}^{N_k} X_i}{N_k} \text{ ----- (2.1)}$$

dimana :

V_k : *centroid* pada *cluster* ke *k*

X_i : Data ke *i*

N_k : Banyaknya objek/jumlah data yang menjadi anggota *cluster* ke *k*

D. Menghitung jarak antara data dengan centroid

Untuk menghitung jarak antara data dengan *centroid* terdapat beberapa cara yang dapat dilakukan yaitu *Manhattan/City Block distance (L1)*, *Euclidean Distance (L2)*. Jarak antara dua titik *X1* dan *X2* pada *manhattan/city block* dihitung dengan menggunakan Persamaan 2.2.

$$D_{L_1}(x_2, x_1) = \|x_2 - x_1\|_1 = \sum_{j=1}^p |x_{2j} - x_{1j}| \text{(2.2)}$$

Dimana P : Dimensi data

$|\cdot|$: Nilai Absolut

Sedangkan untuk *euclidean distance* jarak antara data dengan *centroid* dihitung dengan menggunakan Persamaan 2.3.

$$D_{L_2}(x_2, x_1) = \|x_2 - x_1\|_2 = \sqrt{\sum_{j=1}^p (x_{2j} - x_{1j})^2} \quad \dots\dots$$

(2.3)

Dimana P : Dimensi data

$|\cdot|$: Nilai Absolut

E. Pengalokasian Data ke Masing-masing Cluster

Untuk melakukan pengalokasian data ke dalam masing-masing *cluster* pada saat iterasi dilakukan secara umum dengan dua cara yaitu dengan cara pengalokasian dengan cara *hard k-means*, dimana secara tegas setiap objek dinyatakan sebagai anggota *cluster* satu dan tidak menjadi anggota *cluster* lainnya. Cara lain adalah dengan cara *fuzzy k-means* dimana masing-masing objek diberikan nilai kemungkinan untuk bisa bergabung dengan setiap *cluster* yang ada. Pada paper ini digunakan model *hard - means clustering*.

Hard K-means, pengalokasian kembali objek ke dalam masing-masing *cluster* berdasarkan pada perbandingan jarak antara data dengan *centroid* setiap *cluster* yang ada. Obyek dialokasikan secara tegas ke dalam *cluster* yang mempunyai jarak ke *centroid* terdekat dengan data tersebut. Pengalokasian ini dirumuskan Pada Persamaan 2.4.

$$a_{ik} = \begin{cases} 1 & \text{if } d = \min\{D(x_i, v_j)\} \\ 0 & \text{lainnya} \end{cases} \quad \dots\dots\dots$$

(2.4)

dimana :

a_{ik} : Keanggotaan data atau objek ke *k* pada *cluster* ke-*i*

v_j : Nilai *centroid cluster* ke *i*

F. Jenis Algoritma Data Mining pada SQL Server 2008

SQL Server 2008 memiliki 9 jenis algoritma *data mining* [3]. Tidak semua algoritma tersebut digunakan dalam memecahkan masalah *data mining*, karena setiap algoritma memiliki karakteristik yang mungkin cocok dengan karakteristik masalah yang perlu dicarikan solusi di dalam *data mining*. Sembilan jenis algoritma *data mining* tersebut adalah : (1) *Microsoft Decision Tree*, (2)

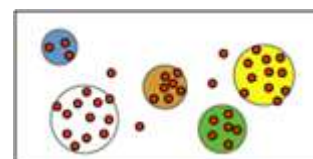
Microsoft Linear Regression; (3) *Microsoft Naïve Bayes*; (4) *Microsoft Clustering* (5) *Microsoft Association Rules*; (6) *Microsoft Sequence Clustering*; (7) *Microsoft Time Series*; (8) *Microsoft Neutral Network*; (9) *Microsoft Logistic Regression Algorithm*. Dalam paper ini dibahas algoritma yang ke empat.

G. Algoritma Microsoft Clustering

Algoritma *Microsoft Clustering* adalah algoritma segmentasi yang disediakan oleh *Analysis Services* [12]. Algoritma ini menggunakan teknik *iteratif* untuk pengelompokan kasus *data set* ke dalam kelompok yang mengandung karakteristik serupa. Pengelompokan ini berguna untuk eksplorasi data, mengidentifikasi anomali dalam data, dan membuat prediksi.

Algoritma *Microsoft Clustering* menyediakan dua metode untuk membuat *cluster* dan menetapkan titik data ke *cluster*. Metode pertama, algoritma *K-means*, adalah metode *hard clustering*. Ini berarti bahwa titik data dapat dikelompokkan pada satu *cluster*, serta probabilitas tunggal dihitung untuk keanggotaan setiap titik data pada *cluster* tersebut. Metode kedua, *Expectation Maximization* (EM), adalah metode *soft clustering*. Ini berarti bahwa titik data dapat menjadi milik beberapa kelompok, dan probabilitas dihitung untuk setiap kombinasi titik data dan *cluster*. EM *Clustering* merupakan algoritma *partitional* yang berbasiskan model yang menggunakan perhitungan probabilitas, bukan jarak seperti pada umumnya algoritma *clustering* yang lain.

Cara Kerja Algoritma *Microsoft Clustering* dijelaskan sebagai berikut. Algoritma *Microsoft Clustering* akan membuat beberapa *cluster* data, dan *cluster* data tersebut terhubung satu dengan lainnya. Setelah *cluster* tersebut terbentuk dan menerima *input* atribut yang akan diprediksi beserta nilainya, algoritma ini akan menganalisis setiap *cluster* berdasarkan *input* yang diterima dan akan menampilkan atribut kontributor prediksi yang paling mempengaruhi dari setiap *cluster*. Algoritma *Microsoft Clustering* pertama mengidentifikasi hubungan dalam *data set* dan menghasilkan serangkaian *cluster* berdasarkan hubungan tersebut. Sebuah *plot* mewakili sekelompok data, seperti yang ditunjukkan pada Gambar 3.



Gambar 3. Clustering [11]

Plot mewakili semua kasus dalam *data set* dan setiap kasus digambarkan sebagai titik pada grafik. Gambar 3 menunjukkan bahwa algoritma telah mengidentifikasi *point* dalam kelompok *cluster*. Setelah mengidentifikasi *cluster*, algoritma menghitung seberapa besar *point* tersebut mewakili kelompok, kemudian mencoba untuk mendefinisikan kembali kelompok untuk membuat *cluster* yang lebih mewakili data.

Teknik clustering juga digunakan untuk keuntungan yang berbeda seperti; mengenali kelas mana yang paling banyak produknya terjual, mengelompokkan pelanggan berdasarkan perilaku pembelian mereka dan kekuatan mereka dalam membeli [5].

H. Promosi

promosi merupakan kegiatan pemasaran dan penjualan dalam rangka menginformasikan dan mendorong permintaan terhadap produk, jasa, dan ide dari perusahaan dengan cara memengaruhi para konsumen agar mau membeli produk dan jasa yang dihasilkan oleh perusahaan.[9]. Tujuan promosi diantaranya adalah: menyebarkan informasi produk kepada target pasar potensial, mendapatkan kenaikan penjualan dan profit/laba, mendapatkan pelanggan baru dan menjaga kesetiaan pelanggan, menjaga kestabilan penjualan ketika terjadi lesu pasar, Membedakan serta mengunggulkan produk dibanding produk pesaing, membentuk citra produk di mata konsumen sesuai dengan yang diinginkan; mengubah tingkah laku dan pendapat konsumen.

Strategi Promosi

Strategi Promosi adalah kegiatan perusahaan untuk mendorong penjualan dengan mengarahkan komunikasi-komunikasi yang menyakinkan kepada para pembeli [5]. Promosi merupakan teknik dan cara-cara yang dipakai oleh suatu perusahaan atau organisasi untuk melakukan promosi dengan tujuan untuk mendapatkan efisiensi promosi yang baik termasuk di dalamnya kegiatan pemasaran untuk menginformasikan, membujuk serta mengingatkan konsumen mengenai produk atau jasa yang ditawarkan perusahaan. Tujuannya agar konsumen dapat mengetahui, menyukai dan akhirnya membeli produk tersebut. Penelitian ini diharapkan menghasilkan sebuah metode untuk pengambilan keputusan penentuan daerah tujuan promosi Universitas Kristen Satya Wacana Salatiga.

III. METODE PENELITIAN

Metode penelitian ini mencakup studi literatur, analisis data dan algoritma, percobaan yang merupakan proses

data mining, serta analisis hasil percobaan. Literatur yang menjadi sumber referensi terdiri dari buku, jurnal, situs *web* dan makalah tentang *data mining* dan *clustering*. Percobaan dilakukan dengan variasi algoritma yang telah diimplementasikan pihak lain, dan variasi atribut yang diikutsertakan untuk kasus-kasus tertentu. Tahapan Penelitian yang dilakukan, dijelaskan sebagai berikut.

Tahap 1 : Pengumpulan data dan analisis permasalahan

Pada tahap satu, dilakukan pengumpulan data terutama data yang berhubungan langsung dengan penerimaan mahasiswa baru di UKSW. Data tersebut antara lain : data pendaftar, data ini dikelola oleh Biro Promosi untuk dilakukan pengolahan, terutama dilakukan komunikasi kepada calon mahasiswa.

Tahap 2 : Analisis Data Pendaftar

Selain informasi dari hasil wawancara, dalam tahap analisis data juga dilakukan analisis terhadap semua data yang berkaitan dengan penelitian, dihubungkan dengan tujuan dari penelitian. Data yang digunakan untuk analisis ini adalah data pendaftar mahasiswa baru.

Tahap 3 : Pemahaman data dan Perbaikan Data

Tahap berikutnya yang dilakukan adalah perbaikan data. Perbaikan yang dilakukan meliputi : melengkapi *record-record* data yang belum lengkap, mengubah *record* data yang tidak konsisten dalam penulisan maupun melengkapi data yang belum lengkap.

Tahap 4 : Percobaan Algoritma

Pada tahap 4, dilakukan percobaan terhadap algoritma yang dipakai untuk melakukan analisis data. Algoritma yang diterapkan dalam pengolahan data adalah algoritma *K-Mean*. Aplikasi yang digunakan untuk melakukan analisis terhadap data promosi adalah: Microsoft Office Excel 2010, SQL Server 2008 *Data Mining Add-in* yang dilengkapi dengan SQL Server 2008 *Analisis Service*, SQL Server *Reporting Service*.

Tahap 5 : Analisis Hasil Percobaan

Pada tahap 5 dari penelitian ini, merupakan tahap penarikan kesimpulan terhadap hasil analisis atas percobaan algoritma. Analisis adalah kontribusi dari penelitian mengenai metode klustering untuk menunjang strategi promosi. Pengambilan keputusan promosi yang tadinya hanya sebatas keputusan rutin, maka setelah dilakukan analisis, diharapkan kebijakan promosi yang diambil bisa berpedoman terhadap hasil analisis data.

Metode Pengumpulan Data

Pengumpulan data dilakukan dengan menggunakan metode sebagai berikut : (1) Observasi; Observasi atau

pengamatan dilakukan ke beberapa bagian yang terlibat dalam kegiatan promosi Universitas. Bagian yang menjadi objek observasi penelitian adalah Bagian Promosi Universitas, Bagian Admisi dan Registrasi, dan Wawancara; Wawancara hanya dilakukan terhadap pejabat pengambil keputusan di bagian BPHL yaitu manajer BPHL.

Analisis Strategi Promosi

Biro Promosi dan Hubungan Luar (BPHL) adalah unit di bawah kantor Pembantu Rektor IV Universitas Kristen Satya Wacana (UKSW), yang bertugas untuk mengelola dan mengembangkan program promosi universitas, serta mengelola *Public Relations* (humas) UKSW, dalam upaya untuk meningkatkan citra universitas. Dalam melaksanakan kegiatannya, BPHL menjalin kemitraan dengan pemerintah, gereja, sekolah, dan alumni. Selain itu BPHL juga membawahi *Satya Wacana Career Centre dan Alumni* (SWCA) yang mengelola dan mengembangkan bursa kerja bagi alumni melalui jejaring dan komunikasi dengan perusahaan dan para pengguna alumni. Komunikasi yang berkesinambungan dengan alumni juga merupakan salah satu hal penting yang dilakukan SWCA untuk menangani pelacakan dan pengelolaan umpan balik dari para alumni. Semua ini dilakukan untuk meningkatkan kualitas dari para lulusannya agar dapat memenuhi kebutuhan yang ada di masyarakat.

Pemahaman Data

Data pendaftar yang dikelola BPHL adalah data yang setelah melewati masa admisi menjadi data yang hanya tersimpan saja. Data tersebut adalah data hasil entri dari bagian Pemanduan Minat Akademik (PEMAMIK) yang akan ditindaklanjuti oleh bagian BPHL untuk dikonfirmasi ulang perihal kelanjutan pendaftaran, memiliki struktur tabel yang ditunjukkan pada Tabel 2.

Tabel 2. Struktur Tabel Data Pendaftar

| <i>Field</i> | <i>Type</i> |
|----------------------|------------------|
| No_formulir | <i>Numeric</i> |
| Nama | <i>Character</i> |
| Pilihan1 | <i>Character</i> |
| Status1 | <i>Character</i> |
| Pilihan2 | <i>Character</i> |
| Status2 | <i>Character</i> |
| No Telepon | <i>Numeric</i> |
| No Telepon Orang tua | <i>Numeric</i> |

| | |
|----------------|------------------|
| Alamat | <i>Character</i> |
| Kelurahan | <i>Character</i> |
| Kecamatan | <i>Character</i> |
| Kabupaten | <i>Character</i> |
| Propinsi | <i>Character</i> |
| Kode Pos | <i>Numeric</i> |
| Nama Ortu | <i>Character</i> |
| Alamat Ortu | <i>Character</i> |
| Nama Ortu | <i>Character</i> |
| Kecamatan Ortu | <i>Character</i> |
| Kabupaten Ortu | <i>Character</i> |
| Propinsi Ortu | <i>Character</i> |
| Kode pos Ortu | <i>Character</i> |
| Asal Sekolah | <i>Character</i> |

Data diolah menggunakan Microsoft Office Excel 2010, yang sudah dilengkapi dengan SQL Server 2008 *Data Mining Add-in for Microsoft Excel 2007*, Microsoft *SQL Server Analysis Service*.

Data pendaftar selanjutnya diolah dengan tahapan sebagai berikut :

1. Pembersihan data (*Data Cleaning*); Pembersihan data yang dilakukan adalah proses menghilangkan *noise* dan data yang tidak konsisten atau data tidak relevan serta melengkapi data yang kurang lengkap.
2. Integrasi data (*data integration*); Integrasi data merupakan penggabungan data dari berbagai database ke dalam satu database baru. Dalam kasus yang diteliti, database hanya berasal dari satu database saja yaitu data Sistem Informasi Akademik.
3. Seleksi Data (*data selection*); Seleksi Data yang dilakukan dalam penelitian adalah memilah *field-field* apa saja yang berhubungan dengan *output* yang diinginkan. Dalam penelitian ini, dipakai data pendaftar mahasiswa baru. Data pendaftar mahasiswa baru memuat atribut seperti yang ditunjukkan pada Tabel 3. Alasan mengapa *field-field* pada table 3 yang dipilih untuk diseleksi, karena *field-field* tersebut yang memuat informasi yang dibutuhkan dalam pengambilan keputusan promosi mahasiswa baru. Pemilihan semua *field-*

field yang akan diolah hanya akan memperlambat proses mining.

Tabel 3. Hasil Seleksi Tabel Data Pendaftar

| <i>Field</i> | <i>Type</i> |
|----------------|-------------|
| thn_daftar | Date |
| Pilihan 1 | Character |
| Pilihan 2 | Character |
| Kelurahan | Character |
| Kecamatan | Character |
| Kabupaten | Character |
| Propinsi | character |
| Kelurahan Ortu | character |
| Kecamatan Ortu | character |
| Kabupaten Ortu | character |
| Propinsi Ortu | character |
| Asal Sekolah | character |

Penjelasan untuk data yang ada dalam Tabel 3, ditunjukkan pada Tabel 4.

Tabel 4. Penjelasan Tabel Data Pendaftar

| <i>Field</i> | <i>Keterangan</i> |
|----------------|--|
| thn_daftar | Tahun mahasiswa mendaftar sebagai calon mahasiswa baru |
| Pilihan 1 | Bersisi nama progdi yang menjadi pilihan pertama |
| Pilihan 2 | Bersisi nama progdi yang menjadi pilihan kedua |
| Kelurahan | Nama kelurahan dari alamat SMU |
| Kecamatan | Nama kecamatan dari alamat SMU |
| Kabupaten | Nama kabupaten dari alamat SMU |
| Propinsi | Propinsi dimana SMU berada |
| Kelurahan Ortu | Nama kelurahan dari alamat asal |
| Kecam | Nama kecamatan dari alamat asal |

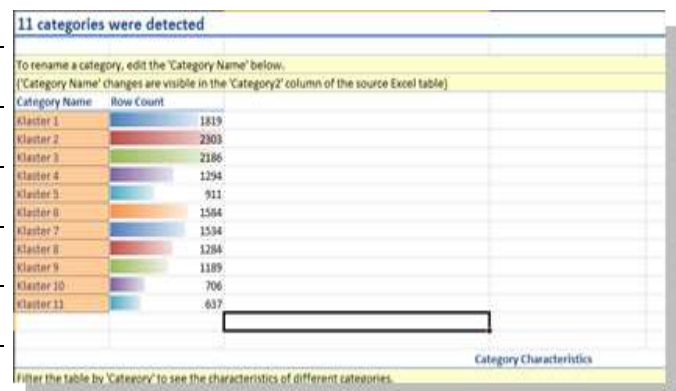
| | |
|----------------|---------------------------------|
| atan Ortu | |
| Kabupaten Ortu | Nama kabupaten dari alamat asal |
| Propinsi Ortu | Propinsi asal calon mahasiswa |
| Asal Sekolah | Nama asal sekolah |

Berdasarkan stuktur tabel pada Tabel 4, maka pengolahan data dengan *data mining* akan lebih efektif dan efisien, karena hanya *field-field* yang berhubungan dengan tujuan saja yang akan diolah.

4. Transformasi Data; Transformasi data merupakan proses perubahan atau penggabungan data ke dalam format yang sesuai untuk diproses dalam *data mining*. Seringkali data yang akan digunakan dalam proses *data mining* mempunyai format yang belum dapat digunakan secara langsung, oleh karena itu perlu dirubah formatnya. Dalam penelitian ini, data yang diolah sudah memenuhi untuk pengolahan *data mining*.
5. Proses Mining; Proses *mining* adalah proses inti untuk melakukan eksekusi terhadap data dalam menemukan pola-pola yang akan digali. Dalam melakukan proses tersebut, pertama yang akan dilakukan adalah : menghubungkan antara tujuan yang ingin dicapai dalam melakukan *data mining* serta memilih algoritma yang sesuai. Berdasarkan tujuan penelitian yang dilakukan, maka algoritma yang sesuai untuk diterapkan adalah algoritma *clustering*.

IV. HASIL PEMBAHASAN

Algoritma *Microsoft Clustering* akan mendeteksi data dan menghasilkan sejumlah klaster. Dalam percobaan ini algoritma menghasilkan 11 klaster dengan karakteristik masing-masing seperti Gambar 4.



Gambar 4. Hasil Klaster

Penjelasan setiap klaster

Tabel 5. Klaster 1

| Category | Column | Value | Relative Importance |
|-----------|--------------|-----------------|---------------------|
| Klaster 1 | Pilihan 1 | S1 PEND.GURU SD | |
| Klaster 1 | Propinsi | Jawa Tengah | |
| Klaster 1 | Pilihan 2 | - | |
| Klaster 1 | Pilihan 1 | D2 PEND.GURU SD | |
| Klaster 1 | Asal Sekolah | Missing | |

Klaster 1 Tabel 5 didominasi oleh calon mahasiswa yang mendaftar di Program Studi S1 Pendidikan Guru Sekolah Dasar (PGSD), yang berasal dari Jawa Tengah. Klaster 1 juga didominasi oleh calon pendaftar yang berminat untuk masuk di Program Kependidikan, terlihat dari pilihan program studi baik pilihan 1 maupun pilihan 2 adalah ProgdI S1 Pendidikan Guru SD dan D2 Pendidikan Guru SD. Asal sekolah berasal dari wilayah Pati, Grobogan, Wonosobo, Purwodadi dan daerah sekitarnya. Dominasi pada klaster 1 lebih jelas dengan pilihan 1 Program Studi S1 PGSD dan tidak memilih pilihan 2. Kondisi seperti ini menandakan bahwa calon mahasiswa yang masuk dalam klaster 1, menunjukkan tingkat keseriusan yang tinggi terhadap progdI ini.

Tabel 6. Klaster 2

| Category | Column | Value | Relative Importance |
|-----------|--------------|------------------------|---------------------|
| Klaster 2 | Asal Sekolah | SMA Negeri 1, Tenganan | |
| Klaster 2 | Propinsi | Jawa Tengah | |
| Klaster 2 | Pilihan 1 | S1 PEND.GURU SD | |
| Klaster 2 | Asal Sekolah | SMA Negeri 2, Salatiga | |
| Klaster 2 | Pilihan 2 | - | |
| Klaster 2 | Kecamatan | Tenganan | |
| Klaster 2 | Kecamatan | Tuntang | |
| Klaster 2 | Asal Sekolah | SMA Negeri 1, Getasan | |
| Klaster 2 | Kecamatan | Getasan | |
| Klaster 2 | Kecamatan | Suruh | |
| Klaster 2 | Pilihan 1 | BIMBINGAN KONSELING | |

Klaster 2 pada tabel 6 nilai dominan adalah asal sekolah, yaitu SMA Negeri 1 Tenganan.

Program Studi pilihan adalah Program Studi S1 PGSD untuk wilayah Salatiga dan sekitarnya, yaitu wilayah Kecamatan Getasan, Kecamatan Bringin, Kecamatan Pabelan. Wilayah Salatiga luar kota masih menjadi wilayah yang dominan untuk masuk di ProgdI S1 PGSD.

Tabel 7. Klaster 3

| Category | Column | Value | Relative Importance |
|-----------|--------------|-------------------------|---------------------|
| Klaster 3 | Asal Sekolah | SMA Kristen 1, Salatiga | |
| Klaster 3 | Propinsi | Jawa Tengah | |
| Klaster 3 | Kecamatan | Sidorejo | |
| Klaster 3 | Kecamatan | Argomulyo | |
| Klaster 3 | Asal Sekolah | SMA Negeri 2, Salatiga | |
| Klaster 3 | Asal Sekolah | SMA Negeri 1, Salatiga | |
| Klaster 3 | Kecamatan | Sidomukti | |
| Klaster 3 | Asal Sekolah | SMA Negeri 3, Salatiga | |
| Klaster 3 | Kecamatan | Tingkir | |
| Klaster 3 | Pilihan 1 | TEKNIK INFORMATIKA | |

Klaster 3 pada tabel 7 didominasi oleh data asal sekolah, yaitu SMA Kristen 1, Salatiga, sedangkan propinsi adalah Jawa Tengah dan kecamatan yang dominan adalah Sidorejo dan Argomulyo. Selain SMA Kristen 1 Salatiga, SMA lain yang dominan adalah SMA Negeri 2 dan SMA Negeri 1 Salatiga.

Tabel 8. Klaster 4

| Category | Column | Value | Relative Importance |
|-----------|--------------|-------------------|---------------------|
| Klaster 4 | Asal Sekolah | Missing | |
| Klaster 4 | Kecamatan | Missing | |
| Klaster 4 | Propinsi | Jawa Timur | |
| Klaster 4 | Propinsi | Jawa Barat | |
| Klaster 4 | Propinsi | Sulawesi Selatan | |
| Klaster 4 | Propinsi | Kalimantan Tengah | |
| Klaster 4 | Propinsi | Bali | |
| Klaster 4 | Propinsi | DKI Jakarta | |
| Klaster 4 | Kecamatan | Rantepao | |
| Klaster 4 | Kecamatan | Jekan Raya | |
| Klaster 4 | Propinsi | Sumatera Selatan | |

Klaster 4 tabel 8, nilai didominasi nilai yang missing, ini dikarenakan data mentah belum sempurna dalam melakukan cleaning data, sehingga masih banyak nama sekolah yang tidak konsisten yang menghasilkan nilai missing. Data yang ditunjukkan pada Tabel 7, terdapat *Missing Value* untuk Asal sekolah dan Kecamatan. *Missing Value* terjadi karena kategori asal sekolah dan kecamatan terlalu banyak variasi nilai, atau nilai tidak bisa memusat dikarenakan tidak banyak *variable* yang bernilai sama. *Cluster 4* berisi nilai yang sangat kecil dan tidak dominan, jika dibandingkan dengan anggota *cluster* yang lain. Selain *Missing value*, untuk propinsi perpusat pada wilayah Jawa Timur, Jawa barat dan Sulawesi Selatan serta Kalimantan Tengah. Tidak ada nilai yang dominan untuk *variable* yang lain pada *cluster 4*. Tidak ada rekomendasi bagi kegiatan promosi untuk *cluster 4*.

Tabel 9. Klaster 5

| Category | Column | Value | Relative Importance |
|-----------|--------------|------------------------|---------------------|
| Klaster 5 | Kecamatan | Missing | |
| Klaster 5 | Asal Sekolah | SMA Yos Soedarso, Pati | |
| Klaster 5 | Propinsi | Jawa Tengah | |
| Klaster 5 | Kecamatan | Pati | |
| Klaster 5 | Pilihan 1 | TEKNIK INFORMATIKA | |
| Klaster 5 | Asal Sekolah | SMA Masehi, Kudus | |
| Klaster 5 | Propinsi | Papua | |
| Klaster 5 | Asal Sekolah | SMA Negeri 1, Papua | |
| Klaster 5 | thn_daftar | 2009 | |
| Klaster 5 | Kecamatan | Kota Kudus | |

Klaster 5 tabel 9, didominasi nilai missing untuk kategori kecamatan. terlalu banyak kecamatan dan tidak ada kesamaan, maka kecamatan-kecamatan yang tidak memiliki kelas yang sama akan mengelompok sendiri dan bernilai *missing*. Untuk asal sekolah, terpusat pada SMA Yos Soedarso Pati, Jawa Tengah. Dalam *cluster 5* ini didominasi oleh pendaftar yang berasal dari Propinsi Jawa Tengah dan dari Papua. Untuk Propinsi Papua, SMA yang memberi kontribusi utama adalah dari SMA Negeri 1 Papua.

Tabel 10. Kluster 6

| Category | Column | Value | Relative Importance |
|-----------|--------------|-----------------------------------|---------------------|
| Kluster 6 | Asal Sekolah | SMA Krist. Satya Wacana, Salatiga | |
| Kluster 6 | Propinsi | Jawa Tengah | |
| Kluster 6 | Kecamatan | Sidorejo | |
| Kluster 6 | Kecamatan | Tingkir | |
| Kluster 6 | Asal Sekolah | SMA Negeri 1, Salatiga | |
| Kluster 6 | Kecamatan | Ungaran | |
| Kluster 6 | Pilihan 1 | AKUNTANSI | |
| Kluster 6 | Asal Sekolah | SMA Negeri 1, Ungaran | |
| Kluster 6 | Kecamatan | Sidomukti | |
| Kluster 6 | Pilihan 1 | DESAIN KOMUNIKASI VISUAL | |
| Kluster 6 | Asal Sekolah | SMA Kristen 1, Salatiga | |

Kluster 6 tabel 10, didominasi oleh asal sekolah yaitu SMA Kristen Satya Wacana. Persebaran calon mahasiswa di Program Studi Akuntansi, Desain Komunikasi Visual, Manajemen, Pendidikan Bahasa Inggris, dimana asal sekolah berasal dari sekolah di Salatiga, adalah SMA Kristen Satya Wacana, SMA Negeri 1 Salatiga, SMA Kristen 1 Salatiga, dan SMA Negeri 3 Salatiga. SMA Negeri 1 Ungaran juga turut menyumbang calon mahasiswa baru untuk kota terdekat dari kota Salatiga. Dapat disimpulkan bahwa SMA Kristen Satya Wacana adalah penyumbang tertinggi untuk program-program studi tersebut.

Tabel 11. Kluster 7

| Category | Column | Value | Relative Importance |
|-----------|--------------|-------------------------------|---------------------|
| Kluster 7 | Propinsi | Jawa Tengah | |
| Kluster 7 | Asal Sekolah | SMA Regina Pacis, Surakarta | |
| Kluster 7 | Asal Sekolah | SMA S. Sapientiae, Semarang | |
| Kluster 7 | Kecamatan | Banjarsari | |
| Kluster 7 | Asal Sekolah | SMA YSKI, Semarang | |
| Kluster 7 | Kecamatan | Pedurungan | |
| Kluster 7 | Asal Sekolah | SMA St. Bernardus, Pekalongan | |
| Kluster 7 | Kecamatan | Jebres | |
| Kluster 7 | Kecamatan | Semarang Utara | |
| Kluster 7 | Asal Sekolah | SMA Karangturi, Semarang | |
| Kluster 7 | Kecamatan | Semarang Timur | |
| Kluster 7 | Asal Sekolah | SMA Krista Mitra, Semarang | |

Kluster 7 pada tabel 11 didominasi propinsi yaitu Jawa Tengah, Asal propinsi, semua berasal dari Jawa Tengah. SMA yang mendominasi pada cluster 7 ini adalah SMA Regina Pacis Surakarta, SMA S. Sapientiae Semarang, serta kecamatan adalah Banjarsari. Dalam cluster 7, terdapat cluster calon mahasiswa yang berminat untuk memilih program studi Akuntansi sebagai pilihan 1, dan program studi manajemen sebagai pilihan 2.

Tabel 12. Kluster 8

| Category | Column | Value | Relative Importance |
|-----------|--------------|---------------------|---------------------|
| Kluster 8 | Propinsi | Nusa Tenggara Timur | |
| Kluster 8 | Propinsi | Maluku | |
| Kluster 8 | Kecamatan | Sirimau | |
| Kluster 8 | Kecamatan | Kota Soe | |
| Kluster 8 | Asal Sekolah | SMA Negeri 1, Ambon | |
| Kluster 8 | Asal Sekolah | SMA Negeri 2, Ambon | |
| Kluster 8 | Kecamatan | Nusanire | |
| Kluster 8 | Asal Sekolah | SMA Negeri 1, Soe | |
| Kluster 8 | Kecamatan | Teluk Ambon Baguala | |
| Kluster 8 | Propinsi | Lampung | |
| Kluster 8 | Kecamatan | Kota Waingapu | |

Kluster 8 tabel 12 didominasi oleh propinsi yang berasal dari Nusa Tenggara Timur. Data pada cluster 8,

menunjukkan bahwa propinsi Nusa Tenggara Timur (NTT) dan propinsi Maluku menjadi pusat cluster. Asal SMA dari mahasiswa pendaftar yang berasal dari kedua propinsi tersebut adalah berasal dari SMA Negeri 1 Ambon, SMA Negeri 2 Ambon, SMA Negeri 1 Soe, dan SMA-SMA lain yang tidak terlalu mendominasi cluster 8 ini. Sekolah-sekolah yang berasal dari luar Jawa, yaitu dari NTT, Maluku, Bandar Lampung, dengan pilihan utama adalah program studi Teologi, dan Keperawatan.

Tabel 13. Kluster 9

| Category | Column | Value | Relative Importance |
|-----------|--------------|------------------------------|---------------------|
| Kluster 9 | Asal Sekolah | SMA Negeri 1, Ambarawa | |
| Kluster 9 | Asal Sekolah | SMA Virgo Fidelis, Bawen | |
| Kluster 9 | Asal Sekolah | SMA Islam Sudirman, Ambarawa | |
| Kluster 9 | Asal Sekolah | SMA Negeri 2, Ungaran | |
| Kluster 9 | Asal Sekolah | SMA Negeri 1, Ungaran | |
| Kluster 9 | Asal Sekolah | SMA Kartika IV-1, Banyubiru | |
| Kluster 9 | Asal Sekolah | SMA Sedes Sapientiae, Jambu | |
| Kluster 9 | Asal Sekolah | SMA Kristen, Purwodadi | |
| Kluster 9 | Asal Sekolah | SMA Negeri 2, Temanggung | |
| Kluster 9 | Asal Sekolah | SMA Negeri 1, Purwodadi | |
| Kluster 9 | Asal Sekolah | SMA Negeri 1, Gubug | |
| Kluster 9 | Asal Sekolah | STM Kristen, Salatiga | |

Tabel 14. Kluster 10

| Category | Column | Value | Relative Importance |
|------------|--------------|--------------------------|---------------------|
| Kluster 10 | Asal Sekolah | Missing | 100 |
| Kluster 10 | Propinsi | Jawa Tengah | 48 |
| Kluster 10 | Kecamatan | Banyuwani | 34 |
| Kluster 10 | Kecamatan | Missing | 30 |
| Kluster 10 | Asal Sekolah | SMA Tarakanita, Magelang | 28 |
| Kluster 10 | Kecamatan | Magelang Selatan | 27 |
| Kluster 10 | Asal Sekolah | SMA Negeri 4, Semarang | 26 |
| Kluster 10 | Kecamatan | Muntian | 22 |
| Kluster 10 | Kecamatan | Magelang Utara | 17 |
| Kluster 10 | Pilihan 1 | TEKNIK ELEKTRO | 16 |
| Kluster 10 | Kecamatan | Martoyudan | 16 |

Kluster 9 Tabel 13 dan kluster 10 tabel 14 didominasi oleh propinsi Jawa Tengah Kluster 9 didominasi oleh asal sekolah di wilayah Jawa Tengah daerah Salatiga dan sekitarnya, yaitu wilayah Ambarawa, Bawen, Ungaran, Banyubiru, Purwodadi dan Temanggung. Kluster 9 berisi SMA-SMA wilayah Salatiga dan sekitarnya.

Tabel 15. Kluster 11

| Category | Column | Value | Relative Importance |
|------------|--------------|-----------------------|---------------------|
| Kluster 11 | Propinsi | Kalimantan Barat | 100 |
| Kluster 11 | Kecamatan | Sengah Temila | 35 |
| Kluster 11 | thn_daftar | 2010 | 29 |
| Kluster 11 | Kecamatan | Ngabang | 27 |
| Kluster 11 | Pilihan 1 | S1 PEND.GURU SD | 20 |
| Kluster 11 | Pilihan 2 | - | 20 |
| Kluster 11 | Asal Sekolah | SMA Negeri 1, Ngabang | 14 |
| Kluster 11 | Asal Sekolah | SMA Maniomas, Ngabang | 10 |

Kluster 11 Tabel 15 berisi data asal sekolah yang berasal dari Kalimantan Barat yang sebagian besar masuk ke Program Studi S1 PGSD, Kecamatan Ngabang.

Kesimpulan analisa dari kluster 1 sampai dengan kluster 11 adalah sebagai berikut :

- a. Propinsi yang paling banyak menyumbang calon mahasiswa adalah Propinsi Jawa Tengah untuk wilayah Jawa

- b. Propinsi di luar Jawa yang dominan adalah Kalimantan Barat, Maluku, dan Nusa Tenggara Timur
- c. Asal sekolah dari Jawa Tengah yang mendominasi adalah dari SMA Kristen Satya Wacana, SMA Negeri 1 Salatiga, SMA Negeri 2 Salatiga, SMA Negeri 3 Salatiga, SMA Negeri 1 Tengaran, SMA Kristen 1 Salatiga, SMA N 1 Ambarawa, SMA di wilayah Semarang, Surakarta dan Pekalongan
- d. Program studi yang paling banyak dituju adalah program studi Pendidikan Guru Sekolah Dasar, Teknik Informatika, S1 Akuntansi

V. SIMPULAN DAN SARAN

A. Simpulan

Dengan menggunakan *Microsoft Excel dengan teknologi Data Mining yaitu SQL Server 2008 Data Mining Add In* dapat ditemukan pola persebaran calon mahasiswa yang mendaftar di UKSW. Pola-pola yang terbentuk menggambarkan profil dari data pendaftar antara lain karakteristik asal daerah dengan program studi yang dituju. Pola-pola tersebut sangat membantu para pengambil keputusan untuk menentukan strategi selanjutnya berdasarkan pola-pola yang terbentuk.

Hasil *clustering* dapat dijadikan bahan pertimbangan rencana promosi di waktu yang akan datang terhadap daerah-daerah yang berpotensi memberikan kontribusi calon mahasiswa baru. Dengan diketahui daerah-daerah yang berpotensi, maka diharapkan kegiatan promosi lebih efektif dan lebih banyak menjangkau calon mahasiswa baru.

Dari sisi biaya promosi, melakukan promosi di daerah yang sudah diketahui potensinya, akan lebih menguntungkan jika dibandingkan di daerah-daerah yang belum diketahui potensinya, sehingga efektifitas biaya akan tercapai.

Secara umum, propinsi Jawa Tengah adalah penyumbang calon mahasiswa tertinggi, dimana Salatiga yang tertinggi calon mahasiswa yang mendaftar dan SMA Kristen Satya Wacana adalah penyumbang terbesar. Faktor kedekatan baik dari sisi jarak maupun organisasi juga mempengaruhi pertimbangan dalam pemilihan Perguruan Tinggi.

Selain pertimbangan umum tersebut, analisis berdasarkan data pada *cluster 1*, menunjukkan bahwa Program Studi S1 Pendidikan Guru Sekolah Dasar (PGSD), masih menjadi Progdil pilihan pertama terbesar dan berasal dari propinsi Jawa Tengah, bagian utara, antara lain wilayah Pati, Grobogan, Wonosobo, Purwodadi dan daerah sekitarnya. *Cluster 2* Program Studi S1 PGSD dan program-program studi kependidikan menjadi pilihan terbanyak tetapi berasal

dari Kota Salatiga dan sekitarnya, antara lain dari Tengaran. Berdasarkan asal SMA pada *Cluster 3* didominasi oleh SMA Kristen 1, Salatiga. Sedangkan propinsi yang dominan adalah Jawa Tengah dan kecamatan yang dominan adalah Sidorejo dan Argomulyo. Selain SMA Kristen 1 Salatiga, SMA lain yang dominan adalah SMA Negeri 2 dan SMA Negeri 1 Salatiga. Sampai pada analisis *cluster 3*, faktor jarak masih menjadi pertimbangan utama. Program studi PGSD terpusat pada dua wilayah yaitu Jawa Tengah dan Kalimantan Barat. Untuk program-program studi yang menjadi favorit lainnya, seperti Teknik Informatika, Bahasa Inggris, Akutansi, Psikologi, Ilmu Keperawatan belum memusat pada daerah tertentu yang mendominasi.

Program studi Akuntansi, Manajemen, Elektro, Desain Komunikasi Visual, didominasi oleh pendaftar yang berasal dari SMA-SMA di wilayah Salatiga, Surakarta, Semarang, dan Pekalongan. Kegiatan promosi difokuskan di SMA Kristen Satya Wacana, SMA Regina Pacis Surakarta, SMA S. Sapientiae Semarang, SMA YSKI Semarang,

Berdasarkan rangkuman hasil *cluster* serta analisis, dapat direkomendasikan strategi promosi sebagai berikut :

1. Untuk efektifitas biaya, kegiatan promosi difokuskan di sekolah-sekolah yang sudah terbukti memberikan kontribusi, dimana informasi terkait didapat dari data tahun-tahun sebelumnya.
2. *Cluster* dimana pilihan pertama PGSD dominan, maka pilihan 2 kosong juga dominan. Hal ini menandakan calon mahasiswa yang mendaftar di progdi PGSD tidak memilih pilihan progdi lain, dan menunjukkan tingkat keseriusan yang tinggi terhadap pilihan progdi, serta juga meningkatkan kepastian untuk melakukan daftar ulang. Progdi PGSD sebaiknya melakukan pemerataan daerah, sehingga tidak terpusat pada Jawa Tengah bagian utara dan Kalimantan Barat saja.
3. Sekolah-sekolah Islam, sebaiknya diberikan perhatian dalam kegiatan promosi. Berdasarkan data untuk wilayah Jawa Tengah pada tahun 2011, terdapat 259 orang pendaftar dari sekolah Islam, sementara *promotion kit* yang dikirimkan untuk sekolah-sekolah Islam hanya untuk 4 sekolah. Jika semua sekolah-sekolah Islam dijadikan tujuan promosi, maka akan lebih banyak jumlah calon pendaftar.
4. Asal sekolah dari propinsi di Luar Jawa yang tingkat responnya tinggi, antara lain : SMA Negeri 1 Papua, SMA Negeri 1 Ambon, SMA Negeri 2 Ambon, SMA Negeri 3 Ambon, SMA

Negeri 1 Soe, SMA Maniamas Ngabang, SMA Imanuel Bandar Lampung, SMA Negeri 1 Kupang, SMA Negeri 1 Waikabubak, SMA Negeri 1 Waingapu, SMA Negeri 1 Rantepao, SMA Negeri 4 Ambon, SMA Kristen Waikabubak, dan SMA Xaverius Bandar Lampung. Kegiatan promosi dapat lebih difokuskan pada SMA-SMA tersebut, untuk pertimbangan efektifitas biaya promosi.

5. Untuk pemerataan pilihan program studi, tim promosi dapat mempromosikan program-program studi yang kurang diminati di daerah tertentu, untuk dipromosikan di daerah lain yang kontribusinya besar.
6. Setiap program Studi memiliki pangsa pasarnya sendiri. Program studi yang secara umum adalah favorit, belum tentu favorit juga di daerah tertentu. Hal ini terjadi pada *cluster* 8, dimana pendaftar progdi S1 Teologi, melebihi pendaftar progdi S1 Teknik Informatika. Peluang ini harus dimanfaatkan oleh progdi Teologi untuk memperhatikan wilayah tersebut sebagai pangsa pasarnya, yaitu didominasi oleh SMA Negeri 1 dan SMA Negeri 2 Ambon.

B. Saran

Pada beberapa *cluster*, masih terdapat *missing value*. *Missing value* merupakan salah satu penyebab suatu data dikatakan tidak valid atau masih merupakan data 'kotor'. *Missing value* dengan jumlah banyak, akan menghilangkan sejumlah informasi yang seharusnya dibutuhkan untuk diolah. Salah satu metode penanganan yang dapat dilakukan pada *missing value* adalah dengan memasukkan suatu nilai kepada data-data yang *missing*. Nilai untuk mengisi *missing value* diperoleh dengan melakukan prediksi terhadap informasi yang masih tersedia pada data. Metode ini disebut dengan metode *imputasi*. Salah satu metode imputasi yaitu *K-Means Imputation* (KMI). Metode ini adalah metode yang didasarkan pada algoritma *K-Means* yang biasa digunakan dalam proses *clustering*. KMI akan melakukan pengisian *missing value* dengan menghitung rata-rata nilai atribut dari data-data yang berada pada suatu *cluster*, dimana data *missing* juga berada; dan data-data tersebut haruslah memiliki kelas yang sama dengan yang dimiliki oleh data *missing*.

UCAPAN TERIMA KASIH

Publikasi artikel penelitian ini bisa dittulis dan dipublikasikan, karena bantuan dari berbagai pihak antara lain Fakultas Teknologi Informasi, Biro Administrasi Akademik, dan Biro Promosi dan Hubungan Luar Universitas Kristen Satya Wacana Salatiga.

VI. Daftar Pustaka

- [1] Azimah, Ariana, & Sucahyo Yudho Giri, Oktober 2007. *Penggunaan Data Warehouse dan Data Mining untuk Data Akademik : sebuah studi kasus pada Universitas Nasional*. Jurnal Sistem Informasi MTI UI Vol.3-No.2.
- [2] Han, Jiawei, & Kamber, Micheline. 2006. *Data Mining : Concept and Techniques*, 2nd ed. Elsevier Publication.
- [3] Kasim, Wirama,; Sudiyanto, Hendrik, dan Hermawan, Yudhi. 2010. *The Essensial Business Intelligence in Microsoft SQL Server 2008*. Indonesia .NET Developer Community
- [4] Ling, C. X. dan Li, C., 1998. *Data Mining for Direct Marketing: Problems and Solutions*, in Proceeding of the Fourth international conference on Knowledge discovery and Data Mining.
- [5] Moekijat. 2000. *Manajemen Pemasaran. Bandung* : Penerbit mandar Maju
- [6] M. Khattak, A. M. Khan, Sungyoung Lee, dan Young-Koo Lee, *Analyzing Association Rule Mining and Clustering on Sales Day Data with XLMiner and Weka*. International Journal of Database Theory and Application. Vol. 3, No. 1, March, 2010.
- [7] P.-N. Tan, M. Steinbach, and V. Kumar, Introduction to Data Mining, (First Edition). Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 2005.
- [8] Rankuti, Freddy. 2009. Strategi Promosi yang kreatif dan analisis kasus Integrated marketing. Jakarta : Gramedia Pustaka Utama.
- [9] Turban, E., dkk 2005. Decision Support System and Intelligence Systems. Yogyakarta : Andi Offset.
- [10] Susanto, Sani, & Suryadi, Dedy. 2010. *Pengantar Data Mining : Menggali Pengetahuan Dari Bongkahan Data*, Jogjakarta: Penerbit Andi.
- [11] Wilarso, Lik. 2008. *Pemanfaatan Data Warehouse di Perguruan Tinggi Indonesia*. Jurnal Sistem Informasi MTI UI Vol.4-No.1.
- [12] [Microsoft Clustering Algorithm. http://msdn.microsoft.com/en-us/library/ms174879.aspx](http://msdn.microsoft.com/en-us/library/ms174879.aspx), diakses tanggal 1 Mei 2012

- [13] Pangkalan data Pendidikan Tinggi.
[http://forlap.dikti.go.id/peguruantinggi/homegr
aphpt](http://forlap.dikti.go.id/peguruantinggi/homegr
aphpt) diakses tanggal 31 Agustus 2016