

# Peran *Bussines Intelligence* Dalam Menentukan Strategi Promosi Penerimaan Mahasiswa Baru

Ika Kurniawati<sup>1</sup>, Richardus Eko Indrajit<sup>2</sup>, Muh. Fauzi<sup>3</sup>

<sup>1</sup>STMIK Nusa Mandiri, <sup>2</sup>ABFI Institut Perbanas, STMIK Bumigora Mataram<sup>3</sup>  
Jakarta<sup>1,2</sup>, Mataram<sup>3</sup> - Indonesia  
[contact.ikakurniawati@yahoo.com](mailto:contact.ikakurniawati@yahoo.com)<sup>1</sup>, [indrajit@post.harvard.edu](mailto:indrajit@post.harvard.edu)<sup>2</sup>, [mr.vawzy@gmail.com](mailto:mr.vawzy@gmail.com)<sup>3</sup>

## ABSTRAK

Institut Sains dan Teknologi Al-Kamal (ISTA) menggunakan sistem penerimaan mahasiswa baru *online* berusaha untuk menarik siswa-siswi SMA dan sederajat untuk menjadi mahasiswa di ISTA. Permasalahan yang terjadi adalah masih besarnya persentase calon mahasiswa yang tidak melakukan registrasi ulang padahal sudah dinyatakan lulus seleksi sehingga target kuota mahasiswa baru tidak tercapai. Penelitian ini bertujuan untuk melakukan mengelompokan data mahasiswa ke dalam beberapa *cluster* berdasarkan kemiripan dari data tersebut. Sehingga data mahasiswa yang memiliki karakteristik yang sama dikelompokkan dalam satu *cluster* dan yang memiliki karakteristik berbeda dikelompokkan dalam *cluster* yang lain, dengan memanfaatkan proses data mining yaitu metode *smart clustering k-means*. Sistem *Bussines intelligence* pada penelitian ini berperan untuk meningkatkan keunggulan kompetitif institusi melalui pendayagunaan berbagai data, informasi dan pengetahuan yang dimiliki oleh institusi sebagai bahan dalam proses pengambilan keputusan. Metode yang digunakan adalah CRISP-DM dengan proses *business understanding, data understanding, data preparation, modeling, evaluation* dan *deployment*. Pengujian dilakukan menggunakan tool *RapidMiner 7.3* untuk membantu menemukan hasil yang akurat untuk memecahkan masalah tersebut. Hasil dari penelitian ini digunakan sebagai salah satu dasar pengambilan keputusan untuk menentukan strategi promosi berdasarkan *cluster* yang terbentuk oleh pihak marketing ISTA.

**Kata Kunci:** *Bussines intelligence, clustering, k-means, strategi promosi*

## 1. PENDAHULUAN

Pendidikan merupakan salah satu aspek yang penting dalam pembangunan bangsa, keberhasilan pembangunan negara adalah dengan tersedianya penduduk yang terdidik dalam jumlah, jenis, dan tingkat yang memadai. Dengan pendidikan seseorang menjadi cerdas, memiliki kemampuan, sikap hidup yang lebih baik dan pada akhirnya dapat meningkatkan taraf hidupnya ke jenjang yang lebih tinggi. Dalam upaya meningkatkan kualitas suatu bangsa, menurut UNESCO (*United Nations, Educational, Scientific and Cultural Organization*) tidak ada cara lain kecuali melalui peningkatan mutu pendidikan. Untuk mencapai tujuan pendidikan yang terarah, diperlukan suatu lembaga pendidikan yang berjenjang dimulai dari pendidikan tingkat dasar, pendidikan menengah dan pendidikan tinggi.

Penyelenggaraan pendidikan tinggi melalui perguruan tinggi dimaksudkan untuk meningkatkan

kemampuan akademik dan keahlian (profesional), siswa-siswi sekolah menengah (SMA, SMK, dan MA) dapat melanjutkan ke perguruan tinggi untuk meningkatkan kemampuan akademik dan keahlian mereka. Perguruan tinggi terbagi menjadi dua, yaitu Perguruan Tinggi Negeri (PTN) dan Perguruan Tinggi Swasta (PTS). Perguruan tinggi negeri dikelola oleh pemerintah sedangkan perguruan tinggi swasta dikelola oleh instansi tertentu (yayasan).

Institut Sains dan Teknologi Al-Kamal (ISTA) Salah satu perguruan tinggi swasta yang dikelola oleh Yayasan Pondok Pesantren Al-Kamal Jakarta berdiri sejak tahun 1989 yang terletak di Jakarta Barat, ISTA memiliki 2 fakultas dan 8 program studi. Untuk masuk menjadi civitas akademika ISTA, siswa-siswi sekolah menengah (SMA, SMK, dan MA) dapat mengikuti seleksi masuk yang diselenggarakan oleh ISTA.

Seleksi masuk mahasiswa di ISTA lebih dikenal dengan istilah Penerimaan Mahasiswa Baru (PMB),

kegiatan PMB adalah salah satu kegiatan rutin tahunan sebagai media untuk merekrut calon mahasiswa baru sesuai dengan kriteria yang ditetapkan oleh ISTA dengan tujuan agar calon mahasiswa yang lulus seleksi atau diterima menjadi mahasiswa dapat menyelesaikan studi tepat waktu sesuai dengan program studi yang mereka tempuh.

Permasalahan yang terjadi dalam perekrutan mahasiswa adalah banyaknya jumlah calon mahasiswa yang tidak melakukan registrasi ulang untuk setiap tahunnya padahal peserta seleksi telah dinyatakan lulus. Hal ini menjadi masalah yang harus diselesaikan oleh pihak kampus, yang berarti kuota mahasiswa baru untuk setiap program studi belum tercapai. Salah satu penyebab terjadinya penurunan jumlah mahasiswa baru pada tahun ini adalah kurang dilakukannya pengolahan data mahasiswa secara tepat berdasarkan data historis oleh pihak admisi ISTA. Hal tersebut dapat mempengaruhi pengambilan keputusan dalam menentukan wilayah promosi yang tepat sasaran. Pengolahan data mahasiswa seharusnya dilakukan agar dapat menentukan wilayah promosi yang tepat sasaran sehingga tidak terjadi penurunan jumlah mahasiswa pada tahun berikutnya.

Dalam mengatasi permasalahan tersebut, peran *Bisnis Intelligence* (BI) dapat membantu suatu organisasi atau perusahaan mendapatkan pengetahuan yang jelas mengenai faktor-faktor yang mempengaruhi kinerja organisasi sehingga dapat membantu organisasi dalam pengambilan keputusan serta sekaligus meningkatkan keunggulannya (*competitive advantage*). Dengan kegiatan BI tersebut maka sebuah organisasi atau perusahaan akan dengan mudah dalam mengambil keputusan secara cepat dan tepat.

Dalam penelitian ini BI digunakan untuk mengelompokkan data mahasiswa dengan memanfaatkan proses data mining menggunakan metode *smart clustering* yaitu *k-means*. *K-means* merupakan salah satu metode data *non-hierarchical clustering* yang dapat mengelompokkan data mahasiswa ke dalam beberapa *cluster* berdasarkan kemiripan dari data tersebut, sehingga data mahasiswa yang memiliki karakteristik yang sama dikelompokkan dalam satu *cluster* dan yang memiliki karakteristik berbeda dikelompokkan dalam *cluster* yang lain. Atribut yang digunakan dalam menentukan pengelompokkan wilayah untuk melakukan promosi di ISTA adalah nama mahasiswa, NIM, kota asal, nilai IPK dan program studi.

Hasil penelitian dengan menerapkan metode *k-means* diharapkan mampu membantu pihak Admisi ISTA untuk mengambil keputusan dalam menentukan wilayah promosi yang tepat sasaran sehingga tidak terjadi penurunan jumlah mahasiswa pada tahun berikutnya dan juga untuk mendapatkan nilai akurasi dari kinerja metode *k-means*.

## 2. LANDASAN TEORI

### *Bussines Intelligence*

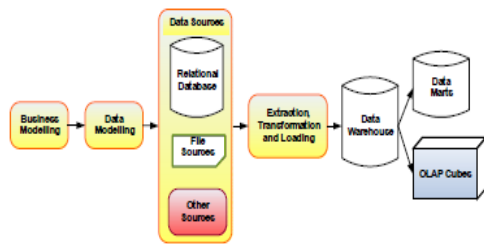
*Business Intelligence* (BI) adalah sebuah proses untuk meningkatkan keunggulan kompetitif perusahaan melalui pendayagunaan berbagai data, informasi, dan pengetahuan (*knowledge*) yang dimiliki oleh perusahaan sebagai bahan baku dalam proses pengambilan keputusan. Berbeda dengan sejumlah aplikasi lain dengan tujuan serupa yang lebih dahulu diperkenalkan sebelumnya, konsep BI menekankan pada penerapan 5 pendayagunaan informasi untuk keperluan spesifik bisnis, masing-masing adalah sebagai berikut.

*Pertama, data sourcing*. Berkaitan dengan kemampuan sistem untuk mengakses berbagai data dan informasi dari sejumlah sumber dengan format yang berbeda. *Kedua, data analysis*. Berkaitan dengan kemampuan sistem untuk membantu proses penciptaan pengetahuan (*knowledge*) melalui aktivitas pengkajian data dan informasi yang dimiliki oleh perusahaan. *Ketiga, situation awareness*. Berkaitan dengan kemampuan sistem untuk mencari dan menyediakan data dan informasi terkait dengan kebutuhan atau konteks bisnis pada saat tertentu, misalnya ketika perusahaan berhadapan dengan peristiwa darurat dan mendesak. *Keempat, risk analysis*. Berkaitan dengan kemampuan sistem untuk melakukan kalkulasi rasio yang akan dihadapi perusahaan terhadap berbagai kecenderungan atau kemungkinan yang dapat terjadi sehubungan dengan kondisi tertentu. *Kelima, decision support*. Berkaitan dengan kemampuan sistem untuk secara aktif membantu manajemen dalam memberikan pertimbangan keputusan-keputusan yang berkualitas berdasarkan sejumlah kalkulasi dan pengolahan terhadap data atau informasi internal maupun eksternal yang dimiliki.

*Business Modelling* menggambarkan gambaran keseluruhan proses bisnis dan masalah-masalah yang harus diselesaikan dan diatasinya, dan juga menggambarkan bagaimana informasi mengalir pada proses-proses bisnis dari sumber informasi ke tujuan. Dalam membuat *business modelling* digunakan *business model* dan diagram yang memberikan informasi secara grafis bagi anggota suatu organisasi atau perusahaan memahami dan mengkomunikasikan *business rule* dan proses-proses bisnisnya.

*Business process modelling, process flow modelling* dan *data flow modelling* adalah sub-bagian dari suatu *business modelling*. *Business modelling strategy* dan *business model* suatu organisasi atau perusahaan berbeda satu dengan yang lainnya, hal ini tergantung pada kebutuhan dan tujuan perusahaan

atau organisasi tersebut. (Barberg, B. 2007).



Gambar 2.1. Business Intelligence Environment  
Sumber: www.learnbi.com (2007)

Seorang kepala biro marketing, dia dapat melakukan kampanye pemasaran dengan segmentasi target yang jelas dan dapat diperhitungkan dalam menyumbang penerimaan mahasiswa baru bagi institusinya dengan mencari strategi yang tepat dalam melakukan promosi, dan masih banyak lagi evaluasi kinerja manajemen dari setiap divisi yang dilakukan pada suatu organisasi atau perusahaan. Jelas sekali bahwa *Business Intelligence* diperlukan di setiap lini dalam suatu organisasi yang harusnya sudah tertanam dalam setiap langkah bisnis, guna mencapai tujuan perusahaan yang lebih baik, lebih efisien dan efektif sehingga mendapat *profit* yang optimal.

### Data Mining

*Data mining* adalah proses analisis terhadap data dengan penekanan menemukan informasi yang tersembunyi pada sejumlah besar data yang disimpan ketika menjalankan bisnis perusahaan. Teknik *data mining* merupakan implementasi yang khusus dengan algoritma, yang digunakan pada operasi *data mining*. Ada 6 (enam) teknik umum *data mining*, antara lain sebagai berikut.

1. *Association*, digunakan untuk mengenali kelakuan dari kejadian-kejadian khusus atau proses. Link asosiasi muncul pada setiap kejadian.
2. *Sequence*, mirip dengan asosiasi namun menghubungkan kejadian-kejadian sepanjang waktu dan menentukan keterhubungan antar item untuk sepanjang waktu.
3. *Classification*, melihat pada kelakuan dan atribut dari kelompok yang telah didefinisikan. Tool *data mining* dapat memberikan klasifikasi pada data baru dengan memanipulasi data yang ada, yang telah diklasifikasi dan dengan menggunakan hasilnya untuk memberikan sejumlah aturan. Aturan-aturan tersebut digunakan pada data-data baru untuk diklasifikasi. Teknik ini menggunakan *supervised induction*, yang memanfaatkan kumpulan pengujian dari *record* yang terklasifikasi untuk menentukan kelas-kelas tambahan.
4. *Cluster*, dapat digunakan untuk menganalisis

pengelompokan berbeda terhadap data. Mirip dengan klasifikasi, namun pengelompokan belum didefinisikan sebelum dijalankannya tool *data mining*. Biasanya menggunakan metode *neural network* atau statistik. Clustering membagi item menjadi kelompok-kelompok berdasarkan yang ditemukan tool *data mining*.

5. *Regression (forecasting)*, menggunakan nilai dari data yang diketahui untuk memperkirakan nilai di masa depan atau kejadian masa depan berdasarkan kecenderungan sejarah dan statistik.
6. *Time series (forecasting)*, perbedaan dengan regresi adalah bahwa *time series* hanya memperkirakan data yang bergantung pada waktu.

### Clustering

Pada dasarnya clustering merupakan suatu metode untuk mencari dan mengelompokkan data yang memiliki kemiripan karakteristik (*similarity*) antara satu data dengan data yang lain. *Clustering* merupakan salah satu metode *data mining* yang bersifat tanpa arahan (*unsupervised*), maksudnya metode ini diterapkan tanpa adanya latihan (*training*) dan tanpa ada guru (*teacher*) serta tidak memerlukan target output. Dalam data mining ada dua jenis metode clustering yang digunakan dalam pengelompokan data, yaitu *hierarchical clustering* dan *non-hierarchical clustering* (Santosa, 2007).

*Hierarchical clustering* adalah suatu metode pengelompokan data yang dimulai dengan mengelompokkan dua atau lebih objek yang memiliki kesamaan paling dekat. Kemudian proses diteruskan ke objek lain yang memiliki kedekatan kedua. Demikian seterusnya sehingga cluster akan membentuk semacam pohon dimana ada hierarki (tingkatan) yang jelas antar objek, dari yang paling mirip sampai yang paling tidak mirip. Secara logika semua objek pada akhirnya hanya akan membentuk sebuah *cluster*. Dendogram biasanya digunakan untuk membantu memperjelas proses hierarki tersebut (Santoso, 2010).

Berbeda dengan metode *hierarchical clustering*, metode *non-hierarchical clustering* justru dimulai dengan menentukan terlebih dahulu jumlah cluster yang diinginkan (dua cluster, tiga cluster, atau lain sebagainya). Setelah jumlah cluster diketahui, baru proses cluster dilakukan tanpa mengikuti proses hierarki. Metode ini biasa disebut dengan *K-Means Clustering* (Santoso, 2010).

### Algoritma K-means Clustering

*K-means clustering* merupakan salah satu metode data clustering non-hirarki yang mengelompokkan

data dalam bentuk satu atau lebih cluster/kelompok. Data-data yang memiliki karakteristik yang sama dikelompokkan dalam satu *cluster*/kelompok dan data yang memiliki karakteristik yang berbeda dikelompokkan dengan *cluster*/kelompok yang lain sehingga data yang berada dalam satu *cluster*/kelompok memiliki tingkat variasi yang kecil (Agusta, 2007).

Menurut Santosa (2007), langkah-langkah melakukan clustering dengan metode *K-Means* adalah sebagai berikut:

- a. Pilih jumlah *cluster*  $k$ .
- b. Inisialisasi  $k$  pusat *cluster* ini bisa dilakukan dengan berbagai cara. Namun yang paling sering dilakukan adalah dengan cara random. Pusat-pusat *cluster* diberiduberi nilai awal dengan angka-angka random,
- c. Alokasikan semua data/ objek ke *cluster* terdekat. Kedekatan dua objek ditentukan berdasarkan jarak kedua objek tersebut. Demikian juga kedekatan suatu data ke *cluster* tertentu ditentukan jarak antara data dengan pusat *cluster*. Dalam tahap ini perlu dihitung jarak tiap data ke tiap pusat *cluster*. Jarak paling antara satu data dengan satu *cluster* tertentu akan menentukan suatu data masuk dalam *cluster* mana. Untuk menghitung jarak semua data ke setiap titik pusat *cluster* dapat menggunakan teori jarak Euclidean yang dirumuskan sebagai berikut:

$$D(i,j) = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + \dots + (x_{ki} - x_{kj})^2} \dots (1)$$

Dimana :

$D(i,j)$  = Jarak antara ke  $i$  ke pusat *cluster*  $j$

$X_{ki}$  = Data ke  $i$  pada atribut data ke  $k$

$X_{kj}$  = Titik pusat ke  $j$  pada atribut ke  $k$

- d. Hitung kembali pusat *cluster* dengan keanggotaan *cluster* yang sekarang. Pusat *cluster* adalah rata-rata dari semua data/ objek dalam *cluster* tertentu. Jika dikehendaki bisa juga menggunakan median dari *cluster* tersebut. Jadi rata-rata (mean) bukan satu-satunya ukuran yang bisa dipakai.
- e. Tugaskan lagi setiap objek memakai pusat *cluster* yang baru. Jika pusat *cluster* tidak berubah lagi maka proses *clustering* selesai. Atau, kembali ke langkah nomor 3 sampai pusat *cluster* tidak berubah lagi.

### Strategi Promosi

Menurut Kotler & Armstrong variabel-variabel yang ada di dalam *promotion mix* ada lima, yaitu :

- a. Periklanan (*advertising*)
- b. Penjualan Personal (*personal selling*)

- c. Promosi penjualan (*sales promotion*)
- d. Hubungan masyarakat (*public relation*)
- e. Pemasaran langsung (*direct marketing*)

Promosi adalah salah satu bagian dari marketing *mix* yang besar peranannya. Promosi merupakan suatu ungkapan dalam arti luas tentang kegiatan-kegiatan yang secara aktif dilakukan oleh perusahaan (penjual) untuk mendorong konsumen untuk membeli produk yang ditawarkan.

Dalam promosi dikenal istilah bauran promosi. Bauran promosi adalah kombinasi strategi yang paling baik dari variabel-variabel periklanan, personal *selling*, dan alat promosi lain, yang semuanya direncanakan untuk mencapai tujuan program penjualan. Promosi adalah salah satu faktor penentu keberhasilan suatu program pemasaran. Betapapun berkualitasnya suatu produk, bila konsumen belum pernah mengenal ataupun mendengarnya dan tidak yakin bahwa produk tersebut akan berguna bagi mereka, maka mereka tidak akan pernah membelinya. (Basu Swastha, 2000)

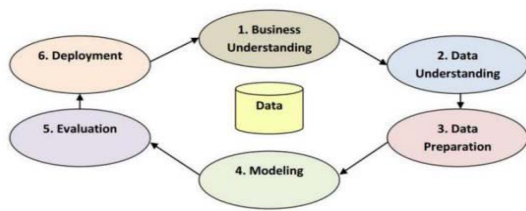
Demi tercapainya tujuan perusahaan, masing-masing perusahaan perlu menyadari faktor-faktor yang menentukan keberhasilan dan kegagalan dalam perjalanannya terhadap persaingan bisnis. Faktor-faktor tersebut bisa bersumber dari dalam perusahaan (*interen*) ataupun dari luar perusahaan (*exteren*). Dalam menyusun strategi pemasaran hendaknya memperhatikan lingkungan pemasaran. Lingkungan internal meliputi sumber daya yang dimiliki oleh masing-masing perusahaan baik dari kemampuan sumber daya manusia, peralatan mesin, kebijakan perusahaan dan sebagainya. Lingkungan eksternal mencakup kondisi konsumen, kebijakan pemerintah, persaingan antar perusahaan. Lingkungan-lingkungan tersebut secara berangsur akan selalu mengalami perubahan, sehingga perusahaan harus dapat mengambil sikap dari setiap perubahan pada elemen lingkungannya.

## 3. METODE PENELITIAN

### Metode Data Mining

Metode data mining yang digunakan dalam eksperimen ini menggunakan model *Cross-Standard Industry for Data Mining* (CRISP-DM) yang terdiri dari 6 fase, yaitu: (Sumathi, 2006)

1. Fase pemahaman bisnis (*Business Understanding*)
2. Fase pemahaman data (*Data Understanding*)
3. Fase pengolahan data (*Data Preparation*)
4. Fase pemodelan (*Modelling*)
5. Fase Evaluasi (*Evaluation*)
6. Fase Penyebaran (*Deployment*)



Gambar 3.1. Tahapan Metodologi CRISP-DM

### **Business Understanding**

Pada tahap ini berfokus pada pemahaman mengenai tujuan dari proyek dan kebutuhan secara persepektif bisnis, kemudian mengubah hal tersebut menjadi sebuah permasalahan *data mining* dan rencana awal untuk mencapai tujuan tersebut. Kegiatan yang dilakukan antara lain: menentukan tujuan dan persyaratan dengan jelas secara keseluruhan, menerjemahkan tujuan tersebut serta menentukan pembatasan dalam perumusan masalah *data mining*, dan selanjutnya mempersiapkan strategi awal untuk mencapai tujuan tersebut.

### **Data Understanding**

Pada tahap ini dilakukan pengumpulan terhadap data, lalu kemudian mempelajari data tersebut dengan tujuan untuk mengenal data, melakukan identifikasi dan mengetahui kualitas dari data, serta mendeteksi subset yang menarik dari data yang dapat dijadikan hipotesa bagi informasi yang tersembunyi.

### **Data Preparation**

Pada tahap ini dilakukan persiapan mengenai data yang akan digunakan pada tahap berikutnya. Kegiatan yang dilakukan antara lain: memilih kasus dan parameter yang akan dianalisis (*Select Data*), melakukan transformasi terhadap parameter tertentu (*Transformation*), dan melakukan pembersihan data agar data siap untuk tahap *modeling* (*Cleaning*).

Data *preprocessing* bertujuan untuk mendapatkan data yang bersih dan siap untuk digunakan dalam penelitian. Tahapan yang dikerjakan adalah melakukan pengabaian atribut pada data mentah yang dianggap tidak relevan dengan hasil pengujian dan perubahan terhadap nilai data bahkan tipe data pada atribut *dataset* dengan tujuan untuk mempermudah pemahaman terhadap isi *record* dengan memperhatikan konsistensi data, *missing value*, dan *redundancy* pada data.

### **Modeling**

Pada tahap ini dilakukan penentuan terhadap teknik *data mining*, alat bantu *data mining*, dan algoritma *data mining* yang akan diterapkan. Lalu selanjutnya adalah melakukan penerapan teknik dan algoritma *data mining* tersebut kepada data dengan bantuan alat bantu. Jika diperlukan penyesuaian data terhadap teknik data mining tertentu, dapat kembali ke tahap persiapan data.

### **Evaluation**

Pada tahap ini dilakukan pengujian terhadap model-model yang dikomparasi untuk mendapatkan informasi model yang paling akurat. Evaluasi dan validasi menggunakan metode *confusion matrix* dan kurva ROC.

### **Deployment**

Setelah pembentukan model dan dilakukan analisa dan pengukuran pada tahap sebelumnya, selanjutnya pada tahap ini diterapkan model yang paling akurat dengan memakai data baru diluar data *training* dan data *testing*.

## **4. HASIL DAN PEMBAHASAN**

### **Pemahaman Bisnis (Business Understanding)**

Pada tahap ini dilakukan pemahaman dari objek penelitian yang dituju, pemahaman ini mencakup antara lain tinjauan objek penelitian yaitu sistem penerimaan mahasiswa baru dan sistem akademik ISTA terutama sub sistem registrasi ulang, tujuan sistem diterapkan, identifikasi permasalahan yang timbul, alternatif pemecahan masalah yang ada dan tujuan diadakannya penelitian ini.

### **Pemahaman Data (Data Understanding)**

Pada tahap *Data Understanding* yang dilakukan adalah mengambil data mentah yang berasal dari sistem informasi akademik ISTA, sumber data ini dikelola oleh Biro layanan teknologi informasi ISTA. Data-data yang digunakan pada penelitian ini didapat dari bagian akademik yang berupa data mahasiswa yang telah lulus dari Universitas tersebut pada tahun 2016. Data yang digunakan adalah sebanyak 332 data mahasiswa yang berisi data diri dari mahasiswa yang telah lulus tersebut, namun dalam penelitian ini hanya beberapa atribut data saja yang digunakan, seperti nama mahasiswa, NIM, kota asal, nilai IPK dan program studi.

Data-data yang telah didapatkan pada tahap pengumpulan data kemudian dilakukan transformasi pada data-data yang berjenis data nominal, yaitu seperti kota asal dan jurusan. Data-data yang berjenis data nominal tersebut diinisialisasikan ke dalam bentuk angka melalui beberapa langkah agar data-data yang berjenis data nominal dapat diolah dengan menggunakan algoritma *K-means Clustering*.

### Persiapan Data (Data Preparation)

Tahap ini dilakukan penyiapan data atau *preprocessing* bertujuan untuk mendapatkan data yang bersih dan siap untuk digunakan dalam penelitian. Tahapan yang dikerjakan adalah melakukan pengabaian atribut pada data mentah yang dianggap tidak relevan dengan hasil pengujian dan perubahan terhadap nilai data bahkan tipe data pada atribut *dataset* dengan tujuan untuk mempermudah pemahaman terhadap isi *record* dengan memperhatikan konsistensi data, *missing value*, dan *redundancy* pada data.

Berikut ini adalah data mahasiswa yang telah lulus pada tahun 2016.

Tabel 1. Data Mahasiswa yang telah lulus dari ISTA Tahun 2016

NO	NIM	NAMA	KOTA	IPK	PROGRAM STUDI
1	201291001	M.Imam Khasan Taufik	Tegal	3.26	Desain Komunikasi Visual
2	200922016	Iswail	Cilacap	2.86	Teknik Informatika
3	201126016	Adi Fandika	Jakarta	3.61	Sistem Informasi
4	201224041	Sujarno	Jakarta	3.17	Teknik Industri
5	201125052	Yessi Angraini	Bogor	2.99	Teknik Kimia
6	201423032	Taufiq Dwi Setiawan	Solo	3.22	Teknik Mesin
7	201231032	Hidayah	Tasikmalaya	3.63	Teknik Sipil
8	201551094	Muhammad Faizon Nopisenle	Bengkulu	3.60	Farmasi
9	201026010	Adelia Roselina Rinjani	Tangerang	3.12	Farmasi
10	201251028	Dewi Lolyta Natalia M.	Jakarta	3.68	Farmasi

### Transformasi Data

Data yang berjenis nominal seperti kota asal dan program studi harus dilakukan proses inialisasi data terlebih dahulu ke dalam bentuk angka/numerikal. Untuk melakukan inialisasi kota asal dapat dilakukan dengan langkah-langkah sebagai berikut:

#### A. Inialisasi Kota Asal

1. Pada data kota asal terlebih dulu dilakukan pembagian wilayah yang menjadi beberapa bagian wilayah, yaitu:
  - a. Wilayah Sumatera yang terdiri dari kota Aceh Besar, Ambarita Samosir, Banda Aceh, Lhokseumawe, Medan, Padang, Palembang, Bengkulu, Jambi, Bandar Lampung, dan Pangkal Pinang.
  - b. Wilayah DKI Jakarta yang terdiri dari kota Jakarta.
  - c. Wilayah Jawa Barat yang terdiri dari kota, Bandung, Bekasi, Bogor, Tasikmalaya, Cikarang, Depok, Garut, Karawang, Pandeglang, Sukabumi, dan Tangerang.
  - d. Wilayah Jawa Tengah yang terdiri dari kota Semarang, Solo, Tegal, Yogyakarta, Purworejo, Kudus.
  - e. Wilayah Jawa Timur yang terdiri dari kota Bojonegoro, Blitar, Lumajang, Magetan, dan Surabaya.

2. Kemudian wilayah-wilayah tersebut diurutkan dari yang terbesar berdasarkan frekuensi mahasiswa yang berasal dari wilayah tersebut.
3. Setelah itu wilayah yang memiliki frekuensi terbesar diberi inisial dengan angka 1 dan wilayah yang memiliki frekuensi terbesar kedua diberi inisial dengan angka 2, begitu seterusnya hingga wilayah dengan frekuensi paling sedikit. Hasil dari inialisasi kota asal dapat dilihat pada tabel 2.

Tabel 2. Inialisasi Data Wilayah Kota Asal

Wilayah	Frekuensi	Inisial
DKI Jakarta	116	1
Jawa Barat	142	2
Jawa Tengah	31	3
Jawa Timur	6	4
Sumatera	4	5

#### B. Inialisasi Program Studi

Selain kota asal, jurusan juga termasuk ke dalam jenis data nominal sehingga perlu diinialisasikan ke dalam bentuk angka. Seperti pada kota asal, pada program studi juga diberikan inialisasi berdasarkan frekuensi jumlah mahasiswa pada program studi. Hasil dari inialisasi tersebut dapat dilihat pada tabel berikut.

Tabel 3. Inialisasi Data Program Studi

Program Studi	Frekuensi	Inisial
Farmasi	232	1
Teknik Kimia	26	2
Teknik Mesin	20	3
Teknik Sipil	14	4
Teknik Informatika	13	5

Teknik Industri	12	6
Sistem Informasi	11	7
Desain Komunikasi Visual	4	8

Berikut ini adalah contoh dataset mahasiswa yang telah dilakukan inisialisasi dan akan diolah dengan *Rapid Miner*.

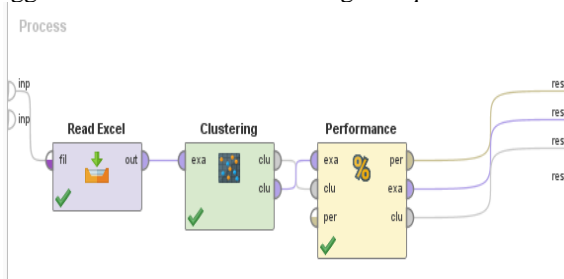
Tabel 4. Dataset Mahasiswa yang Telah dilakukan Inisialisasi

NO	NIM	KOTA	IPK	PROGRAM STUDI
1	201291001	3	3.26	8
2	200922016	3	2.86	5
3	201126016	1	3.61	7
4	201224041	1	3.17	6
5	201125052	2	2.99	2
6	201423032	3	3.22	3
7	201231032	2	3.63	4
8	201551094	5	3.60	1
9	201026010	2	3.12	1
10	201251028	1	3.68	1

### Pemodelan (*Modeling*)

Pemodelan adalah fase yang secara langsung melibatkan teknik *data mining* dengan melakukan pemilihan teknik *data mining* dan menentukan algoritma yang akan digunakan yaitu algoritma *k-means* yang akan diolah dengan *frame work Rapid Miner* versi 7.3.

Berikut ini adalah implementasi pengolahan data menggunakan metode *k-means* dengan *Rapid Miner*.



Gambar 4.1. Pemodelan *k-means* pada *Rapid Miner*

Gambar 4.2. *Result Overview k-means* pada *Rapid Miner*

Dengan menggunakan pemodelan *k-means clustering* seperti gambar 4.1 diatas, dengan inisialisasi jumlah *cluster* sebanyak 3 *cluster*, maka didapatkan hasil dengan *cluster* yang terbentuk adalah 3, sesuai dengan pendefinisian nilai *k* dengan jumlah *cluster\_0* ada 10 item, *cluster\_1* ada 147 item, *cluster\_2* ada 175 item, dengan total item sejumlah 332 item.

Tabel 5. Hasil *cluster K-Means* dalam implementasi *Rapid Miner*

Model Cluster	
Cluster 0	10 items
Cluster 1	147 items
Cluster 2	175 items
Total number of items	332 items

Berikut ini hasil perhitungan antara jarak *cluster* dengan *centroid* menggunakan *Rapid Miner*

Attribute	cluster_0	cluster_1	cluster_2
Nim	200917029.900	201447326.027	201232137.057
Kode kota	1.800	1.891	1.674
IPK	3.373	3.293	3.256
Kode Prodi	7.100	5.469	6.600

Gambar 4.3. Hasil Perhitungan Antara Jarak cluster dengan centroid menggunakan Rapid Miner

### Evaluasi (Evaluation)

Berdasarkan hasil pengolahan data dengan metode *k-means clustering* menggunakan *Rapid Miner* dapat dilakukan pengelompokan data dari hasil cluster yang diperoleh yaitu sebagai berikut.

#### 1. Hasil Cluster 1

Hasil cluster 1 terdiri dari 10 mahasiswa, dapat dilihat pada tabel 6, bahwa karakteristik mahasiswa pada cluster 1 didominasi mahasiswa dari program studi Farmasi sebanyak 6 mahasiswa. Sedangkan berdasarkan wilayah kota asal, didominasi oleh mahasiswa dari wilayah kota asal DKI Jakarta dan Jawa Barat. Sehingga dapat disimpulkan bahwa rata-rata mahasiswa pada cluster 1 berasal dari wilayah kota asal DKI Jakarta dan Jawa Barat, peminatan program studi Farmasi, dengan rata-rata IPK 3,26.

Tabel 6. Hasil Analisis Cluster 1

Hasil Cluster 1	
Cluster 1 terdiri dari 10 mahasiswa yang berasal dari :	
Program Studi : a. Farmasi : 6 b. Desain Komunikasi Visual : 3 c. Teknik Industri : 1	Wilayah Kota Asal : a. DKI Jakarta : 4 b. Jawa Barat : 4 c. Jawa Tengah : 2
Dengan rata-rata nilai IPK 3.26	

#### 2. Hasil Cluster 2

Hasil cluster 2 terdiri dari 147 mahasiswa, dapat dilihat pada tabel 7. Karakteristik mahasiswa pada cluster 2 didominasi mahasiswa dari program studi Farmasi sebanyak 68 dan Teknik Kimia 26 mahasiswa. Sedangkan berdasarkan wilayah kota asal, didominasi oleh mahasiswa dari wilayah kota asal Jawa Barat dan DKI Jakarta. Sehingga dapat disimpulkan bahwa rata-rata mahasiswa pada cluster 2 berasal dari wilayah kota asal Jawa Barat dan DKI Jakarta, peminatan program studi Farmasi dan Teknik Kimia, dengan rata-rata IPK 3,27.

Tabel 7. Hasil Analisis Cluster 2

Hasil Cluster 2	
Cluster 2 terdiri dari 147 mahasiswa yang berasal dari :	
Program Studi : a. Farmasi : 68 b. Teknik Kimia : 26 c. Teknik Mesin : 20 d. Teknik Sipil : 14 e. Teknik Informatika : 12 f. Teknik Industri : 3 g. Sistem Informasi : 3 h. Desain Komunikasi Visual : 1	Wilayah Kota Asal : a. Jawa Barat : 66 b. DKI Jakarta : 52 c. Jawa Tengah : 25 d. Sumatera : 3 e. Jawa Timur : 1
Dengan rata-rata nilai IPK 3.27	

#### 3. Hasil Cluster 3

Hasil cluster 3 terdiri dari 175 mahasiswa, dapat dilihat pada tabel 8. Karakteristik mahasiswa pada cluster 3 didominasi mahasiswa dari program studi Farmasi sebanyak 158 mahasiswa. Sedangkan berdasarkan wilayah kota asal, didominasi oleh mahasiswa dari wilayah kota asal Jawa Barat dan DKI Jakarta. Sehingga dapat disimpulkan bahwa rata-rata mahasiswa pada cluster 3 berasal dari wilayah kota asal Jawa Barat dan DKI Jakarta, peminatan program studi Farmasi, dengan rata-rata IPK 3,26.

Tabel 8. Hasil Analisis Cluster 3

Hasil Cluster 3	
Cluster 3 terdiri dari 175 mahasiswa yang berasal dari :	
Program Studi : a. Farmasi : 158  b. Sistem Informasi : 9  c. Teknik Industri : 8	Wilayah Kota Asal : a. Jawa Barat : 89 b. DKI Jakarta : 76 c. Jawa Tengah : 4 d. Jawa Timur : 3 e. Sumatera : 3
Dengan rata-rata nilai IPK 3.26	

### Penyebaran (Deployment)

Dari data hasil clustering yang telah dilakukan di atas, maka dapat diterapkan pada beberapa strategi promosi yang dapat dilakukan oleh pihak marketing ISTA agar promosi yang dilakukan lebih efektif dan efisien, yaitu:

- Promosi dengan mengirim tim marketing yang sesuai dengan program studi yang paling banyak diminati.



- b. Promosi pada persebaran wilayah kota asal berdasarkan tingkat akademik calon mahasiswa.

Selanjutnya dilakukan analisis strategi promosi pada tabel 9, berdasarkan *promotion mix* pada masing-masing *cluster* yang terbentuk. Diharapkan dengan strategi yang diberikan pada masing-masing *cluster* dapat sesuai dengan tujuan bisnis.

Tabel 9. Strategi Promosi berdasarkan *Promotion Mix*

No	Strategi Promosi	Cluster 1	Cluster 2	Cluster 3
1	Periklanan	√	√	√
2	Penjualan Personal	√	√	√
3	Promosi Penjualan	√	√	√
4	Hubungan Masyarakat	√		
5	Pemasaran Langsung	√		

## 5. KESIMPULAN

Dari hasil penelitian yang telah dilakukan dapat disimpulkan sebagai berikut :

- Setelah dilakukan pengelompokan data mahasiswa melalui persebaran wilayah kota asal, peminatan program studi dan potensi akademik menggunakan *K-Means clustering* dan diolah dengan *Rapid Miner*, terbentuk 3 (tiga) cluster yaitu :
  - Cluster 1 dengan jumlah 10 mahasiswa berasal dari wilayah kota asal DKI Jakarta dan Jawa Barat dengan peminatan program studi Farmasi dan rata-rata IPK 3,26.
  - Cluster 2 dengan jumlah 147 mahasiswa berasal dari wilayah kota asal Jawa Barat dan DKI Jakarta dengan peminatan program studi Farmasi dan Teknik Kimia dan rata-rata IPK 3,27.
  - Cluster 3 dengan jumlah 175 mahasiswa berasal dari wilayah kota asal Jawa Barat dan DKI Jakarta, peminatan program studi Farmasi, dengan rata-rata IPK 3,26.
- Strategi promosi bagi calon mahasiswa baru yang tepat sasaran untuk setiap wilayah berdasarkan *cluster* yang terbentuk adalah dengan mengirim tim admisi marketing ISTA yang sesuai dengan program studi yang paling banyak diminati dan melakukan promosi berdasarkan potensi akademik mahasiswa

dengan melakukan penyesuaian menggunakan strategi *promotion mix*.

## DAFTAR PUSTAKA

- Arifin.Muhammad. (2014). "*Bussiness Intelligence Untuk Prediksi Customer Churn Telekomunikasi*", Prosiding SNATIF, Vol.1, Agusta, Y., *K-means - Penerapan, Permasalahan dan Metode Terkait*. Jurnal Sistem dan Informatika Vol. 3: 47-60.
- Bryan L. Garton, James E. Dyer, James E. Dyer. (2000). "*The Use Of Learning Styles And Admission Criteria In Predicting Academic Performance And Retention Of College Freshmen*", *Journal of Agricultural Education*, Vol. 41 Issue 2.
- Han. Jiawei, Micheline Kamber, Jian Pei.(2012). *Data Mining Concept and Techniques*, San Fransisco: Morgan Kauffman.USA,
- Institut Sains dan Teknologi Al-Kamal. (2014).Buku Panduan Akademik dan Kemahasiswaan. Jakarta.
- Irwan Budiman, Toni Prahasto, and Yuli Christyono. (2012). *Data Clustering Menggunakan Metodologi Crisp-Dm Untuk Pengenalan Pola Proporsi Pelaksanaan Tridharma*, SNATI, Yogyakarta.
- Kusrini,&Luthfi, E.T. (2009). *Algoritma Data Mining*. Yogyakarta: Andi Publishing.
- Larose, Daniel T. (2005). *Discovering Knowledge in Data*. New Jersey: John Willey & Sons, Inc.
- Larose, Daniel T.,(2006). *Data Mining Methods and Models*. Hoboken New Jersey: Jhon Wiley & Sons, Inc,
- Miranda. Eka. (2008). *Pengembangan Bussiness Intelligence Bagi Perkembangan Bisnis Perusahaan*, Comm IT, Vol.2, No. 2.
- Mustafid. (2013). *Sistem Bussiness Intelligence Untuk Mendukung Perguruan Tinggi Yang Kompetitif*, Jurnal Sistem Informasi Bisnis, Vol.1, No.1.
- Philip dan Gary Armstrong Kotler. (2001). *Prinsip-Prinsip Pemasaran*, Alih Bahasa Imam Nurmawan, Ed. Jakarta: Salemba Empat.
- Ong. Johan Oscar. (2013). *Implementasi Algoritma K-Means Clustering Untuk Menentukan Stategi Marketing President University*, Jurnal Ilmiah Teknik Industri, Vol. 12, No. 1.
- Rajagopal. Sankar. (2011). *Customer Data Clustering Using Data Mining Technique*, *International Journal of Database Management Systems*, Vol.3, No.4.
- Ranjan.Jayanthi.(2009).*BusinessIntelligence: Concepts, Components, Techniques And Benefits*, *Journal of Theoretical and Applied Information Technology*.

- Suprihatin, (2011). Klastering K-Means Untuk Penentuan Nilai Ujian, *JUSI*, Vol. 1, No. 1.
- Santosa, B. (2007). *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisni* Yogyakarta: Graha Ilmu.
- Safuan, Wahono, R.S., Catur S. (2015). Penanganan Fitur Kontinyu dengan *Feature Discretization* Berbasis *Expectation Maximization Clustering* untuk Klasifikasi *Spam Email* Menggunakan Algoritma ID3, *Journal of Intelligent Systems*, Vol.1, No.2.
- Turban, E. dkk. (2005). *Decicion Support Systems and Intelligent Systems.*: Andi Offset.
- Wahono, R.S., Herman, N.S., Ahmad, S. (2014). *A Comparison Framework of Classification Models for Software Defect Prediction*, *Advanced Science Letters*, 20(10-12).