

KAJIAN DATA MINING PROFIL SISWA BARU DALAM PENENTUAN STRATEGI PROMOSI DENGAN METODE TWO STEP CLUSTERING

Iis Ismawati

Prodi Teknik Informatika

Sekolah Tinggi Teknologi Bandung

Jl. Soekarno Hatta No.378 Bandung

Cuiez86@gmail.com

Abstrak

Data yang disimpan dalam *database* pada umumnya belum dimanfaatkan secara maksimal. Sekolah setiap tahun melakukan promosi dengan biaya yang besar untuk kegiatan penerimaan siswa baru, tapi belum diketahui tingkat efektifitas dari kegiatan promosi yang dilakukan. *Database* penerimaan siswa baru dapat dimanfaatkan untuk mengetahui profil siswa baru dengan menerapkan metode *two step clustering* diakrenakan memiliki tipe data campuran yaitu tipe data numerik dan kategorikal. Klaster yang terbentuk kemudian dianalisis untuk dapat mengidentifikasi strategi promosi yang tepat. Data yang akan diteliti adalah data siswa baru tahun 2015, 2016 dan 2017 yang berjumlah 1399, dengan metodologi *Cross-Industry Standart Process for Data Mining (CRISP-DM)* yang terdiri dari enam tahapan yaitu *Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation* dan *Deployment*.

Melalui tahapan *data preparation* variabel yang digunakan berjumlah delapan variabel yaitu jenis kelamin, agama, kerja ayah, kerja ibu, penghasilan orang tua, kode jurusan, jarak ke sekolah dan jarak SMP. Pada penelitian ini data dikelompokkan menjadi dua sampai empat klaster dengan tiga skema variabel, sehingga total didapatkan sembilan skema, yaitu skema 1.2 sampai skema 1.4 dengan 8 variabel, 2.2 sampai 2.4 dengan 6 variabel, dan 3.2 sampai 3.4 dengan 5 variabel. Pengujian dengan jumlah klaster yang berbeda dan variabel yang berbeda ditujukan untuk mendapatkan klaster yang menggambarkan profil siswa baru dengan tepat untuk strategi promosi yang akan direkomendasikan kepada pihak sekolah.

Hasil pengujian menunjukkan bahwa siswa baru dengan klaster terbesar didominasi oleh laki-laki, jurusan TKJ, beragama islam dengan jarak rumah rata-rata 6.25-6.29 KM, Jarak SMP antara 12.61-12.69 KM, pekerjaan ayah wiraswasta, ibu tidak bekerja dengan penghasilan orang tua 2.000.000-4.999.999. Sedangkan klaster terkecil didominasi oleh perempuan pada jurusan Akuntansi, laki-laki pada jurusan RPL dan TKR, jarak rumah antara 5.11-736 KM, jarak SMP antara 8.16-20.12 KM, pekerjaan ayah wiraswasta, ibu tidak bekerja dengan penghasilan orang tua 2.000.000-4.999.999 untuk jurusan akuntansi, 500.000-999.999 untuk jurusan TKR, dan 1.000.000-1.999.999 untuk jurusan RPL. Strategi yang dilakukan berdasarkan analisis profil siswa baru yang didapatkan melalui metode two step clustering dan pemetaan menggunakan matriks BCG berdasarkan sembilan skema yaitu dengan membuat acara tahunan turnamen futsal, pelatihan perakitan laptop, entrepreneur day, perlombaan peragaan busana, pelatihan beauty class, dan pelatihan pemrograman dasar dimana pesertanya merupakan siswa SMP, dan pemberian potongan biaya pendidikan bagi siswa yang bertempat tinggal dalam satu kelurahan yang sama dengan sekolah.

Kata kunci :

Data Mining, Two Step Clustering, Strategi Promosi, Siswa Baru, Matriks BCG

Abstract

The stored Data in Data base is generally not fully utilized. The school conducts promotions every year for new student admission while there is no being known effectively level of carried out promotion. Its database can be used for finding out

Iis Ismawati

Jurnal Ilmiah Teknologi Informasi Terapan

Volume 5, No 3, 15 Agustus 2019

the new student profiles by using two step clustering method deemed to have mixed data types namely numeric and categorical data types. The clusters that are formed are then analyzed to be able to identify the right promotional strategies. The researched data are some new student data in 2015, 2016 and 2017 students are totally 1399 people, with the Processing Standard for Cross Industry-Data Mining (CRISP-DM) methodology which consists of six stages, which are Business Understanding, Understanding Data, Preparation Data, Modeling, Evaluation and Deployment.

Through the preparation data variable stage used into eight variables namely gender, religion, father's job, mother's job, parents income, department code, distance to school and junior high school distance. The mixed data is running with two step clustering methods. In this Research, the data is grouped into two or four clusters with three variable schemes so a total of nine schemes that the naming schemes begins in 1.2 until 1.4 for 8 variable, 2.2 till 2.4 for 6 variable, then 3.2 till 3.4 for 5 variable. The calibration for the amount of different clusters and variables is objected to obtain for the ideal cluster to depict the proper profile through the promoting strategy which is recommended to the school.

The result shows that new students with the largest are dominated by male in majoring TKJ, moslems; the range distance 6.25 – 6.29 kilometers, Junior School distance at around 12.61- 12.69 kilometers, and entrepreunering father position with mother of housewife position among salary at Rp. 2.000.000- Rp. 4.999.999,-. Whereas the smallest cluster is dominated by female in majoring Accountancy, male in majoring RPL and TKR, the range distance at 5.11-736 kilometers, school range distance at 8.16-20.12 kilometers, enterpreunering father position with housewife of mother position among salary at Rp. 2.000.000-4.999.999 in majoring Accountancy, Rp. 1.000.000-1.999.999 in Majoring RPL, and Rp. 500.000 - 999.999 for TKR Major. The strategy, that can be treated after looking at the output of the new student's profiles, obtained through the two-step clustering and mapping methods using a BCG matrix based on nine schemes, is planned annual program as futsal tournament, laptop assembly training, entrepreneur day, fashion show competition, beauty class training, and basic programming on training activities, which some participants are from junior high school students,

giving tuition fees for students residing in the same kelurahan as the school.

Keywords :

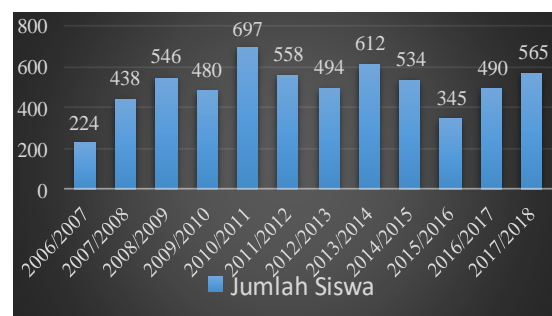
Data Mining, Two Step Clustering, Promotion Strategy, New Students, BCG Matrix.

I. PENDAHULUAN

Data yang disimpan pada umumnya belum dimanfaatkan secara maksimal. Data tersebut dapat diolah menggunakan sebuah teknologi, salah satunya dengan teknik data mining. Data mining sendiri adalah suatu istilah untuk menemukan pengetahuan yang tersembunyi di dalam database.

Metode yang dapat digunakan untuk menganalisa profil dari database yaitu dengan metode clustering. Permasalahan utama dalam menerapkan analisis clustering adalah jika tipe data dasar clustering bersifat campuran yaitu kategorik dan numerik. Salah satu metode clustering yang didesain untuk menangani objek yang mempunyai tipe data kategorik dan numerik adalah dengan menggunakan metode two step clustering.

SMK Medikacom adalah salah satu sekolah swasta di kota Bandung. Setiap tahun SMK Medikacom melakukan penerimaan siswa baru yang membutuhkan biaya yang besar tetapi belum diketahui tingkat efektifitas dari kegiatan promosi yang dilakukan. Berikut adalah data jumlah pendaftar pada SMK Medikacom dari tahun pelajaran 2006/2007 sampai 2017/2018 yang ditampilkan pada gambar 1 :



Gambar 1 Grafik Jumlah Pendaftar
(Sumber : Dokumen Profile SMK Medikacom Bandung)

Gambar 1 menunjukkan bahwa jumlah pendaftar setiap tahunnya yang mengalami perubahan, terkadang mengalami kenaikan atau bahkan

penurunan jumlah pendaftar tanpa diketahui penyebabnya. Berdasarkan latar belakang tersebut, maka dibutuhkan sebuah “**Kajian Data Mining Profil Siswa Baru Dalam Penentuan Strategi Promosi Dengan Metode Two Step Clustering**” pada SMK Medikacom Bandung.

II. KAJIAN LITERATUR

II.1 Data Mining

Data mining merupakan bidang dari beberapa bidang keilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, *database*, dan visualisasi untuk pengenalan permasalahan pengambilan informasi dari *Data mining* merupakan bidang dari beberapa bidang keilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, *database*, dan visualisasi untuk pengenalan permasalahan pengambilan informasi dari *database* yang besar. Seperti yang diutarakan Larose dalam buku yang berjudul *Discovering Knowledge In Data* yang isinya :

“*Data mining is an interdisciplinary field bringing together techniques from machine learning, pattern recognition, statistics, databases, and visualization to address the issue of information extraction from large data bases*”. (Larose, 2005 : 2)

Menurut Gartner Group *data mining* adalah suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam media penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika (Larose, 2005: 2).

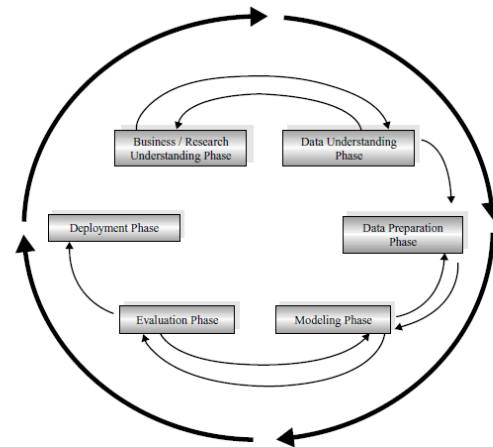
Menurut (Fajar, 2013 :1). “*Data mining berisi pencarian trend atau pola yang diinginkan dalam database yang besar untuk membantu pengambilan keputusan diwaktu yang akan datang*”.

Data mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual. Perlu diingat bahwa *data mining* sendiri berarti usaha untuk mendapatkan sedikit data berharga dari sejumlah besar data dasar. Karena itu *data mining* sebenarnya memiliki akar yang panjang dari bidang ilmu seperti kecerdasan buatan (*artificial intelligent*), *machine learning*, statistik dan *database*.

A. Cross-Industry Standard Process for Data Mining (CRISP-DM)

CRISP-DM yang di kembangkan tahun 1996 oleh analisis dari beberapa industri seperti *Daimler Chrysler*, *SPSS* dan *NCR*. CRISP-DM menyediakan standar proses *data mining* sebagai strategi pemecahan masalah secara umum dari bisnis atau unit penelitian.

Dalam CRISP-DM sebuah proyek data mining memiliki siklus hidup yang terbagi dalam enam fase yang ditunjukkan pada gambar 2.



Gambar 2
CRISP-DM (Sumber : Larose, 2005 : 5)

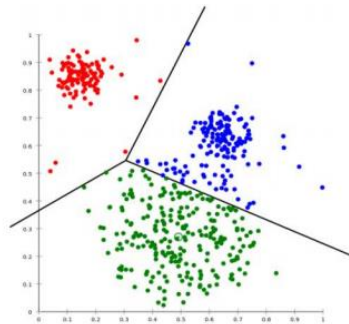
Enam fase CRISP-DM (*Cross Industry Standard Process for Data Mining*): (Larose, 2005 : 5)

1. Fase Pemahaman Bisnis (*Business Understanding Phase*)
2. Fase Pemahaman Data (*Data Understanding Phase*)
3. Fase Pengolahan Data (*Data Preparation Phase*)
4. Fase Pemodelan (*Modeling Phase*)
5. Fase Evaluasi (*Evaluation Phase*)
6. Fase Penyebaran(*Deployment Phase*)

II.2 Clustering

Clustering adalah suatu metode yang digunakan untuk mengelompokkan dokumen-dokumen ke dalam kelompok-kelompok atau *cluster* berdasarkan kemiripan dokumen, sehingga dokumen yang saling berhubungan ditempatkan pada *cluster* yang sama,

ilustrasi *clustering* pada *data mining* ditunjukkan pada gambar 3.

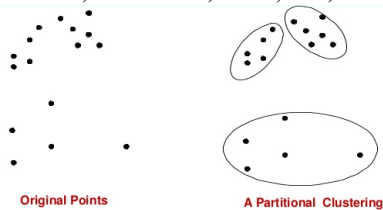


Gambar 3
Ilustrasi clustering
(Sumber : Ejaz, 2017 : 32)

Berikut penjelasan beberapa pendekatan dan algoritma yang digunakan dalam *clustering*:

1. *Partitioning-based*

Clustering dengan partisi akan membagi objek ke dalam subset (*cluster*) sehingga tiap satu data objek akan dikelompokkan tepat dalam satu subset saja, seperti yang ditunjukkan oleh gambar 4. Contoh: *kmeans*, *Fuzzy c-means (FCM)*, *k-medoids*, *CLARA*, *CLARANS*, *PAM*, *EM*, *Two Step*

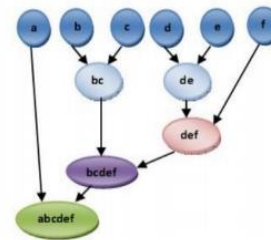


Gambar 4
Partitional Clustering
(Sumber : Ejaz, 2017 : 35)

2. *Hierarchical-based*

Pendekatan hirarki akan mengelompokkan objek yang mirip dengan meletakkannya pada hirarki yang berdekatan dan sebaliknya objek yang tidak mirip akan berada pada

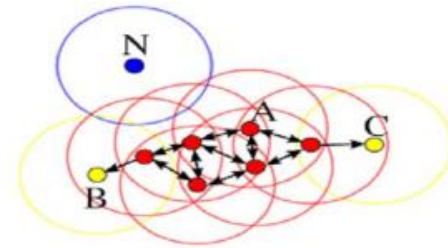
hirarki yang berjauhan, seperti yang ditunjukkan gambar 5. Contoh: *BIRCH*, *AGNES*, *DIANA*



Gambar 5 Hierarchical Clustering
(Sumber : Ejaz, 2017 : 36)

3. *Density-based*

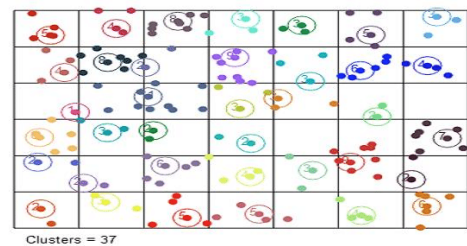
Pendekatan ini cocok digunakan untuk klustering objek yang memiliki *noise* dan *outlier*, namun sulit menemukan kluster yang memiliki bentuk yang berubah-ubah. *Density-based* ditunjukkan pada gambar 6. Contoh: *DBSCAN*, *OPTICS*, *DENCLUE*, *CLIQUE*



Gambar 6
Density-based Clustering
(Sumber : Ejaz, 2017 : 36)

4. *Grid-based*

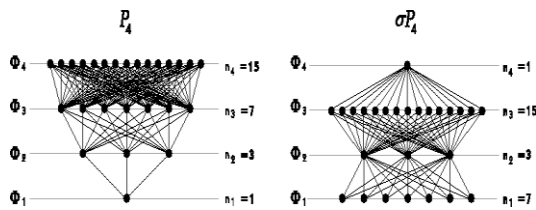
Pendekatan yang mebagi objek ke dalam sejumlah sel yang membentuk struktur grid. Seperti yang ditampilkan pada gambar 7. Contoh: *STING*, *WaveKluster*



Gambar 7 Grid Based Clustering
(Sumber : Ejaz, 2017 : 36)

5. *Model-based clustering*

Pendekatan didasarkan pada dugaan-dugaan mengenai sebuah model untuk tiap kluster, kemudian mencari data objek yang sesuai untuk model tersebut. Contoh: *SOM*, *COBWEB* yang ditunjukkan pada gambar 8.



Gambar 8 Cobweb Clustering
 (Sumber : Andrzej, 2009 : 6)

II.3 Two Step Clustering

“Metode two step cluster adalah metode yang didesain untuk menangani jumlah objek yang besar, terutama pada masalah objek yang mempunyai peubah campuran, kontinu dan kategorik” (Latifurrahman, 2014 : 44).

Metode penggerombolan lain yang dapat digunakan untuk data berskala campuran yaitu metode *Laten Class Model* pada software *Laten GOLD* dan *ALMO*. Dari ketiga metode tersebut, metode *two step cluster* merupakan metode yang paling efisien dari segi waktu, hal ini karena metode *two step cluster* didesain untuk data yang berukuran besar.

Jarak antara dua gerombolan didefinisikan sebagai jarak antar pusat dari masing-masing peubahnya. Jarak yang digunakan pada metode *two step cluster* adalah jarak *Log-Likelihood* dan jarak *Euclidean*.

“Prosedur penggerombolan dengan metode *two step cluster* mempunyai dua tahapan yaitu tahap *preclustering* (penggerombolan awal) objek kedalam *subcluster-subcluster* kecil dan tahap *penggerombolan akhir*” (Bacher : 2004 : 103).

Jarak yang digunakan dalam metode *two step cluster* adalah jarak *Log-Likelihood* dan jarak *Euclidean*. Prosedur penggerombolan dengan metode *two step cluster* mempunyai dua tahapan yaitu tahap *preclustering* (penggerombolan awal) objek ke dalam *subcluster-subcluster* kecil dan tahap penggerombolan akhir.

1. Penggerombolan Awal (*Preclustering*)

Tahap ini dilakukan dengan pendekatan sekuensial, yaitu objek diamati satu persatu berdasarkan ukuran jarak yang kemudian ditentukan apakah objek tersebut masuk dalam kelompok yang telah terbentuk atau harus membentuk kelompok baru. Algoritma melakukan

prosedur yang sangat mirip dengan algoritma *k-means*. Pada langkah ini, diimplementasikan dengan pembentukan *Cluster Feature (CF) Tree*. *CF-tree* terdiri dari beberapa tingkat cabang/simpul dan masing-masing cabang berisikan individu objek (entries) dari kelompok awal. Tingkatan daun atau daun entri yang terdapat pada cabang merepresentasikan anak kelompok (*subcluster*). Prosedur *CF-tree* dilakukan dengan memilih satu amatan awal secara acak yang akan diukur jaraknya satu persatu dengan amatan lainnya menggunakan ukuran jarak yang telah ditentukan.

2. Tahap Kedua/Pengelompokkan Akhir

Hasil dari *CF-tree* dikelompokkan dengan analisis kelompok berhierarki dengan metode *agglomerative*. Untuk menghitung banyaknya kelompok dapat dilakukan dengan dua tahapan, yang pertama adalah menghitung *Schwarz's Bayesian Criterion (BIC)* untuk tiap kelompok. Rumus BIC untuk kelompok adalah sebagai berikut:

$$BIC(j) = -2 \sum_{j=1}^j \xi_j + m_j \log(N) \quad \dots [1]$$

Dengan :

$$m_j = j \left\{ 2K^A + \sum_{k=1}^{K^B} (L_k - 1) \right\}$$

$$\xi_j = N \left(\sum_{k=1}^{K^A} \frac{1}{2} \log(\hat{\sigma}_k^2 + \hat{\sigma}_{jk}^2) \sum_{k=1}^{K^B} \sum_{l=1}^{L_k} \frac{N_{jkl}}{N_j} \log \left(\frac{N_{jkl}}{N_j} \right) \right)$$

Solusi banyaknya kelompok yang optimal adalah yang memiliki nilai BIC terkecil, tetapi ada beberapa kasus dalam pengelompokkan dimana BIC akan terus meningkat nilainya bila jumlah kelompok semakin meningkat. Maka dalam situasi tersebut, *ratio BIC changes* (rasio perubahan BIC) dan *ratio of distance measure changes* (rasio perubahan jarak) digunakan untuk mengidentifikasi solusi banyaknya kelompok optimal. Solusi untuk banyaknya kelompok optimal akan memiliki *ratio BIC changes* dan *ratio distance measure* yang besar. Jumlah kelompok yang terbentuk dapat diketahui dengan menggunakan perbandingan antar jarak untuk gerombol, dengan rumus perbandingannya sebagai berikut:

$$R(k) = d_{k-1} / d_k$$

$$dk = l_{k-1} - l_k \quad \dots [2]$$

dengan:

$$l_v = (r_v \log n - BIC_v)/2, \text{ atau}$$

$$l_v = (2r_v - AIC_v)/2$$

$$v = k, k - 1$$

d_{k-1} = jarak jika kelompok k digabungkan dengan k-1 kelompok

R(k) = rasio perubahan jarak.

Menurut (Anderberg, 1973 : 9) : “Ukuran jarak dibutuhkan untuk setiap pasang objek yang akan dikelompokkan”. Beberapa metode pengukuran jarak antar dua objek, yaitu:

1. Jarak Euclidean

Jarak ini merupakan jarak yang umum digunakan dan dapat digunakan apabila semua peubahnya berskala kontinu. Jarak ini harus memenuhi asumsi bahwa peubah-peubah yang diamati tidak berkorelasi dan antar peubah memiliki satuan yang sama. Dalam metode ini, pengukuran jarak dilakukan dengan menghitung akar kuadrat dari penjumlahan kuadrat selisih dari nilai masing-masing peubah. Jarak *euclidean* dapat dirumuskan sebagai berikut:

$$d_{Euclidean} = (x, y) = \sqrt{\sum_i (x_i - y_i)^2} \dots [3]$$

2. Jarak Manhattan (City Block/Minkowski)

Jarak ini merupakan bentuk umum dari jarak *euclidean*. Jarak *manhattan* digunakan jika peubah yang diamati berkorelasi atau tidak saling bebas. Dalam metode ini, pengukuran jarak dilakukan dengan menghitung jumlah absolut perbedaan objek untuk masing-masing peubah. Jarak *manhattan* dapat dirumuskan sebagai berikut:

$$d_{Cityblock} = (x, y) = \sum_i |x_i - y_i| \dots [4]$$

3. Jarak Mahalanobis

Jarak ini sangat berguna dalam menghilangkan atau mengatasi perbedaan skala pada masing-masing peubah. Jarak *Mahalanobis* dapat dirumuskan sebagai berikut:

$$d_{mahalanobis} = \sqrt{(X_{ik} - X_{jk})' S^{-1} (X_{ik} - X_{jk})} \dots [5]$$

Dengan:

$d_{mahalanobis}$ = jarak antara objek dengan objek

X_{ik} = nilai objek i pada peubah ke-

X_{jk} = nilai objek pada peubah ke-

S = matriks kovarian

4. Jarak Log-Likelihood

Jarak ini digunakan untuk peubah berskala kontinu dan kategorik. Jarak antara gerombol dengan gerombol dapat dirumuskan sebagai berikut:

$$d(j, s) = \xi_j + \xi_s - \xi_{j-s} \dots [6]$$

Dengan :

$$\xi_j = -N \left(\sum_{k=1}^{K^A} \frac{1}{2} \text{Log}(\hat{\sigma}_k^2 + \sigma_{jk}^2) - \sum_{k=1}^{K^B} \sum_{l=1}^{L_k} \frac{N_{jkl}}{N_j} \text{Log}\left(\frac{N_{jkl}}{N_j}\right) \right)$$

$$\xi_s = -N \left(\sum_{k=1}^{K^A} \frac{1}{2} \text{Log}(\hat{\sigma}_k^2 + \sigma_{sk}^2) - \sum_{k=1}^{K^B} \sum_{l=1}^{L_k} \frac{N_{skl}}{N_s} \text{Log}\left(\frac{N_{skl}}{N_s}\right) \right)$$

$$\xi_{(js)} = -N \left(\sum_{k=1}^{K^A} \frac{1}{2} \text{Log}(\hat{\sigma}_k^2 + \sigma_{(js)k}^2) - \sum_{k=1}^{K^B} \sum_{l=1}^{L_k} \frac{N_{jlskl}}{N_{js}} \text{Log}\left(\frac{N_{jlskl}}{N_{js}}\right) \right)$$

N = banyaknya objek

N_j = jumlah objek di dalam gerombol

N_{jkl} = jumlah objek di gerombol j untuk peubah kategorik ke- k dengan kategori ke- l

$\hat{\sigma}_k^2$ = ragam dugaan untuk peubah kontinu ke- k untuk keseluruhan objek

$\hat{\sigma}_{jk}^2$ = ragam dugaan untuk peubah kontinu ke- k untuk keseluruhan objek dalam gerombol j

K^A = banyaknya peubah kontinu

K^B = banyaknya peubah kategorik

L_k = banyaknya kategori untuk peubah kategorik ke- k

II.4 Penerimaan Siswa Baru

Perencanaan terhadap siswa menyangkut perencanaan penerimaan siswa baru, kelulusan,

jumlah putus sekolah dan kepindahan. Khusus mengenai perencanaan siswa akan langsung berhubungan dengan kegiatan penerimaan dan proses pencatatan atau dokumentasi data pribadi siswa, yang kemudian tidak dapat dilepaskan kaitannya dengan pencatatan atau dokumentasi data hasil belajar dan aspek-aspek lain yang diperlukan dalam kegiatan kurikuler dan ko-kurikuler. (Dokumen ISO SMK MedikaCom Tentang PSB, 2014:50)

Langkah yang pertama yaitu perencanaan terhadap siswa, yang meliputi kegiatan;

- 1) Analisis kebutuhan siswa
- 2) Rekrutmen siswa
- 3) Seleksi siswa
- 4) Orientasi
- 5) Penempatan siswa
- 6) Pencatatan dan pelaporan

II.5 Matriks Boston Consulting Grup (BCG)

Boston Consulting Group (BCG) adalah perusahaan konsultan manajemen swasta yang berbasis di Boston. *Boston Consulting Grup* merupakan perusahaan yang berkecimpung dalam hal perkembangan pangsa pasar. Matriks BCG membantu perusahaan mengalokasikan sumber daya dan digunakan sebagai alat analitis dalam merek pemasaran, manajemen produk, manajemen strategis, dan analisis portofolio.

“ *Matrik BCG adalah matrik dan internal-external (IE) matrik yang dibentuk secara khusus dalam rangka meningkatkan usaha-usaha perusahaan yang memiliki multidivisi dengan merumuskan strategi yang paling cocok*”. (Husein Umar,1999:200).

Metode analisis BCG matrix membantu unit bisnis untuk mengetahui posisi dirinya di dalam empat kategori, yakni kategori Anjing (*Dog*), Tanda Tanya (*Question Mark*), Bintang (*Star*), dan Kas Sapi (*Cash Cow*). Matriks ini juga dapat digunakan untuk memetakan atribut produk strategis yang dapat menghasilkan keuntungan bagi perusahaan.

Matriks BCG terdapat 4 kuadran yang menggambarkan posisi suatu unit bisnis dipandang dari segi pertumbuhan pasar serta pangsa pasarnya, keempat kuadran atau kategori tersebut adalah :

1. *Dog* (Anjing)

Kategori Anjing memiliki pangsa pasar yang rendah dan tingkat pertumbuhan yang rendah dan

dengan demikian tidak menghasilkan atau mengkonsumsi uang dalam jumlah besar.

2. *Question Mark* (Tanda Tanya)

Kategori ini memiliki posisi pangsa pasar relatif yang rendah, tetapi mereka bersaing dalam industri yang bertumbuh pesat.

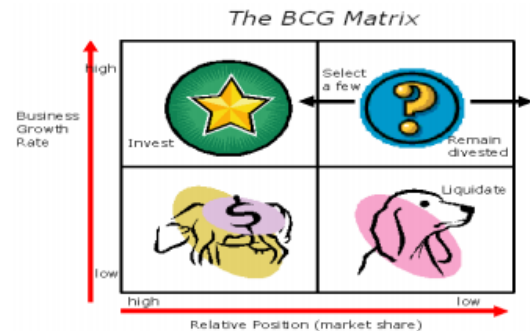
3. *Star* (Bintang)

Kategori ini memiliki peluang jangka panjang terbaik untuk pertumbuhan dan profitabilitas bagi organisasi.

4. *Cash Cow* (Kas sapi)/Sapi Perah

Kategori ini memiliki pangsa pasar relatif yang tinggi tetapi bersaing dalam industri yang pertumbuhannya lambat.

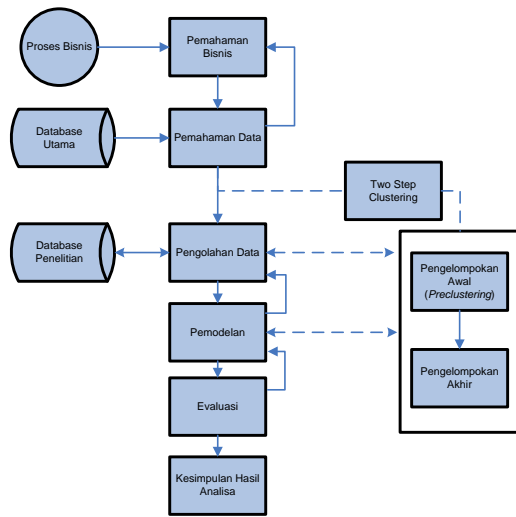
Gambar 9 menunjukkan matriks BCG yang terbagi menjadi empat kuadran.



Gambar 9
Matriks Boston Consulting Grup (BCG)
(Sumber : Yanuar, 2014 : 56)

III. ANALISIS DAN PERANCANGAN

Penelitian yang dilakukan menggunakan metodologi CRISP-DM (*Cross Industry Standard Proses For Data Mining*). Metodologi penelitian yang dilakukan digambarkan secara mendetail pada gambar 10, untuk memberikan gambaran informasi yang utuh mengenai tahap-tahap yang dilakukan dari awal penelitian sampai pada tahap kesimpulan dari hasil analisis.



Gambar 10

Metodologi dan tahapan penelitian kajian data mining menggunakan metode two step clustering

1. Fase Pemahaman Bisnis (Business Understanding Phase)

Business understanding merupakan cara untuk memahami tujuan proyek dan kebutuhan tujuan bisnis. Tahap ini mencakup :

- Penentuan tujuan proyek dan kebutuhan secara detail dalam lingkup bisnis atau unit penelitian secara keseluruhan. Penelitian ini bertujuan untuk menganalisa profil dari siswa baru yang telah mendaftar sehingga dapat mengidentifikasi strategi promosi yang tepat bagi SMK Medikacom Bandung. Kebutuhan untuk dapat melaksanakan penelitian ini adalah data siswa baru SMK Medikacom, data yang akan diteliti adalah data siswa baru pada tahun pelajaran 20015/2016, 2016/2017 dan 2017/2018.
- Menerjemahkan tujuan dan batasan menjadi formula dari permasalahan data mining. Penelitian dilakukan menggunakan metode two step clustering, dikarenakan dalam data memiliki dua tipe data yaitu nominal dan kategorikal seperti yang dapat dilihat pada gambar 11 yang menunjukkan untuk atribut jenis kelamin (JK), asal SMP dan jurusan bertipe kategorikal, sedangkan untuk atribut jarak rumah bertipe nominal.

No	Nama	JK	Alamat	Jarak semp	jenis_rumah	Kelurahan	Kecamatan	Penghasilan	jurusan
1	ABDUL AKEP ROZENDRO	L	PAKIGALAH	SHUPI 42	KARAYARAN	2,7 MARGALARI	Kec. Buahbatu	Rp. 1.400.000 - Rp. 1.999.999	TKJ
2	ABDUL LATIF	M	ANTAYANI	WISWANTARA	WISWANTARA	4,2 ANTAYANI LATA	Kec. Antayani	Rp. 200.000 - Rp. 999.999	ISI
3	ABDUL LATIF	L	IL. CIGANTRI SUKTI 03	WISWANTARA	WISWANTARA	7 CIPARIGALD	Kec. Bojongsari	Rp. 1.000.000 - Rp. 1.999.999	TKJ
4	ABDUL RAHMAN HALLILAH	L	IL. KARANGREJA	WISWANTARA	WISWANTARA	1 KARANGREJA	Kec. Bojongsari	Rp. 1.000.000 - Rp. 1.999.999	TKJ
5	ABDUL RAHMAN SIDDIQ	L	CILANTRE	LAHANTA	WISWANTARA	8,5 LENGKONGE	Kec. Bojongsari	Rp. 1.000.000 - Rp. 1.999.999	TKJ
6	ABDUL WAHEDI RAHMAY	L	MP. BARABANG KELLIPA	LAHANTA	KARAYARAN	10 NEKAWALERE	Kec. Tanggung Bidad	TKJ	
7	ABDULLA PUTRI	P	IL. TEKSTIL NO 27	SOMATTA 40	KARAYARAN	18 CICALADI	Kec. Cikarang Bidad	Rp. 2.000.000 - Rp. 4.999.999	TKJ

Gambar 11

Jenis tipe data pada data siswa baru

(Sumber : Database siswa baru SMK Medikacom Bandung)

- Menyiapkan strategi awal untuk mencapai tujuan. Strategi awal yang dilakukan adalah dengan memahami tujuan penelitian, menerapkan metodologi CRIPS-DM pada penelitian, dan menggunakan metode two step clustering sehingga data-data yang mirip atau memiliki karakteristik sama akan dikelompokkan dalam satu kelompok untuk kemudian dianalisis dan dievaluasi untuk mempermudah analisa profil siswa sehingga dapat dilakukan strategi yang tepat untuk menambah jumlah siswa baru yang mendaftar.
2. Fase Pemahaman Data (Data Understanding Phase)

Data understanding adalah fase awal pengumpulan data dan menemukan insight pertama. Tahap ini terdiri dari beberapa tahap diantaranya adalah mengumpulkan data, menggunakan analisis penyelidikan data untuk mengenali lebih lanjut data dan pencarian pengetahuan awal, mengevaluasi kualitas data, jika diinginkan, pilih sebagian kecil kelompok data yang mungkin mengandung pola dari permasalahan.

- Tahap pengumpulan data dilakukan ketika proses penerimaan siswa baru berlangsung, siswa mengisi formulir pendaftaran, kemudian operator pendaftaran melakukan entry data pada database. Sehingga dapat dikatakan bahwa objek yang diteliti merupakan data primer, yaitu data merupakan data yang diambil langsung dari pihak SMK Medikacom yang menjadi objek penelitian.
- Data yang digunakan dalam penelitian adalah data siswa baru pada tahun pelajaran 2015/2016, 2016/2017 dan 2017/2018. Para kurun waktu tiga tahun ini data yang ada dalam database adalah sebanyak 1399 item, dengan 68 atribut berupa data diri siswa, orang tua / wali dan data pendidikan sebelumnya.
- Mengidentifikasi kualitas data
Berdasarkan data yang tersimpan dalam database siswa baru terdapat 1399 item siswa baru. Dengan jumlah atribut sebanyak 68 atribut. Dari semua atribut yang melekat pada data siswa baru dipilih beberapa atribut yang memiliki kualitas baik dilihat dari kelengkapan data/ tidak ada data bernilai null, tidak ada redundansi data, tidak mengandung outlier, data

tidak usang, memiliki data yang konsisten dan yang memiliki hubungan dengan penelitian.

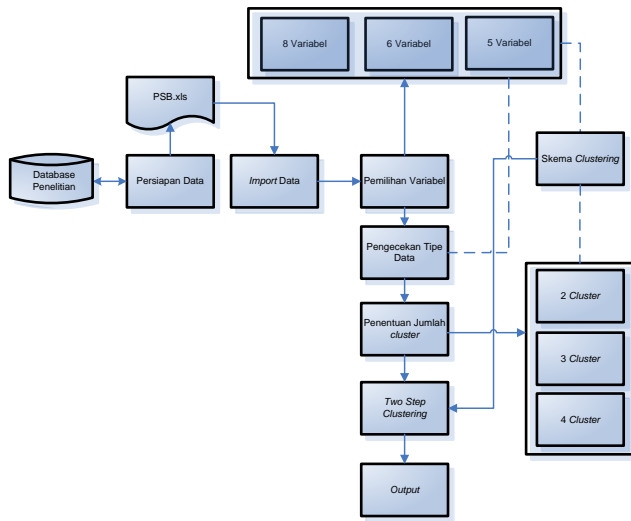
3. Fase Pengolahan Data (*Data Preparation Phase*)
Berdasarkan data yang sudah ada, kemudian akan diolah kedalam metode *two step clustering*. Namun sebelumnya data tersebut akan diproses untuk mendapatkan data yang sesuai dengan keinginan, sehingga dilakukan proses berikut :
 - a. Seleksi data
Memilih data yang akan digunakan dalam proses *data mining*. Data yang digunakan adalah data yang menjadi acuan pokok untuk melakukan analisis profil siswa baru yaitu jarak rumah ke sekolah, asal SMP/MTS, pekerjaan ayah, jenis kelamin, jurusan(kompetensi keahlian) yang dipilih, dan penghasilan orang tua. Data tersebut yang kemudian menjadi atribut yang digunakan pada proses *clustering* dengan metode *two step clustering*.
 - b. *Data Preprocessing*
Proses pembersihan data dilakukan dengan mencari dan memperbaiki kesalahan atau permasalahan dalam data seperti *missing values*, menghilangkan *outlier* atau *noisy* dan memperbaiki data dengan nilai yang tidak konsisten. Ada beberapa cara untuk mengatasi *missing values* diantaranya dengan mencari dari data sumber atau mengisi dengan modus atau rata-rata. Modus digunakan pada data yang berskala kontinu/numerik.
4. Fase Pemodelan (*Modeling Phase*)
Pada Fase pemodelan, metode yang akan digunakan adalah metode *two step clustering*. Metode ini dipilih karena metode yang didesain untuk menangani jumlah objek yang besar, terutama pada masalah objek yang mempunyai tipe data campuran, kontinu/numerik(data berupa hasil pengukuran) dan kategorik (dimana variable-variablenya dapat dikelompokkan menjadi beberapa kelompok) seperti jenis kelamin dan kompetensi keahlian. Data yang sudah dipilih kemudian diuji menggunakan metode *two step clustering* dengan menggunakan aplikasi SPSS *Modeler* 18.0. Pengujian dilakukan lebih dari dua kali pengujian dimulai dari pembentukan dua *cluster* sampai ditemukan jumlah *cluster* yang cocok dan menggambarkan profil yang sesungguhnya yang tidak mengalami perubahan-perubahan yang signifikan,
5. Fase Evaluasi (*Evaluation Phase*)

Model sepenuhnya dievaluasi dan diteliti untuk meyakinkan kalau model tersebut memenuhi tujuan bisnis untuk mendapatkan tujuan yang diharapkan dan menyiapkan strategi awal untuk mencapai tujuan. Langkah – langkah pada fase ini adalah :

- a. Mengevaluasi model yang digunakan, untuk mendapatkan kualitas dan efektivitas pada *cluster* yang terbentuk.
 - b. Menetapkan apakah memenuhi tujuan pada fase awal yaitu mengetahui profil siswa baru pada SMK Medikacom Bandung sebagai analisa dari strategi yang akan dilakukan untuk proses penerimaan baru. Jika tidak maka akan kembali kepada fase sebelumnya.
 - c. Menentukan apakah terdapat permasalahan penting dari bisnis atau penelitian yang tidak tertangani dengan baik.
 - d. Mengambil keputusan berkaitan dengan penggunaan hasil dari *data mining*.
6. Fase Penyebaran(*Deployment Phase*)
Pengetahuan yang didapat butuh untuk diorganisasikan dan dipresentasikan dengan cara dimana *customer* mudah memahaminya. Fase ini menghasilkan dari laporan hasil penelitian yang telah dilakukan dan kesimpulan dari penelitian.

IV. PEMODELAN DAN PEMBAHASAN

Pemodelan yang pada *database* siswa baru terdiri dari beberapa tahap dimulai dari tahap persiapan data, *import* data, pemilihan variabel, pengecekan tipe data, penentuan jumlah *cluster* dan pemodelan dengan metode *two step clustering*. Langkah-langkah pemodelan ini dapat dilihat pada gambar 12.



Gambar 12
Langkah-langkah pemodelan

1. Persiapan Data

Tahap pertama adalah persiapan data. Database yang sudah siap untuk dimodelkan disimpan dalam bentuk excel atau .xls untuk kemudian diproses untuk diimport kedalam aplikasi SPSS Modeler 18.0. Gambar 13 menunjukkan database penelitian yang sudah dalam bentuk excel.

No	Nama Lengkap	JK	Agama	Jarak_Ke_Sekolah	Jarak_SMP	Kerja_Ayah	Kerja_Ibu	Kode_Jurusan	Penghasilan
1	MARCEL SMP Bandung	Laki-laki	Protestan	2,7	1000000	Tidak bekerja	Tidak bekerja	TKR	Rp. 1.000.000 - Rp. 1.500.000
2	ALBERTUS LUTHER	Laki-laki	Katolik	4,2	1000000000	Agustiner	Agustiner	TKR	Rp. 1.000.000 - Rp. 1.500.000
3	ALBERTUS LUTHER	Laki-laki	Katolik	5,0	1000000000	Tidak bekerja	Tidak bekerja	TKR	Rp. 1.000.000 - Rp. 1.500.000
4	MARCEL SMP Bandung	Laki-laki	Protestan	2,7	1000000	Tidak bekerja	Tidak bekerja	TKR	Rp. 1.000.000 - Rp. 1.500.000
5	MARCEL SMP Bandung	Laki-laki	Protestan	2,7	1000000	Tidak bekerja	Tidak bekerja	TKR	Rp. 1.000.000 - Rp. 1.500.000

Gambar 13
Data dalam format excel

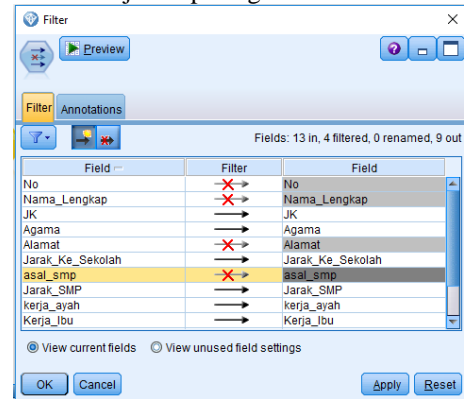
2. Import Data

Proses import data dari excel pada SPSS Modeler merupakan proses pengambilan data penelitian yaitu data PSB.xls pada SPSS Modeler. Tahapan ini dibutuhkan sebagai sumber data penelitian yang akan dimodelkan menggunakan SPSS Modeler. Setelah proses import data tahapan selanjutnya adalah pemilihan variabel yang akan dimodelkan.

3. Pemilihan variabel

Pada tahap pemilihan variabel terbagi menjadi tiga skema, skema kesatu menggunakan 8 buah variabel yaitu jenis_kel, agama, jarak_ke_sekolah, jarak_SMP, kerja_ayah, kerja_ibu, kode_jurusan dan penghasilan, skema kedua menggunakan 6 buah variabel yaitu dengan menghilangkan variabel agama dan kerja_ibu, dan pada skema ke 3 menggunakan 5 buah variabel dengan menghilangkan variabel radius_SMP. Pada

masing-masing skema akan dilakukan pengujian menggunakan pengclusteran dengan membagi menjadi 2 sampai 4 cluster. Tahap pemilihan variabel ini ditunjukkan pada gambar 14.



Gambar 14
Tahap pemilihan variabel yang digunakan dalam pemodelan

4. Pengecekan Tipe Data

Pengecekan tipe data diperlukan untuk menganalisa tipe data dari variabel-variabel yang akan dimodelkan dan sebagai pemeriksaan ketika data yang akan dimodelkan mengalami kesalahan pembacaan tipe data yang biasanya terjadi pada kesalahan penulisan variabel dengan tipe data numerik pecahan menggunakan koma atau titik. Karena kesalahan pembacaan tipe data akan menyebabkan kesalahan pada proses memodelkan data, terutama dikarenakan metode yang digunakan menggunakan metode two step clustering yang memisahkan data dengan variabel numerik dan kategorik. Gambar 4.8 menunjukkan jenis tipe data yang akan dimodelkan pada penelitian.

Field	Measurement	Values	Missing	Check	Role
JK	Flag	PL	None	In...	In...
Agama	Nominal	Budha, Islam, Katolik, Kristen	None	In...	In...
Jarak_Ke_S...	Continuous	[1,0, 135,0]	None	In...	In...
Jarak_SMP	Continuous	[1,0, 50,0]	None	In...	In...
Kerja_ayah	Nominal	BUMN/BUMD, BURUH, GUR...	None	In...	In...
Kerja_Ibu	Nominal	Buruh, "Karyawan Swasta", L...	None	In...	In...
Kode_Jurusan	Nominal	AKN, FAR, RPL, TKJ, TKR, TP...	None	In...	In...
Penghasilan	Nominal	"Kurang dari Rp. 500.000", "...	None	In...	In...

Gambar 15
Jenis tipe data dari variabel-variabel yang akan dilakukan pemodelan

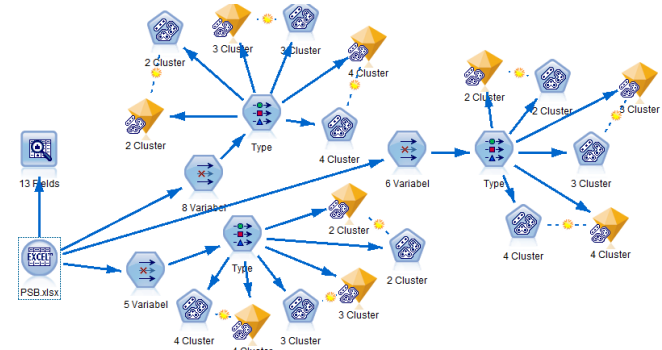
Gambar 15 menunjukkan data dengan tipe data kategorik yang memiliki dua jenis karakter akan dibaca sebagai data *nominal*, sedangkan variabel dengan tipe data kategorik yang memiliki dua jenis karakter dibaca sebagai *flag*. Untuk data *numerik* dibaca sebagai *continuous*.

5. Penentuan Jumlah *Cluster*

Pada tahapan ini ditentukan jumlah *cluster* yang akan diuji pada masing-masing skema. Pada skema kesatu dengan variabel berjumlah delapan buah dilakukan pengclusteran dengan menggunakan 2 sampai 4 *cluster*, untuk mempermudah pengidentifikasian maka penamaan skema dimulai dari skema 1.2 sampai skema 1.4. Pada skema kedua dengan variabel berjumlah enam dilakukan pengclusteran menjadi 2 sampai 4 *cluster* dengan nama 2.2 sampai 2.4. Begitu pula pada skema ketiga dengan variabel berjumlah 5 jumlah *cluster* sebanyak 2 sampai 4 *cluster* dengan nama 3.2 sampai 3.4. Pengujian dengan jumlah *cluster* yang berbeda dan variabel yang berbeda ditujukan untuk mendapatkan *cluster* yang ideal dan menggambarkan profil yang tepat untuk dilakukan analisis lanjutan dan analisis strategi promosi yang tepat untuk dijadikan usulan atau masukan kepada pihak sekolah dalam hal ini SMK Medikacom Bandung.

6. Pemodelan/Modeling

Pemodelan dengan menggunakan metode *two step clustering*, pada tahapan ini Data dengan tipe data kategorik akan dikelompokkan pada tahap penggerombolan awal / *preclustering* dengan melakukan perhitungan jarak menggunakan jarak *log-likelihood* sedangkan data dengan tipe data numerik akan dikelompokkan pada tahap kedua/ pengelompokan akhir dengan perhitungan jarak menggunakan jarak *euclidean*. Sedangkan untuk penentuan pusat *cluster* ditentukan secara acak. Proses akan terus berulang sampai pusat *cluster* tidak mengalami perubahan. Gambar 16 menunjukkan proses pemodelan dengan metode *two step clustering* dari tahap awal hingga tahap akhir menggunakan tiga skema pengclusteran.



Gambar 16
Pemodelan dengan metode *two step clustering*

Profil siswa baru berdasarkan *cluster* terbesar. Pada skema dengan dua *cluster* mendapatkan hasil yang memiliki kemiripan satu sama lain. Siswa baru didominasi oleh laki-laki jurusan TKJ, beragama islam dengan jarak rumah rata-rata 6.25-6.29 KM, Jarak SMP ke SMK antar 12.61-12.69 KM, pekerjaan ayah wiraswasta, ibu tidak bekerja/ ibu rumah tangga dengan penghasilan 2.000.000-4.999.999.

Pada Skema dengan 3 buah *cluster* menunjukkan hasil yang cukup berbeda ketika menggunakan variabel agama dan pekerjaan ibu dibandingkan ketika tidak menggunakan variabel tersebut. Dengan menghilangkan dua buah variabel maka hasil yang didapatkan yaitu siswa didominasi oleh laki-laki jurusan TKJ, dengan jarak rumah rata-rata 6.42-6.92 KM, pekerjaan ayah sebagai pegawai swasta, dengan penghasilan 2.000.000-4.999.999.

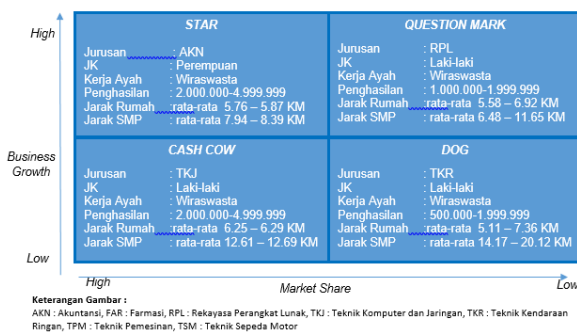
Pada Skema dengan 4 buah *cluster* menunjukkan hasil yang sangat berbeda dengan skema sebelumnya, tetapi masing-masing *cluster* yang terbentuk memiliki kemiripan yaitu sama-sama didominasi oleh perempuan dengan jurusan akuntansi, jarak rumah ke sekolah 5.76-5.87 KM, pekerjaan wiraswasta dengan penghasilan 2.000.000-4.999.999.

Analisis dilanjutkan berdasarkan *cluster* terkecil pada masing-masing skema yang telah dilakukan.

profil dari siswa baru dengan jumlah paling sedikit, pengujian pada masing-masing skema menunjukkan beberapa kemiripan terutama ketika mengelompokkan menjadi dua *cluster*. Sedangkan secara keseluruhan siswa baru pada *cluster* terkecil didominasi oleh perempuan dengan jurusan Akuntansi, laki-laki jurusan RPL dan TKR, jarak rumah antara 5.11-736 KM, jarak kesekolah antara 8.16-20.12 KM, pekerjaan ayah wiraswasta, ibu tidak bekerja dengan penghasilan orang tua 2.000.000-

4.999.999 untuk jurusan akuntansi, 500.000-999.999 untuk jurusan TKR, dan 1.000.000-1.999.999 untuk jurusan RPL.

Untuk mendapatkan strategi yang tepat selain strategi yang dilakukan berdasarkan hasil analisa pada masing-masing variabel dengan metode *two step clustering* dilakukan juga analisis dari seluruh skema pemodelan yang menunjukkan minat dan segmentasi pasar berdasarkan *cluster* yang terbentuk yang dipetakan kedalam matriks *Boston Consulting Group (BCG)*. Matriks BCG dibentuk secara khusus dalam rangka meningkatkan usaha-usaha SMK Medikacom yang memiliki banyak jurusan dengan merumuskan strategi yang paling cocok berdasarkan empat kuadran, yaitu kuadran *Dog/Anjing*, *Question Mark/Tanda Tanya*, *Star/ Bintang* dan *Cash Cow/Sapi Perah*. Pemetaan profil siswa baru pada masing-masing kuadran yang didapatkan dari hasil analisis pada masing-masing skema pada *cluster* terbesar dan *cluster* terkecil dapat dilihat pada gambar 17.



Gambar 17
Matriks BCG profil siswa baru berdasarkan sembilan skema pemodelan

V. KESIMPULAN DAN SARAN

Metode *two step clustering* diterapkan untuk variabel dengan tipe data campuran yaitu kategorikal dan numerik. Variabel dengan tipe data kategorikal akan dikelompokkan pada tahap pengelompokan awal dengan menggunakan perhitungan jarak log-likelihood sedangkan data dengan tipe data numerik akan dikelompokkan pada tahap kedua/ pengelompokan akhir dengan perhitungan jarak euclidean.

Pengujian pada data siswa baru menggunakan 9 skema pemodelan, berdasarkan analisa yang dilakukan dari masing-masing skema pemodelan didapatkan profil siswa baru yang didominasi oleh

laki-laki, beragama islam, jurusan TKJ, pekerjaan ayah wiraswasta, ibu tidak bekerja, penghasilan orang tua Rp. 2.000.000,- sampai Rp.4.000.000,- dengan jarak rumah kesekolah rata-rata 5.58 sampai 6.92 KM sedangkan untuk jarak SMP rata-rata 6.48 sampai 12.69 KM.

Strategi yang dilakukan berdasarkan analisis profil siswa baru yang didapatkan melalui metode *two step clustering* dan pemetaan menggunakan matriks BCG berdasarkan sembilan skema pemodelan adalah dengan membuat acara tahunan dimana pesertanya merupakan siswa SMP seperti kegiatan turnamen futsal, pelatihan perakitan laptop, entrepreneur day, perlombaan peragaan busana, pelatihan beauty class, pelatihan pemrograman dasar, pemberian potongan biaya pendidikan bagi siswa yang bertempat tinggal dalam satu kelurahan yang sama dengan sekolah.

REFERENSI

- Anderberg, M.R., *Cluster Analysis for Applications*. New York : Academic Press. 1973
- E. Mongi, Charles. Penggunaan Analisis *Two Step Clustering* untuk Data Campuran . JdC, Vol. 4, No. 1. 2015
- Han, J., Kamber, M.,Pei, J. *Data Mining Concept and Techniques*, 3rd ed. Amsterdam : Morgan Kaufmann-Elsevier. 2012
- Haq, Ejaz Ul. and Huarong, Xu. Khattak, dan M. Irfan. 2017. *A Review of Various Clustering Techniques*, *International Journal of Engineering and Applied Computer Science (IJEACS)*. ISBN: 978-0-9957075-2-8. 2015
- Krzysztof, Andrzej. *On natural join of posets properties and first applications*. Poland : Institute of Combinatorics and its Applications. 2009
- Kotler, Philip. Manajemen Pemasaran Edisi Milenium Jilid 1, Jakarta : Prehallindo. 2002
- Kusnawi. Pengantar Solusi Data Mining. Seminar Nasional Teknologi 2007 (SNT). Yogyakarta: STMIK AMIKOM. 2013
- Larose, Daniel T. *Discovering knowledge in data : an introduction to data mining*. New Jersey : A John Wiley & Sons , INC.Publication. 2005
- Latifurrahman. Perbandingan hasil penggerombolan *k-means*, *Fuzzy k-means*, dan *Two Step Clustering*. JPM IAIN Antasari. 2014

-
- Lupiyoadi, Rambat. *Manajemen Pemasaran dan Jasa*, Jakarta : Salemba Empat. 2013
- Muwardah, Fitroh Rizky. Pramuhendar, Ricardus. *Penentuan Penerimaan Siswa Baru Menggunakan Decision Tree*, Semarang : 15165 Universitas Dian Nuswantoro. 2015
- Nasari, Fina. Darma, Surya. *Penerapan K-Means Clustering Pada Data Penerimaan Mahasiswa Baru*. Medan : ISSN : 2302-3805. 2015
- Nisfiannoor, Muhammad. *Pendekatan Statistika Modern Untuk Ilmu Social*, Salemba Humanika. 2009
- N. Sharma, A. Bajpai, and R. Litoriya,. *Comparison the various klustering algorithms of weka tools*, Int. J. Emerg. Technol. Adv. Eng., vol. 2, no. 5. 2012
- Ong, Johan Oscar. *Implementasi Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing President University*. Bekasi : President University ISSN 1421-6869. 2013
- Putra, Yanuar S. *Analisis Matriks Boston Consulting Grup (BCG) Pada Pada Sepeda Motor Merk Honda*. Among Makarti, Vol.7 No.13. 2014
- Saifudin, Aries. *Metode Data Mining Untuk Seleksi Calon Mahasiswa pada Penerimaan Siswa Baru Di Universitas Pamulang*. Tangerang : urnal.umj.ac.id/Index.php/jurtext. 2017
- Widaningsih, Sri. Dan Agus Suheri. *Pengelompokkan Data Profil Dosen Kopertis Wilayah IV Menggunakan Metode Two Step Cluster Analysis* . Konferensi Nasional Sistem Informasi 2018 STMIK Atma Luhur Pangkalpinang. 2018