

## Teknik Klustering untuk Menganalisa Peminatan Sekolah pada Penerimaan Peserta Didik Online Tingkat SMA di Kota Semarang

Eko Nur Wahyudi

Program Studi Sistem Informasi, Universitas Stikubank Semarang

email : ekowahyudi157@gmail.com

### Abstrak

Sejak pemerintah kota Semarang menyelenggarakan penerimaan peserta didik baru secara online, khususnya untuk tingkat SMA Negeri di Kota Semarang, banyak permasalahan dapat diatasi dengan adanya sistem tersebut. Orang tua/wali dan calon peserta didik tidak lagi harus banyak mengeluarkan tenaga dan pikiran untuk melakukan proses pendaftaran, seperti pergi jauh menuju lokasi sekolah dan menunggu antrian yang cukup lama, tetapi cukup dengan melakukan proses secara online melalui internet. Yang sangat perlu diperhatikan adalah bagaimana proses yang berjalan secara real-time tersebut dapat dipantau secara transparan oleh masyarakat luas, sehingga keputusan yang harus diambil oleh orang tua/wali maupun calon peserta didik sepenuhnya tepat berdasarkan kesadaran dan kondisi yang ada pada saat proses penerimaan siswa baru secara *online* berlangsung.

Menjadi unik ketika mengamati bagaimana jumlah SMA yang ada di Kota Semarang diserbu oleh calon peserta didik yang begitu banyak, yang berasal dari negeri maupun swasta dan dari dalam kota maupun luar kota. Kuota yang ada pada setiap sekolah tentunya tidak dapat menampung semua jumlah pendaftar, sehingga ketentuan pada proses seleksi yang akan menjadi dasar akhir penerimaan. Tidak semua calon peserta didik yang berasal dari sekolah negeri bisa diterima di sekolah yang dituju, demikian pula calon peserta didik yang berasal dari sekolah swasta tetap memiliki kesempatan yang sama untuk bisa diterima di sekolah negeri tujuannya.

Menggunakan metode data mining dan teknik klustering diperoleh gambaran hasil bagaimana proses seleksi masuk di SMA Negeri di Kota Semarang yang cukup ketat namun transparan. Data yang diolah dapat mengetahui sejauh mana calon peserta didik berkompetisi memperebutkan kursi di sekolah yang diidamkannya tersebut. Gambaran hasil berikutnya adalah bagaimana porsi sebenarnya antara calon peserta didik laki-laki dan perempuan yang berkesempatan untuk melanjutkan studi di SMA Negeri. Prosentase pendaftar dari Dalam Rayon, Luar Rayon dan Luar Kota juga tampak jelas dalam hasil analisa. Terakhir adalah bagaimana dengan data mining dan teknik klustering diperoleh gambaran kelompok calon peserta didik dengan nilai tinggi, sedang dan rendah, serta kelompok jumlah calon peserta didik dari asal sekolah mereka.

**Kata kunci:** peserta didik, online, klustering

### PENDAHULUAN

Setiap tahun selalu ada masa dimana para orang tua/wali dan pihak sekolah, khususnya orang tua/wali dari peserta didik yang lulus dari Sekolah Menengah Pertama (SMP) akan mendaftar di Sekolah Menengah Atas/Kejuruan Negeri (SMA/K N), atau siswa dari Sekolah Dasar (SD) yang akan masuk ke SMP Negeri, dibuat dibuat sibuk dengan adanya proses Penerimaan Peserta Didik Baru yang dilaksanakan secara Online oleh Dinas

Pendidikan Kota Semarang. Dengan adanya sistem online ini proses pendaftaran dapat dilakukan dimana saja dan kapan saja, sebelum sampai dengan batas waktu calon peserta didik untuk datang langsung ke sekolah yang dipilihnya. Yang cukup menarik perhatian tentunya adalah transparansi informasi yang disajikan melalui laman [ppd.semarangkota.go.id](http://ppd.semarangkota.go.id) tersebut yang dapat dipantau setiap saat mengenai perkembangan jumlah pendaftar dan bagaimana pergerakan posisi pendaftar. Tak

jarang para orang tua/wali merasa was-was dengan perubahan posisi anaknya dalam daftar tabel pendaftar, apalagi bagi calon peserta didik yang memiliki nilai cukup, yang memilih di sekolah favorit. Jumlah SMP Negeri dan SMA/K Negeri di kota Semarang relatif cukup, namun demikian jumlah calon peserta didik yang sangat banyak tentunya tidak dapat semuanya ditampung di sekolah negeri. Karena itulah maka proses seleksi melalui sistem online menjadi sangat penting agar dapat dilaksanakan sesuai dengan ketentuan yang telah ditetapkan oleh Dinas Pendidikan Kota Semarang.

Hal yang tak kalah penting dengan adanya sistem penerimaan peserta didik baru secara online adalah mengetahui bagaimana sebenarnya sebaran atau peminatan siswa terhadap sekolah yang dipilihnya. Hal tersebut berkaitan erat dengan tujuan pemerintah dalam melakukan pemerataan pendidikan di semua tingkatan. Meskipun pemerintah kota telah menetapkan ketentuan berupa rayonisasi, nilai akademik, nilai prestasi dan lain-lain, namun tidak dapat dipungkiri bahwa ada sekolah yang sangat banyak peminatnya sebagai pilihan utama namun kapasitas penerimaannya sangat terbatas, tetapi masih ada juga sekolah yang peminatnya juga banyak namun menjadi pilihan kedua atau alternatif terakhir bagi calon peserta didik.

### **Perumusan Masalah**

Berdasarkan uraian singkat latar belakang di atas, maka permasalahan yang dapat dirumuskan adalah bagaimana mengetahui peminatan siswa terhadap sekolah yang dipilihnya menggunakan teknik klustering pada metode data mining.

### **Batasan Masalah**

Penelitian ini membatasi pengolahan data pada calon peserta didik yang akan mendaftar ke SMA Negeri di Kota Semarang pada tahun pelajaran 2014/2015, baik yang berasal dari sekolah negeri, swasta, dalam rayon, luar rayon, maupun yang berasal dari luar kota Semarang.

### **TUJUAN PENELITIAN**

Tujuan yang ingin dicapai dalam penelitian ini adalah mengetahui sejauh mana calon peserta didik SMA Negeri, orang tua/wali

dan pihak sekolah dalam memprediksi jumlah calon peserta didik yang mendaftar, asal calon peserta didik dan kecenderungan calon siswa dalam menentukan pilihan sekolahnya, sehingga untuk waktu yang akan datang dapat diambil langkah-langkah evaluasi yang tepat untuk menyempurnakannya.

### **MANFAAT PENELITIAN**

Manfaat yang bisa diambil dari hasil penelitian ini nantinya antara lain adalah :

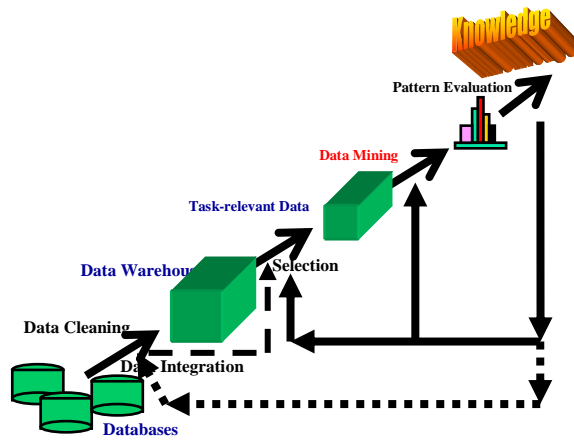
1. Dapat membantu orang tua/wali dan calon peserta didik SMA Negeri dalam merencanakan menentukan pilihan sekolah berdasarkan kelompok nilai yang dicapai pada penerimaan peserta didik baru sebelumnya
2. Dapat membantu pihak penyelenggara atau dalam hal ini adalah dinas pendidikan kota Semarang untuk memperbaiki sistem, prosedur atau syarat yang berlaku berdasarkan kelompok wilayah atau rayonisasi pada penerimaan peserta didik baru sebelumnya
3. Dapat membantu pihak SMA Negeri dalam mengatur strategi promosi agar peserta didik yang diterima sesuai dengan harapan yang akan dicapai.

### **TELAAH PUSTAKA**

#### **Data Mining**

Data mining merupakan suatu metode menemukan suatu pengetahuan dalam suatu database yang cukup besar. Data mining adalah proses menggali dan menganalisa sejumlah data yang sangat besar untuk memperoleh sesuatu yang benar, baru, sangat bermanfaat dan akhirnya dapat dimengerti suatu corak atau pola dalam data tersebut (Han & Kamber, 2006).

Data mining adalah bagian integral dari penemuan pengetahuan dalam database (KDD), yang merupakan proses keseluruhan mengubah data mentah menjadi informasi yang bermanfaat, seperti yang ditunjukkan pada Gambar 1.



Gambar 1. Proses penemuan pengetahuan dalam database / KDD (Han & Kamber, 2006)

Proses KDD ini terdiri dari serangkaian langkah-langkah transformasi, dari proses data *preprocessing* dan proses data *postprocessing* dari data yang merupakan hasil penggalian. Input data dapat disimpan dalam berbagai format (flat file, spreadsheet, atau relasional tabel) dan mungkin berada dalam penyimpanan data terpusat atau didistribusikan di beberapa alamat. Tujuan dari proses data *preprocessing* adalah untuk mengubah data input mentah menjadi format yang sesuai untuk analisis selanjutnya. Langkah-langkah yang dilakukan antara lain dengan memperbaiki data yang kotor atau ganda, dan memilih catatan dan fitur yang relevan dengan proses pengelolaan data selanjutnya. Karena banyak cara data dapat dikumpulkan dan disimpan, maka proses pengolahan data mungkin akan melelahkan dan memakan waktu yang lama dalam keseluruhan proses penemuan pengetahuan (Tan, dkk, 2006).

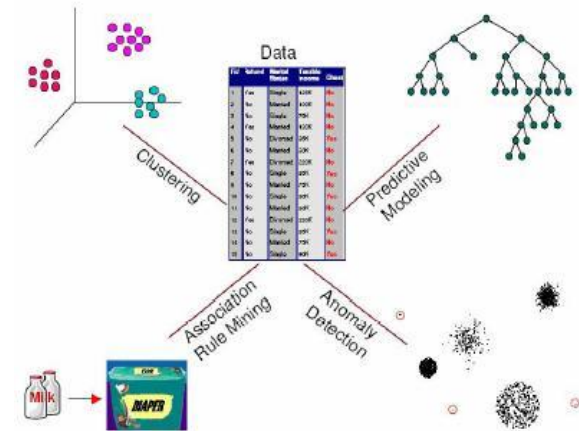
Terdapat empat tugas utama data mining seperti terlihat pada gambar 2, yaitu :

**Cluster Analysis**

Tidak seperti klasifikasi yang menganalisa kelas data obyek yang mengandung label. Clustering menganalisa objek data tanpa memeriksa kelas label yang diketahui. Label-label kelas dilibatkan di dalam data training. Karena belum diketahui sebelumnya. Clustering merupakan proses pengelompokkan sekumpulan objek yang sangat mirip.

**Anomaly Detection**

Anomaly detection merupakan metode pendeteksian suatu data dimana tujuannya adalah menemukan objek yang berbeda dari sebagian besar objek lain. Anomaly dapat di deteksi dengan menggunakan uji statistik yang menerapkan model distribusi atau probabilitas untuk data.



Gambar 2. Tugas utama data mining (Tan, dkk, 2006)

Entin Martiana, Nur Rosyid Muftada'i dan Edi Purnomo dalam penelitiannya yang berjudul Penggunaan Metode Pengklasteran untuk Menentukan Bidang Tugas Akhir Mahasiswa Teknik Informatika PENS Berdasarkan Nilai, memberikan hasil bahwa bidang tugas akhir dapat ditentukan berdasarkan nilai dari beberapa mata kuliah yang sudah dikelompokkan ke arah salah satu bidang tugas akhir. Metode clustering yang digunakan dalam penelitian tersebut adalah Single Linkage Hierarchical, Centroid Linkage Hierarchical dan K-Means. Hasilnya penelitian menunjukkan bahwa metode centroid linkage memiliki nilai variance paling kecil yang menandakan bahwa metode tersebut menghasilkan cluster yang paling baik dibandingkan dengan metode single linkage dan K-Means pada penelitian tersebut.

Selanjutnya Rahanimi dalam penelitiannya yang berjudul Peramalan Jumlah Mahasiswa Pendaftar PMDK Jurusan Matematika Menggunakan Metode Automatic Clustering dan Relasi Logika Fuzzy, bertujuan untuk memprediksi jumlah pendaftar jurusan

Matematika ITS melalui jalur PMDK regular. Metode yang digunakan adalah automatic clustering dan relasi logika fuzzy. Hasil yang diperoleh dari penelitian tersebut adalah untuk memperkenalkan metode automatic clustering dan relasi logika fuzzy dalam menyelesaikan masalah peramalan, sebagai referensi untuk pengembangan metode peramalan selanjutnya dan untuk mengetahui gambaran prediksi jumlah pendaftar PMDK jurusan matematika ITS untuk tahun yang akan datang.

Sedangkan dalam penelitian ini metode klustering digunakan dalam mengolah data calon peserta didik untuk mengetahui kelompok peserta didik yang dapat diterima di SMA Negeri berdasarkan nilai dan asal sekolah sehingga pihak penyelenggara penerimaan peserta didik online, yang dalam hal ini adalah dinas pendidikan kota Semarang dapat melakukan evaluasi dan perbaikan terhadap proses yang telah berjalan menjadi lebih baik lagi di masa yang akan datang. Sedangkan bagi orang tua/wali peserta didik dan calon peserta didik dapat digunakan sebagai acuan untuk merencanakan lebih awal terhadap pilihan sekolah yang diharapkan.

## METODE PENELITIAN

### Tahap Pengumpulan Data

Langkah pertama yang diambil sebagai awal untuk mendapatkan data valid agar dapat diproses lebih lanjut adalah dengan mengambil semua data pendaftar online calon peserta didik melalui laman [ppd.semarangkota.go.id](http://ppd.semarangkota.go.id). Data diambil dari setiap SMA Negeri yang dipilih, dalam hal ini terdapat 16 SMA Negeri di Kota Semarang. Data yang diambil memiliki struktur

asal seperti pada tabel 1.

Kemudian untuk melengkapi kebutuhan data dari tabel yang sudah ada ditambahkan 3 (tiga) buah kolom baru seperti terlihat pada tabel 2.

Tabel 2. Tambah Struktur Tabel

SMA	PILIHAN	DAYA TAMPUNG
1	1	522
1	1	522
1	1	522
...	...	...

Adapun tujuan dari penambahan kolom tersebut adalah ketika semua data digabung maka akan diketahui dengan pasti sekolah yang dituju oleh calon peserta didik dan prioritas pilihannya, yang nantinya akan disesuaikan dengan data tampung dari setiap sekolah.

### Tahap *Cleaning Data*

Data yang diperoleh dari setiap sekolah kemudian digabung menjadi satu dan dilakukan proses *Cleaning Data*, yaitu data yang masih belum valid akan dibersihkan terlebih melalui beberapa proses, antara lain :

- Edit Data, dimana data yang masih salah dalam penulisan akan disesuaikan, misalnya data seharusnya tertulis Semarang, masih ada yang ditulis Smg, Semrang, Smarang, atau Semaraang. Atau ada juga penulisan yang tidak standar juga akan disesuaikan, antara lain misalnya data yang seharusnya tertulis SMA Negeri, masih banyak yang ditulis SMAN, SMA N atau SMA Negeri.

Tabel 1. Struktur Tabel Asal

NO.	NO. PPD	NAMA	ASAL SEKOLAH	NILAI AKHIR	VERIFIKASI	RAYON
1	4010338	KHAFIDHA ERNING ARIYANTI	SMP NEGERI 15 SEMARANG	39.00	25/06/2014 10:30	DR
2	4010123	HAMDAN RAMADHANI	SMP IBU KARTINI	38.93	24/06/2014 11:02	DR
3	4010358	FRIDA RISVIANI CHAIRUNNISSA	SMP NEGERI 12 SEMARANG	38.60	25/06/2014 10:55	DR
4	4010786	AULIA NUR HIDAYATI	SMPN 1 BALIKPAPAN	38.42	27/06/2014 8:27	LK
...	...	...	...	...	...	DR

- b. Tambah Data, dimana proses yang dilakukan adalah dengan menambahkan data yang belum lengkap, misalnya SMA 14 menjadi SMA Negeri 14 Semarang, supaya dikenali perbedaan jika terdapat nama sekolah yang sama namun berasal dari kota yang berbeda.
- c. Hapus Data, jika terdapat data yang tidak sesuai dan tidak bisa di edit atau ditambah untuk menjadi valid, misalnya nama asal sekolah yang tidak jelas

4	SMA NEGERI 4	396
5	SMA NEGERI 5	384
6	SMA NEGERI 6	430
7	SMA NEGERI 7	353
8	SMA NEGERI 8	315
9	SMA NEGERI 9	354
10	SMA NEGERI 10	358
11	SMA NEGERI 11	432
12	SMA NEGERI 12	359
13	SMA NEGERI 13	287
14	SMA NEGERI 14	320
15	SMA NEGERI 15	349
16	SMA NEGERI 16	213
<b>Jumlah :</b>		<b>6.063</b>

**Tahap Pengolahan Data dan Visualisasi**

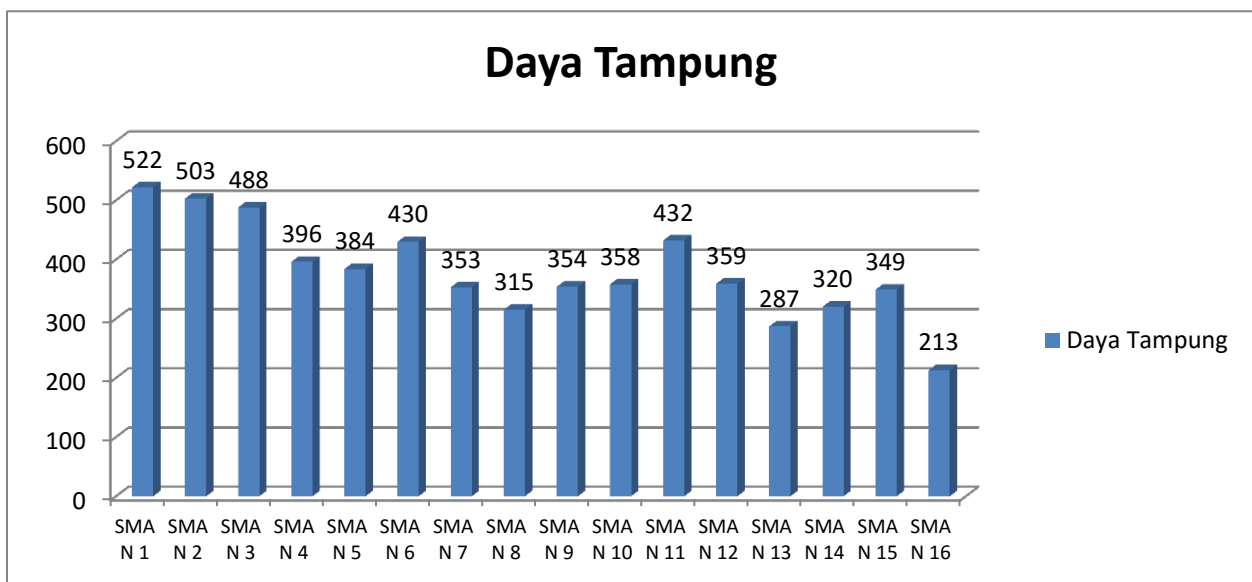
Data yang sudah valid kemudian diolah menggunakan rumus statistik dan teknik klustering untuk mendapatkan beberapa hasil, antara lain :

- a. Jumlah peserta didik yang akan diterima oleh SMA Negeri di Kota Semarang adalah sebanyak 6.063 orang seperti pada tabel 3 dan gambar 3.

Tabel 3. Daya Tampung Peserta Didik

No	Nama Sekolah	Daya Tampung
1	SMA NEGERI 1	522
2	SMA NEGERI 2	503
3	SMA NEGERI 3	488

- b. Jumlah seluruh calon peserta didik baru SMA Negeri di Kota Semarang. Nantinya hanya akan dipilih data Pilihan 1 yang menggambarkan minat utama pilihan calon peserta didik. Data tampak terlihat seperti pada tabel 4.

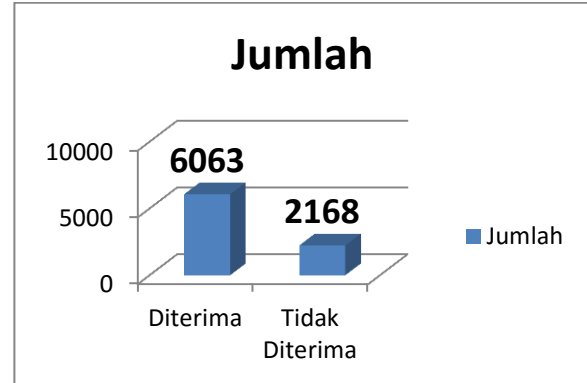


Gambar 3. Daya Tampung Peserta Didik

Tabel 4. Prosentase Pilihan 1 Terhadap Kuota

SMA	Calon	Kuota	%	Pilihan 1	%
1	1.208	522	231%	741	142%
2	972	503	193%	636	126%
3	891	488	183%	789	162%
4	803	396	203%	505	128%
5	1.188	384	309%	505	132%
6	1.186	430	276%	630	147%
7	1.113	353	315%	517	146%
8	783	315	249%	411	130%
9	1.207	354	341%	499	141%
10	838	358	234%	376	105%
11	1.469	432	340%	653	151%
12	854	359	238%	460	128%
13	1.043	287	363%	398	139%
14	946	320	296%	387	121%
15	1.398	349	401%	532	152%
16	689	213	323%	192	90%

Kota Semarang Pilihan 1 sebanyak 8.231, dari jumlah tersebut yang akan diterima sebanyak 6.063 (74%) dan yang tidak diterima sebanyak 2.168 (26%) terlihat pada tabel 5.

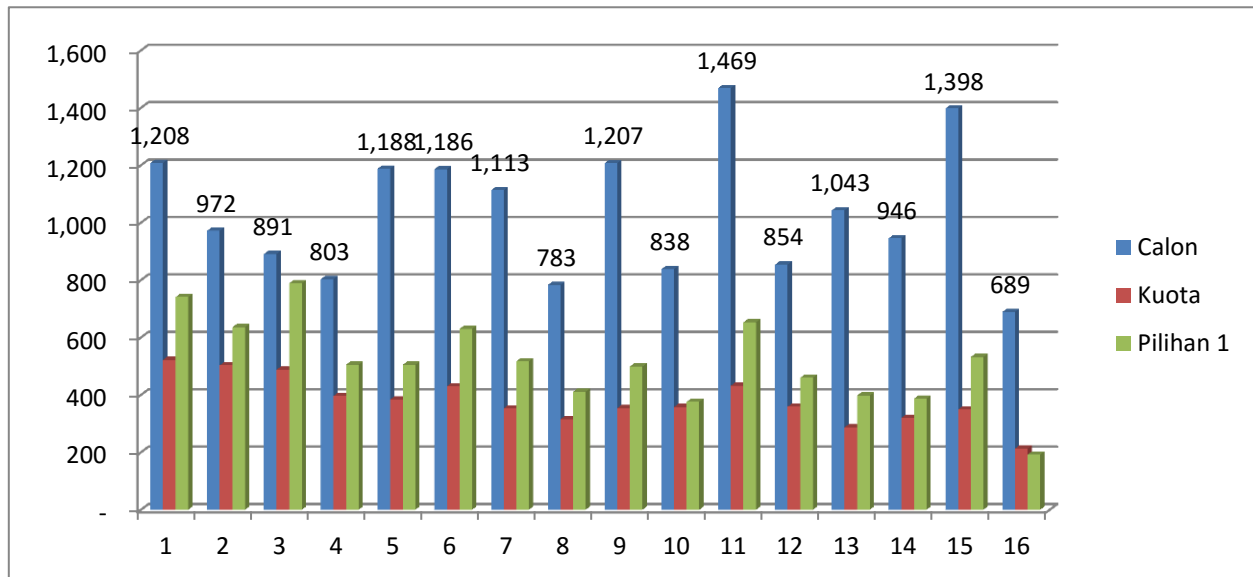


Gambar 5. Calon Peserta Didik Diterima/Tidak Diterima

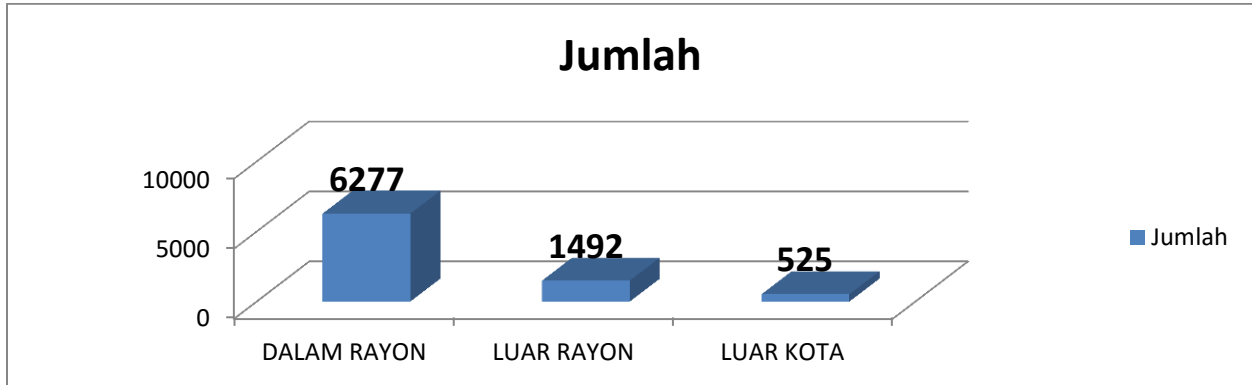
Jumlah calon peserta didik yang berasal dari Dalam Rayon sebanyak 6.277 (75,68%), dari Luar Rayon sebanyak 1.492 (17,99%), dan dari Luar Kota sebanyak 525 (6,33%) terlihat pada gambar 6.

Jumlah asal sekolah calon peserta didik yang mengikuti pendaftaran secara online sebanyak 501 sekolah

Jumlah calon peserta didik baru SMA Negeri di

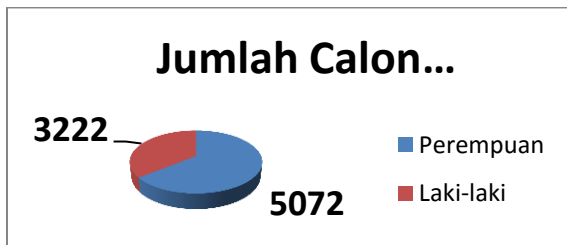


Gambar 4. Prosentase Pilihan 1 Terhadap Kuota



Gambar 6. Calon Peserta Didik Dalam/Luar Rayon dan Luar Kota

c. Jumlah calon peserta didik yang berjenis kelamin laki-laki sebanyak 3.222 (38,85%) dan perempuan sebanyak 5.072 (61,15%) terlihat pada gambar 7.



Gambar 7. Calon Peserta Didik Berdasar Jenis Kelamin

Hasil Kluster berdasarkan asal sekolah dengan jumlah calon peserta didik seperti pada gambar 8

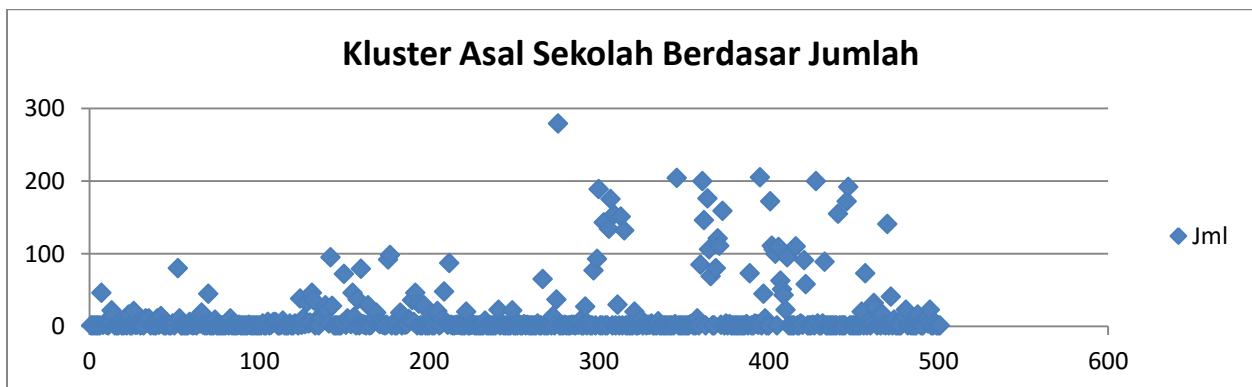
- Kluster 1 sebanyak 20 sekolah, dengan jumlah calon peserta didik antara 132 – 279 orang
- Kluster 2 sebanyak 38 sekolah, dengan

jumlah calon peserta didik antara 41 – 121 orang

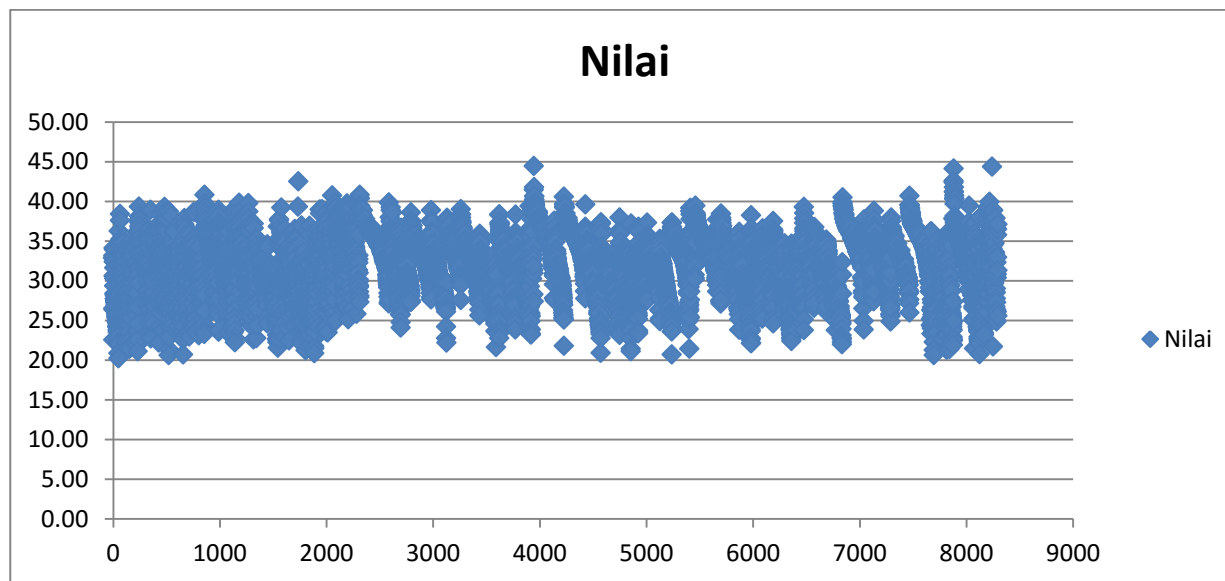
- Kluster 3 sebanyak 443 sekolah, dengan jumlah calon peserta didik antara 1 – 38 orang

Hasil Kluster berdasarkan nilai calon peserta didik seperti pada gambar 9.

- Kluster 1 sebanyak 2.835 orang (34%), dengan nilai antara 33,57 – 44,45
- Kluster 2 sebanyak 3.608 orang (44%), dengan nilai antara 29,10 – 33,56
- Kluster 3 sebanyak 1.851 orang (22%), dengan nilai antara 20,23 – 29,00
- Nilai Terendah calon peserta didik adalah 20,23
- Nilai Tertinggi calon peserta didik 44,45
- Nilai Rata-rata calon peserta didik 31,82



Gambar 8. Hasil Kluster Asal Sekolah Berdasar Nilai



Gambar 9. Hasil Kluster Berdasar Nilai

## HASIL DAN PEMBAHASAN

Hasil teknik *clustering* yang diterapkan terhadap data calon peserta didik yang masuk ke SMA Negeri di kota Semarang menunjukkan adanya pola peminatan secara umum, yaitu bagi calon peserta didik yang memiliki nilai tinggi akan cenderung memilih SMA Negeri dengan peringkat atas pula sebagai pilihan pertama, sedangkan pilihan kedua akan dijatuhkan pada SMA Negeri peringkat di bawahnya sebagai cadangan, meskipun diantaranya lebih memilih SMA Negeri dengan peringkat bawah untuk posisi lebih aman karena kemungkinan diterima menjadi lebih pasti seandainya pilihan pertama gagal. Hasil kluster juga menunjukkan bahwa calon peserta didik dengan nilai cukup atau kurang tidak sedikit yang berspekulasi memilih SMA Negeri dengan peringkat atas, dengan harapan bahwa pada saat proses seleksi berlangsung sebagian calon peserta didik akan mundur dan tempatnya akan diisi oleh mereka.

Proses *clustering* sendiri dibagi menjadi 3(tiga) kelompok nilai dengan harapan dapat menggambarkan pola pilihan mereka terhadap 3(tiga) kelompok peringkat SMA Negeri yang tersedia. Namun demikian tidak menutup kemungkinan untuk dikembangkan menjadi lebih banyak kelompok jika dikehendaki.

Persyaratan komponen penilaian bagi calon peserta didik, seperti anak guru atau kependidikan dan prestasi di tingkat wilayah, berdampak cukup signifikan terhadap nilai akhir, sedangkan rayonisasi hanya sedikit berpengaruh terhadap prediksi jumlah siswa yang akan diterima oleh setiap SMA Negeri.

Pada bagian akhir proses, setelah diperoleh hasil kelompok, masih ada sebuah penilaian yang hendaknya juga sudah dapat diperhitungkan sejak awal pada proses online, yaitu kuota sebesar 10% bagi keluarga kurang mampu. Harapannya adalah calon peserta didik tetap diseleksi dari mereka yang memiliki nilai akhir tinggi setelah dilakukan proses ranking, sehingga sekolah yang dipilih juga masih memiliki kualitas baik, di samping menjaga agar peserta didik juga tetap bisa mengikuti proses pembelajaran dengan baik sesuai standar yang diterapkan oleh masing-masing sekolah.

## SIMPULAN DAN SARAN

Simpulan yang dapat diambil dari penelitian ini sebagai berikut:

1. SMA Negeri di Kota Semarang mempunyai peminat calon peserta didik yang sangat tinggi, namun belum terdistribusi secara merata, terbukti dari sebaran asal sekolah



2. SMA Negeri dengan jumlah calon peserta didik terbanyak belum sepenuhnya berasal dari Nilai tertinggi, mengingat terdapat 10% kuota berasal dari masyarakat kurang mampu yang bisa jadi memiliki Nilai yang tidak tinggi
3. Hampir di semua SMA Negeri didominasi oleh calon peserta didik Perempuan dengan jumlah rasio 61,15% berdasarkan Nilai Tertinggi, sedangkan calon peserta didik Laki-laki hanya 38,85%.

Simpulan yang dapat diambil dari penelitian ini sebagai berikut:

1. Perlu diperoleh data mengenai Peminatan di SMK Negeri sebagai pembandingan peminatan terhadap semua SMA/SMK Negeri di Kota Semarang
2. Perlu dikembangkan ke arah orientasi peminatan asal sekolah terhadap sekolah yang menjadi tujuan calon peserta didik

#### DAFTAR PUSTAKA

- Yu, C.H., Digangi, S., Jannasch-Pennell, A.K., Kaprolet, C. (2008). *Profiling Students Who Take Online Courses Using Data Mining Methods*. Arizona State University
- Romero, C., Ventura, S., Expejo, P.G., Hervas, C. (2008). *Data Mining Algorithms to Classify Students*. Cordoba University
- Han, J., Kamber, M. (2000). *Data mining: Concepts and Techniques*. New York: Morgan-Kaufman.
- Otto, J., Sanford, D., Wagner, W. (2005). *Analisis Of Online Student Ratings Of University Faculty*. Towson University & Villanova University
- Al-Radaideh, Q.A., Al-Shawakfa, E.M., Al-Najjar, M.I. (2006). *Mining Student Data Using Decision Trees*. Yarmouk University
- Erdogan, S.Z., Timor, M. (2005). *A Data Mining Application In A Student Database*. Maltepe University & Istanbul University.

Trivedi, S., Pardos, Z.A., Heffernan, N.T. (2011). *Clustering Students to Generate an Ensemble to Improve Standard Test Score Predictions*. Worcester Polytechnic Institute, United States