

KOMPARASI ALGORITMA K-MEANS DAN K-MEDOIDS UNTUK MENANGANI STRATEGI PROMOSI DI POLITEKNIK TEDC BANDUNG

Novita Lestari Anggreini¹⁾ Shandy Tresnawati²⁾
Teknik Komputer, Politeknik TEDC Bandung^{1) 2)}
Email: vienovie@poltektedc.ac.id¹⁾, shandy.tresnawati@poltektedc.ac.id²⁾

Abstrak

Penelitian ini membahas mengenai perbandingan antara Algoritma *K-Means* dan Algoritma *K-Medoids* dalam klusterisasi, dan hasilnya dapat diusulkan menjadi pertimbangan untuk memutuskan dalam menangani strategi promosi di Politeknik TEDC Bandung dalam kegiatan Penerimaan Mahasiswa Baru. Peneliti melakukan pengamatan langsung dan juga melalui sesi wawancara di bagian Promosi Penerimaan Mahasiswa Baru Politeknik TEDC Bandung untuk mengumpulkan data-data dan informasi yang dibutuhkan. Wawancara tersebut dilakukan langsung dengan mahasiswa tahun akademik 2015 sampai dengan tahun akademik 2017 mengenai seberapa mudah mereka mendapatkan informasi tentang Penerimaan Mahasiswa Baru di Politeknik TEDC Bandung. Proses klusterisasi memerlukan beberapa data sebagai data masukan yang meliputi Nama Pendaftar, Tempat dan Tanggal lahir, Alamat, Agama, Nomor Telepon, Asal Sekolah, Program Studi Pilihan, dan Pembawa Informasi, yang diperoleh dari data promosi tahun akademik 2015 sampai dengan tahun akademik 2017. Setelah dilakukan pembersihan dan perbaikan data, dapat diperoleh suatu *dataset* yang harus dibuat lebih detail, seperti data Alamat yang kemudian dipecah lagi menjadi Kecamatan, Kota dan Propinsi agar dapat memberikan informasi yang lebih spesifik dalam pengelompokan data. Ada juga beberapa data yang tidak digunakan sehingga oleh peneliti dibuang agar lebih efektif dalam melakukan pengujian seperti Nama Pendaftar, Tempat dan Tanggal lahir, Agama, dan Nomor Telepon. Dalam penelitian ini juga peneliti menambahkan atribut baru seperti Jenis Sekolah yang dapat dibagi menjadi dua yaitu SMA dan SMK. Atribut ini dibuat agar dapat terlihat lebih jelas mengenai jenis sekolah yang mana yang lebih banyak mendaftar ke Politeknik TEDC Bandung.

Kata Kunci: *k-means*, *k-medoids*, promosi, klusterisasi, komparasi.

Abstract

This study discusses the comparison between the K-Means Algorithm and the K-Medoids Algorithm in clustering, and the results can be proposed to consider a decision for handling promotional strategies at Politeknik TEDC Bandung in New Student Admissions activities. The researcher made direct observations also through an interview session in the Promotion Section for Politeknik TEDC Bandung New Student Admissions to collect the data and information needed. The interview was conducted directly with students in the 2015 academic year up to the 2017 academic year about how easily they got information about Admission to New Students at the Politeknik TEDC Bandung. The clustering process requires some data as input data which includes the Registrant Name, Place and Date of Birth, Address, Religion, Telephone Number, School of Origin, Elective Study Program, and Information Carrier, obtained from the 2015 academic year promotion data up to 2017 academic year. After cleaning and repairing the data, a dataset can be obtained that is more detailed, such as Address data which is then broken down into Sub-Districts, Cities and Provinces to provide more specific information in grouping data. Some data are not used so that researchers are discarded to be more effective in conducting tests such as Registrant Name, Place and Date of Birth, Religion, and Telephone Number. In this study, researchers also added new attributes such as the type of school that can be divided into two: high school and vocational school. This attribute is made so that it can be seen more clearly about the types of schools which are more registered with Politeknik TEDC Bandung.

Keywords: *k-means*, *k-medoids*, promotion, clustering, comparison.

I. PENDAHULUAN

Banyaknya Perguruan Tinggi di Propinsi Jawa Barat menjadi faktor utama dalam mencari strategi promosi terbaik untuk Politeknik TEDC Bandung. Politeknik TEDC Bandung merupakan salah satu Perguruan Tinggi Vokasi yang sedang berkembang di Indonesia khususnya Jawa Barat. Politeknik TEDC Bandung selalu melakukan yang terbaik untuk memberikan pendidikan yang layak kepada calon-calon mahasiswanya. Seiring dengan perkembangan zaman modern ini, serta teknologi dan informasi khususnya di lingkungan perguruan tinggi vokasi, maka daya saing antar perguruan tinggi pun semakin meningkat untuk mencari dan mendapatkan

mahasiswa baru. Politeknik TEDC Bandung juga melakukan promosi ke seluruh Indonesia untuk menambah jumlah penerimaan mahasiswa baru. Setelah diamati dapat diketahui bahwa teknik promosi yang dilakukan selama ini tidak efektif sehingga kurang meratanya informasi mengenai Politeknik TEDC Bandung yang menyebabkan masih banyak sekali yang tidak serta belum mengetahui mengenai Politeknik TEDC Bandung.

Pada penelitian ini, peneliti melakukan penelitian dengan mengambil data calon mahasiswa baru dari tahun akademik 2015 sampai tahun akademik 2017. Pada kisaran tahun tersebut ada beberapa promosi yang dilakukan oleh Politeknik TEDC Bandung,

antara lain melalui Koran Tribun Jabar, Presentasi ke SMA maupun ke SMK, memasang baliho di beberapa titik kota Cimahi, Kota Bandung, Kab. Bandung Barat dan Kab. Bandung, memasang informasi tentang Politeknik TEDC Bandung di laman *website*, bekerjasama dengan perguruan tinggi lain seperti Politeknik Negeri Bandung dan Politeknik Manufaktur Bandung, menghubungi beberapa alumni yang bekerja di industri maupun di SMA dan SMK, serta menawarkan program beasiswa yang diberikan oleh pemerintah. Pada seluruh kegiatan promosi yang dilakukan selama ini adanya ketidakseimbangan dalam pemilihan program studi sehingga jumlah mahasiswa yang masuk tidak sama atau tidak rata.

Ada beberapa penelitian terdahulu yang menjadi referensi dalam penelitian ini antara lain:

- a. Pada Penelitian Asril, dkk (2015) yang berjudul Analisis Data Lulusan dengan *Data Mining* untuk Mendukung Strategi Promosi Universitas Lancang Kuning memiliki tujuan untuk menentukan strategi promosi yang tepat pada sasaran agar dapat mengurangi biaya untuk promosi dan mencapai target promosi yang tepat pula, sehingga mendapatkan hasil yang berupa informasi penting sehingga dapat mendukung strategi promosi dalam mencari calon mahasiswa baru. Penelitian yang dilakukan oleh Asril, dkk (2015) menjadi salah satu motivasi untuk peneliti agar dapat menentukan strategi terbaik dalam menarik calon mahasiswa baru sehingga tepat sasaran sehingga dapat mengurangi biaya promosi yang terkadang diluar Rencana Anggaran Biaya (RAB).
- b. Pada penelitian yang dilakukan oleh Ong Johan Oscar (2015) bertujuan untuk memperlihatkan bahwa hasil dari pengolahan data mahasiswa dapat membantu pihak *Marketing President University* untuk melakukan promosi dan mencari calon mahasiswa baru di berbagai kota yang ada di Indonesia, sehingga menghasilkan informasi yang cukup efisien serta efektif. Metode dalam mengolah data yang digunakan dalam penelitian Ong (2015) ini merupakan *Data Mining*, data calon mahasiswa yang telah didapatkan, pada tahap pengumpulan data di awal kemudian diolah untuk didapatkan pola dari data-data tersebut sehingga terlihatlah informasi yang tersembunyi.
- c. Data yang didapatkan dalam jumlah besar yang memungkinkan penggunaan *Data Mining* dengan teknik pengelompokan atau klusterisasi teori ini disebutkan dalam Penelitian Nugraha, dkk (2016). Pada Penelitian ini peneliti mendapatkan data yang cukup banyak, sebanyak 1411 *dataset* sehingga memungkinkan untuk dapat melakukan teknik *clustering*.

Dari beberapa penelitian yang telah dilakukan oleh beberapa peneliti lainnya, peneliti merasa sangat tertarik untuk melakukan penelitian dengan menggunakan Algoritma *K-Means* dan Algoritma *K-Medoids* untuk menangani Strategi Promosi di Politeknik TEDC Bandung.

II. LANDASAN TEORI

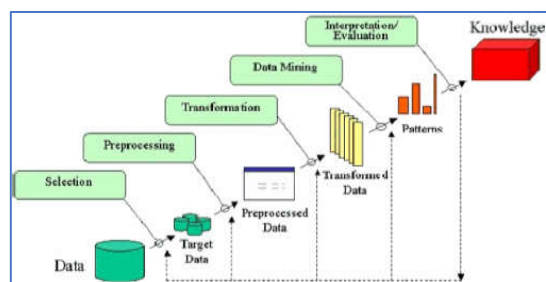
Data Mining

Istilah *Data Mining* digunakan untuk menjelaskan atau memaparkan penemuan ilmu pengetahuan di dalam *database*. *Data Mining* merupakan proses yang menggunakan teknik statistik, *mathematics*, *Artificial Intelligence* (AI), dan *machine learning* untuk meringkas dan menentukan informasi yang berguna dan pengetahuan yang terakit dari berbagai *database* besar (Kusrini dkk, 2009).

Beberapa proses untuk mengeluarkan nilai tambah dari suatu kelompok data yang berupa pengetahuan yang selama ini tidak diketahui secara manual disebut dengan *Data Mining*. Di mana saat mendapatkan hasil dari proses-proses yang telah dilakukan tersebut akan membentuk beberapa pola berupa kumpulan data, yang sering disebut dengan pengenalan pola (*pattern recognition*). Ada beberapa bagian pada *data mining* salah satunya adalah pengenalan pola.

Data Mining juga sering disebut *knowledge discovery* yang merupakan proses pengambilan pola pada data yang akan di proses lalu mendapatkan *keluaran* berupa informasi yang sangat pokok dan bernilai. Dimana data tersebut merupakan kumpulan fakta dan dapat memberikan gambaran, jadi pada saat kita melakukan sesuatu dalam pengambilan data maka data tersebut langsung tersimpan dan pola-pola data itu akan diteliti secara manual sehingga kita dapat mengetahui hal-hal yang akan terjadi. *Data mining* sering disebut *Knowledge Discovery in Database* (KDD), yaitu kegiatan yang mencakup pengumpulan data, pemakaian data historis untuk menemukan kesamaan keadaan, pola atau hubungan dalam set data yang berukuran besar (Santosa, 2007).

Proses logika *Knowledge Discovery in Database* bisa dilihat pada gambar dibawah ini.



Gambar 1. Proses Logika *Knowledge Discovery in Database* (KDD)

Berikut penjelasannya:

- a. *Data Selection*
Pemilihan data (seleksi data) dari beberapa data pokok perlu dilakukan sebelum memulai tahap mengeluarkan informasi dalam *Knowledge Discovery in Database* (KDD). Data yang telah didapatkan dari seleksi akan digunakan untuk proses *data mining* dan juga dapat disimpan secara terpisah dari basis data operasional.
- b. *Pre-processing/Cleaning*

Perlunya proses pembersihan data sebelum proses *data mining* dilakukan yang menjadi inti dalam *Knowledge Discovery in Database* (KDD).

c. *Transformation*

Perubahan yang terjadi pada data yang telah dipilih disebut juga dengan *Coding*, sehingga data tersebut sudah sesuai untuk proses *data mining*. Proses *coding* pada *Knowledge Discovery in Database* (KDD) merupakan sebuah proses telaah untuk menciptakan suatu kreasi dan sangat tergantung pada jenis atau pola informasi yang akan dicari di basis data.

d. *Data Mining*

Di dalam proses mencari pola atau informasi dalam data terpilih dapat menggunakan teknik tertentu yang disebut *Data Mining*. Teknik, metode ataupun algoritma dalam *data mining* sangat beraneka ragam. Penetapan cara atau teknik, algoritma yang tepat sangat bergantung penuh pada tujuan dan proses *Knowledge Discovery in Database* (KDD).

e. *Interpretation/Evaluation*

Proses pada tahap ini disebut *interpretation/Evaluation*. Proses di tahap ini mencakup pemeriksaan pola atau informasi data yang ditemukan apakah sudah benar dan tidak bertentangan dengan fakta atau hipotesis yang telah ada sebelumnya.

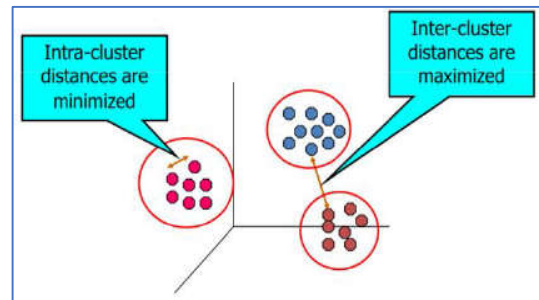
f. *Knowledge Presentation* (Presentasi Pengetahuan)

Pada tahap ini akan terlihat gambaran dan penyajian pengetahuan mengenai Teknik atau metode yang digunakan untuk memperoleh pengetahuan yang didapatkan oleh pengguna.

Clustering

Clustering adalah suatu Teknik pengelompokan *record*, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki 18 kemiripan. *Cluster* adalah kumpulan beberapa rekaman yang memiliki suatu kemiripan dengan yang lainnya dan memiliki ketidakmiripan dengan beberapa rekaman di *cluster* yang lain. *Clustering* tidak dapat disamakan dengan klasifikasi, yaitu tidak ada variabel target dalam *clustering*. *Clustering* tidak akan pernah mencoba untuk melakukan klasifikasi, menduga, atau memperkirakan nilai dari variabel target. Tetapi, pada algoritma *clustering* mencoba untuk melakukan pembagian terhadap semua data yang ada menjadi beberapa kelompok yang memiliki kedekatan/kesamaan (homogen), di mana kesamaan *record* dalam satu kelompok akan bernilai lebih besar, sedangkan kemiripan dengan *record* di kelompok lain akan bernilai lebih kecil (Kusrini et al., 2009).

Clustering melakukan pengelompokan data yang didasarkan pada berapa banyak persamaan antar satu objek ke objek yang lain, oleh karena itu klasterisasi dikelompokkan sebagai metode *unsupervised learning* bukan *supervised learning* (Santosa, 2007). Logika dari konsep *clustering* diperlihatkan pada Gambar 2 dibawah ini.



Gambar 2. Konsep Clustering (Han, 2006)

K-Means

Algoritma *K-Means* merupakan suatu Algoritma yang digunakan untuk menyelesaikan masalah *clustering* (Fang, 2015). *K-Means* juga merupakan algoritma untuk mengelompokkan atau mengklasifikasi objek berdasarkan atribut/fitur ke dalam *k* kelompok. *k* merupakan bilangan positif. Tata cara dalam pengelompokan dengan meminimalkan jumlah kuadrat dari jarak antara data dan *cluster centroid* yang bersangkutan. Dengan begitu, tujuan pokok dari *k-means clustering* adalah untuk mengelompokkan data. Efisiensi serta keautentikan algoritma *K-Means* sangat bergantung pada *centroid* awal. Berikut ini langkah kerja yang dilakukan oleh algoritma *K-Means* adalah : **1)** Tentukan terlebih dahulu ada berapa *cluster k* dan *centroid* awal, **2)** Melakukan penghitungan jarak antara satu objek ke objek lain kemudian ke *centroid*, **3)** Mengelompokkan objek berdasarkan jarak minimum objek ke *centroid*, **4)** Menetapkan *centroid* kembali apabila terjadi perpindahan kluster dengan melakukan perhitungan nilai rata-rata dari seluruh anggota pada masing-masing kluster, **5)** Lakukan pengulangan lagi langkah (1) sampai langkah (4) sampai tidak akan terjadi lagi perpindahan kluster.

Untuk mengakhiri algoritma adalah ketika semua *record* yang dimiliki oleh masing-masing pusat *cluster* tetap dalam *cluster* itu. Atau, algoritma bisa berhenti ketika beberapa kriteria konvergensi terpenuhi, seperti ada penyusutan yang tidak signifikan dalam jumlah kuadrat *error* (*sum of squared errors*).

K-Medoids

Teknik *Partitional Clustering* yang meminimalkan jarak antara titik berlabel dalam *cluster* dan titik yang ditunjuk sebagai pusat kluster tersebut, Algoritma *K-Medoids* salah satunya. Algoritma *K-Medoids* sangat berbeda dengan algoritma *K-Means*, *K-Medoids* memilih data points sebagai pusat (*medoids*).

K-Medoids adalah teknik atau metode partisi klasik *Clustering* yang melakukan pengelompokan *dataset* dari objek ke dalam kelompok *k* yang dikenal dengan apriori. Perbandingan *K-Medoids* dengan *K-Means*, didapatkan bahwa *K-Medoids* lebih kuat dalam menangani kebisingan (*noise*) serta pencilan (*outlier*) karena dapat meminimalkan beberapa *dissimilarities* yang berpasangan, bukan jumlah kuadrat jarak *Euclidean*. Suatu *medoid* bisa juga diartikan sebagai objek *cluster* yang rata-rata

perbedaan untuk semua objek dalam *cluster* minimal yaitu titik paling berlokasi di *cluster*.

Strategi Promosi

Ada beberapa pengertian mengenai strategi promosi, yaitu bahwa strategi promosi adalah kegiatan perusahaan/lembaga untuk menggerakkan penjualan dengan mengarahkan komunikasi-komunikasi yang meyakinkan kepada para pemakai (Moekijat, 2000). Ada juga pendapat lain dari Lamb et al. (2001) yang mengatakan bahwa strategi promosi adalah merencanakan penggunaan yang maksimal dari elemen-elemen promosi seperti periklanan, hubungan kepada masyarakat, transaksi pribadi dan *marketing* penjualan.

Dari beberapa pengertian itu dapat diketahui bahwa strategi promosi adalah kegiatan yang disusun secara matang dengan maksud menggoda, mendorong pembeli agar mau atau memiliki keinginan untuk membeli produk perusahaan, sehingga tujuan dalam meningkatkan penjualan diharapkan dapat dicapai.

Terdapat pula dua macam promosi yang ada, yaitu promosi pasif dan promosi aktif. Sesuai dengan judul penelitian ini, promosi pasif digunakan untuk meningkatkan kualitas internal seperti melalui karyawan ataupun dosen Politeknik TEDC Bandung, alumni, serta kakak tingkat, Sedangkan promosi aktif adalah dengan cara melakukan promosi seperti memasang baliho di beberapa tempat, mencetak brosur dan menyebarkannya, melakukan kunjungan ke SMA maupun SMK dalam rangka mempromosikan Politeknik TEDC Bandung.

RapidMiner

RapidMiner merupakan salah satu *tool* analisis yang berupa *software* untuk pengolahan *data mining*. RapidMiner juga sebuah lingkungan (*environment*) untuk *machine learning*, *data mining*, *text mining*, dan *predictive analytics*. *Machine learning* adalah suatu algoritma di mana perilaku komputer berevolusi berdasarkan data empiris, seperti sensor ataupun *database*. *Data mining* merupakan suatu proses merangkai pola-pola dari *dataset* yang besar dengan mengombinasikan metoda statistika, kecerdasan buatan dan *database*.

III. METODE PENELITIAN

Metode penelitian yang digunakan pada penelitian ini adalah metode analisis deskriptif dengan pendekatan kuantitatif, artinya penelitian ini menekankan analisisnya pada data-data yang terdapat di bagian Promosi Penerimaan Mahasiswa Baru yang bertujuan untuk mendapatkan visualisasi yang jelas mengenai strategi promosi yang paling baik serta sesuai untuk meningkatkan Penerimaan Mahasiswa Baru di Politeknik TEDC Bandung.

Pengumpulan data dalam penelitian ini menggunakan beberapa cara sebagai berikut :

a. Observasi

Observasi yang dilakukan adalah dengan cara mengamati secara langsung di bagian Promosi Politeknik TEDC Bandung, serta secara cermat

dan sistematis mengumpulkan data calon mahasiswa baru tahun akademik 2015 sampai dengan tahun akademik 2017. Peneliti juga mengumpulkan informasi berapa persen target yang tercapai dalam promosi penerimaan mahasiswa baru dalam rentang waktu tiga tahun.

b. Wawancara

Peneliti juga melakukan wawancara secara langsung pada bagian promosi Politeknik TEDC Bandung terkait pendataan Penerimaan Mahasiswa Baru. Wawancara dilakukan oleh peneliti terhadap para mahasiswa tahun akademik 2015 sampai tahun akademik 2017 mengenai seberapa mudah mereka memperoleh informasi mengenai penerimaan mahasiswa baru di Politeknik TEDC Bandung, maupun mencari tahu dari mana mereka memperoleh informasi penerimaan mahasiswa baru tersebut.

c. Studi Literatur

Peneliti melakukan studi literatur dengan membaca buku-buku, jurnal, dan paper yang berkaitan serta dapat mendukung penelitian ini.

Data yang akan dianalisis disini adalah data calon mahasiswa baru dari tahun akademik 2015 sampai dengan tahun akademik 2017 yang didapatkan dari bagian Promosi Penerimaan Mahasiswa Baru Politeknik TEDC Bandung yang kemudian data tersebut disesuaikan dengan kebutuhan dari penelitian ini sehingga akan didapat strategi promosi yang lebih efektif dan efisien. Ada beberapa atribut yang dibutuhkan dalam penelitian ini adalah : **1)** Kecamatan, **2)** Kota, **3)** Propinsi, **4)** Asal Sekolah, **5)** Jenis Sekolah, **6)** Program Studi Pilihan, dan **7)** Pembawa Informasi.

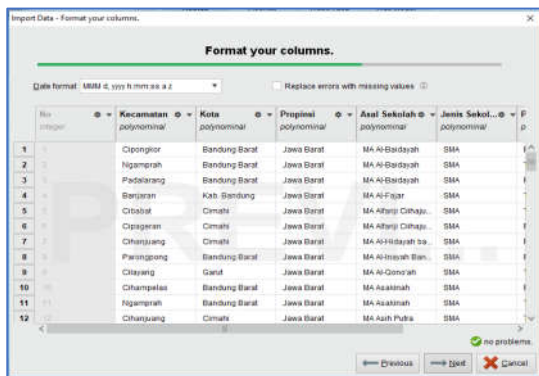
IV. HASIL DAN PEMBAHASAN

Pada penelitian ini Pengumpulan Data yang diperoleh pada bagian Promosi Politeknik TEDC Bandung adalah data mahasiswa baru dari tahun akademik 2015 sampai tahun akademik 2017 sebanyak 1.411 mahasiswa. Data yang didapatkan oleh peneliti terdiri dari Nama Pendaftar, Tempat Tanggal Lahir, Alamat, Agama, Nomor HP/Telepon, Asal Sekolah, Program Studi Pilihan, dan Pembawa Informasi.

Setelah dilakukan *cleaning* dan *filtering* maka diperoleh *dataset* yang lebih detail, Data Alamat dipecah menjadi tiga buah atribut baru, yaitu Kecamatan, Kota, dan Propinsi. Hal ini dilakukan agar peneliti dapat mengetahui lebih spesifik daerah mana yang berpotensi besar dalam penyebaran informasi Promosi Penerimaan Mahasiswa Baru di Politeknik TEDC Bandung. Ada juga beberapa data yang tidak efektif sehingga dibuang oleh peneliti seperti Nama Pendaftar, Tempat Tanggal Lahir, Agama dan Nomor HP/Telepon. Pada penelitian ini juga menambahkan sebuah atribut baru, yaitu Jenis Sekolah, yang dibedakan menjadi SMA dan SMK, dengan alasan agar dapat mengetahui jenis sekolah mana yang lebih banyak mendaftar di Politeknik TEDC Bandung.

Dataset yang akan dimasukkan ke dalam Rapidminer adalah dataset yang dibuat dengan aplikasi *Microsoft Excel*, kemudian pilih *file* yang akan dipakai dalam penelitian ini. Setelah *file* dimasukkan, kemudian tentukanlah masing-masing indikator atribut yang terdapat di dalamnya. Karena semua atribut berisi lebih dari dua klasifikasi maka tipe dari atribut tersebut adalah *polynomial*.

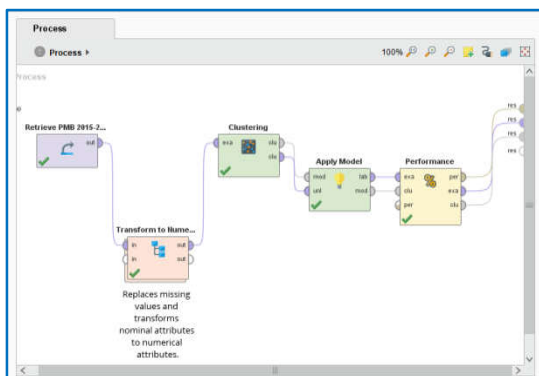
Pada Gambar 3 dibawah ini dapat dilihat cara pengaturan dataset dalam melakukan masukkan ke dalam *tool RapidMiner*.



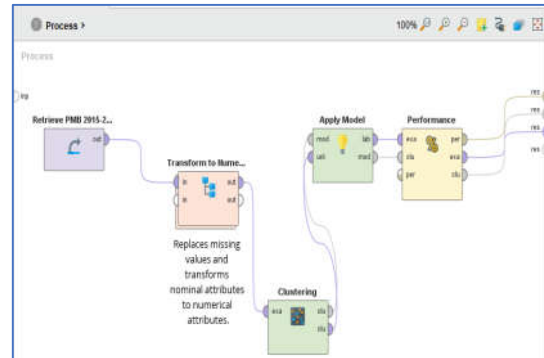
Gambar 3. Pengaturan dan Proses Impor Dataset ke dalam *RapidMiner*

Peneliti juga membuat desain clustering menggunakan algoritma *K-Means* dan *K-Medoids* seperti yang ditunjukkan pada Gambar 4 dan 5.

Pada Gambar 4 dan 5 dapat dilihat bahwa proses pengujian dimulai dengan *retrive* data PMB 2015-2017, artinya dataset disimpan dengan nama PMB 2015-2017 Deal. Dataset itu kemudian disambungkan terlebih dahulu ke *Transform to Numerical* sebelum dilakukan pengujian menggunakan *Algoritma Clustering K-Means* dan *K-Medoids*. Setelah disambungkan ke *Algoritma Clustering K-Medoids* maka selanjutnya disambungkan lagi ke *Apply Model*, lalu ke *Performance* dan terakhir disambungkan ke *Result*.

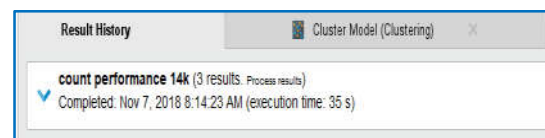


Gambar 4. Desain *Clustering* Menggunakan Algoritma *K-Means*

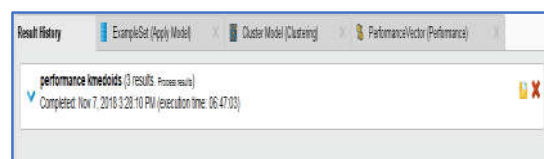


Gambar 5. Desain *Clustering* Menggunakan Algoritma *K-Medoids*

Setelah pengujian berjalan maka didapatkan informasi bahwa lama waktu pemrosesan clustering dengan Algoritma *K-Means* sebesar 35 detik sedangkan pemrosesan clustering dengan Algoritma *K-Medoids* sebesar 6 Jam lebih, seperti yang tercatat pada Gambar 6 dan 7.

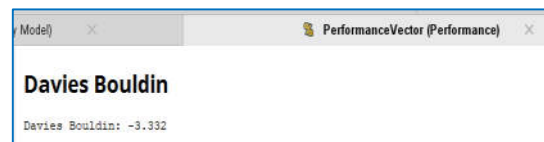


Gambar 6. *Result History Clustering* menggunakan Algoritma *K-Means*

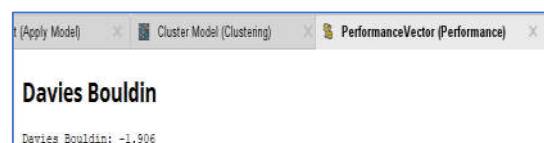


Gambar 7. *Result History Clustering* menggunakan Algoritma *K-Medoids*

Dapat dilihat juga *Davies Bouldin Index* (DBI) yang ada didalam *Performance Vector K-Means* pada gambar 8. DBI merupakan metode validasi cluster dari hasil *clustering*. Semakin rendah nilai DBI, maka cluster tersebut semakin baik.



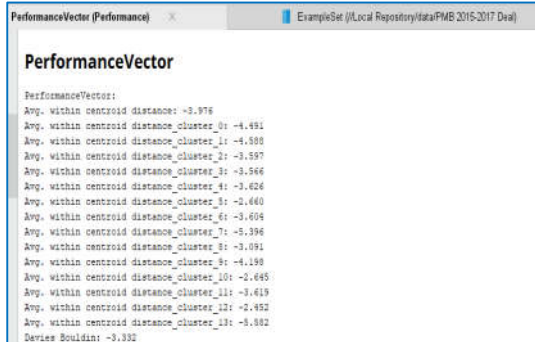
Gambar 8. *Davies Bouldin Index* untuk *Performance Vector K-Means*



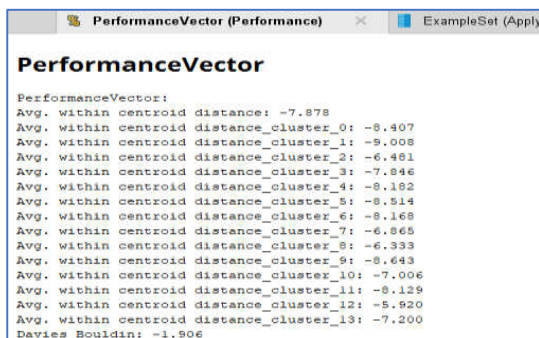
Gambar 9. *Davies Bouldin Index* untuk *Performance Vector K-Medoids*

Hasil dari nilai *Davies Bouldin Index* (DBI) yang terdapat di dalam *Performance Vector K-Medoids* ditunjukkan melalui Gambar 8 dan 9.

Untuk menunjukkan deskripsi Performance yang menghasilkan Centroid Distance pada setiap cluster dapat dilihat pada Gambar 10 dan 11.

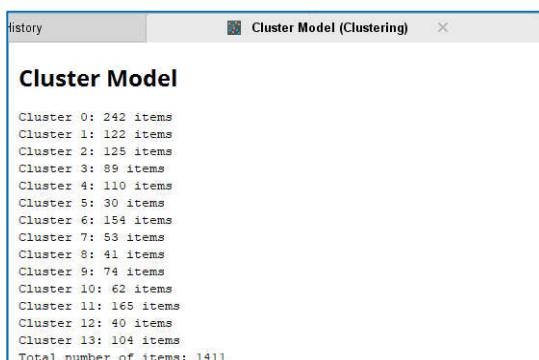


Gambar 10. Centroid Distance K-Means

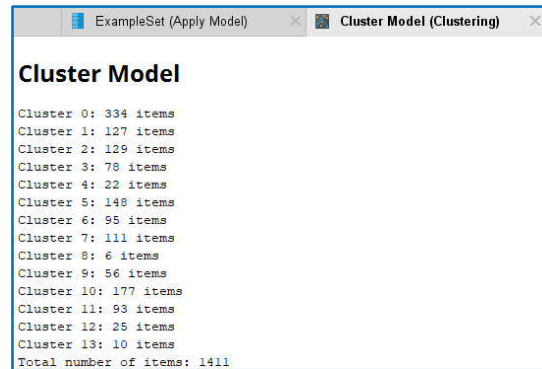


Gambar 11. Centroid Distance K-Medoids

Pada Gambar 12 dan 13 Menunjukkan bahwa pada penelitian ini *cluster* yang digunakan berjumlah 14, dimulai dari *cluster* 0 dan diakhiri dengan *cluster* 13, total *dataset* yang di-*clustering* sebanyak 1.411 dengan jumlah *items* pada tiap-tiap *cluster*.

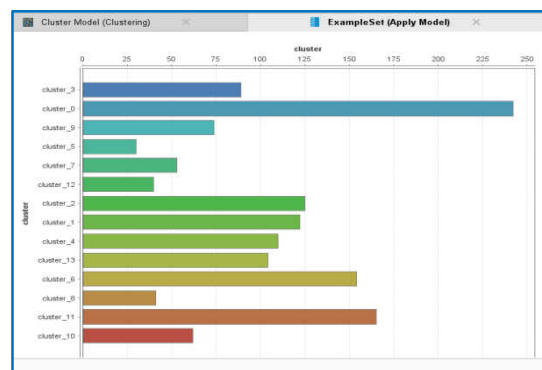


Gambar 12. Hasil Cluster Model K-Means



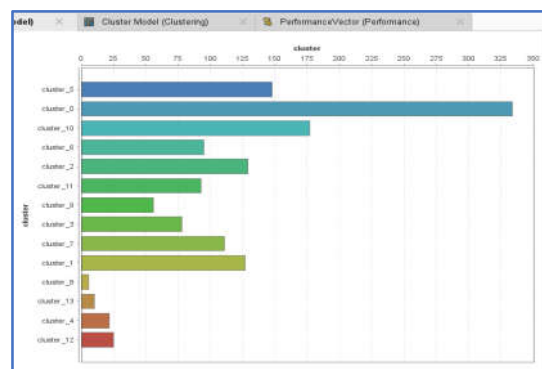
Gambar 13. Hasil Cluster Model K-Medoids

Pada penelitian ini hasil dari *Cluster Model K-Means* dapat dilihat dalam bentuk *Charts Bars* seperti Gambar 14, *cluster* Kerjasama Perguruan Tinggi memiliki nilai paling tinggi, dan *cluster* Orang Tua memiliki nilai paling rendah dalam *set* data Pembawa Informasi.



Gambar 14. Charts Bars Hasil Cluster Model dengan Metode K-Means

Pada penelitian ini juga dapat dilihat hasil dari *Cluster Model K-Medoids* dapat digambarkan dalam bentuk *Charts Bars* seperti pada Gambar 15, data Jenis Sekolah memiliki nilai paling tinggi dan kota Sumedang memiliki nilai paling rendah.



Gambar 15. Charts Bars Hasil Cluster Model dengan Metode K-Medoids

Peneliti memakai *Confusion Matrix* untuk menguji presisi dan akurasi dari algoritma *K-Means* dan algoritma *K-Medoids* yang dibandingkan dalam penelitian ini. Dengan pengujian melalui *Confusion*

Matrix dapat diketahui performansi suatu algoritma berdasarkan nilai akurasi algoritma itu.

Pengujian dengan *Confusion Matrix* untuk algoritma *K-Means* dan algoritma *K-Medoids* diperinci pada Tabel 1 dan Tabel 2.

Tabel 1. *Confusion Matrix* Pengujian Analisis *K-Means*

Analisis dengan		<i>K-Means</i>	
		<i>Cluster 0</i>	<i>Cluster 5</i>
<i>K-Medoids</i>	<i>Cluster 0</i>	242	334
	<i>Cluster 5</i>	30	148

Informasi-informasi yang dapat diperoleh dari nilai-nilai di Tabel 1. meliputi:

$$\begin{aligned} \text{True Positive} &= 148 / (30 + 148) * 100\% \\ &= 148 / 178 * 100\% \\ &= 83.14607\% \end{aligned}$$

$$\begin{aligned} \text{False Positive} &= 334 / (242 + 334) * 100\% \\ &= 334 / 576 * 100\% \\ &= 57.98611111\% \end{aligned}$$

$$\begin{aligned} \text{True Negative} &= 242 / (242 + 334) * 100\% \\ &= 242 / 576 * 100\% \\ &= 42.01388889\% \end{aligned}$$

$$\begin{aligned} \text{False Negative} &= 30 / (30 + 148) * 100\% \\ &= 30 / 178 * 100\% \\ &= 16.85393\% \end{aligned}$$

$$\begin{aligned} \text{Presisi} &= 148 / (334 + 148) * 100\% \\ &= 148 / 482 * 100\% \\ &= 30.70539419\% \end{aligned}$$

$$\begin{aligned} \text{Akurasi} &= (242 + 148) / (242 + 334 + 30 + 148) * 100\% \\ &= 390 / 754 * 100\% \\ &= 51.72413793\% \end{aligned}$$

Tabel 2. *Confusion Matrix* Pengujian Analisis *K-Medoids*

Analisis dengan		<i>K-Means</i>	
		<i>Cluster 0</i>	<i>Cluster 8</i>
<i>K-Medoids</i>	<i>Cluster 0</i>	242	334
	<i>Cluster 8</i>	41	6

Informasi-informasi yang dapat diperoleh dari nilai-nilai di Tabel 2. meliputi:

$$\begin{aligned} \text{True Positive} &= 6 / (41 + 6) * 100\% \\ &= 6 / 47 * 100\% \\ &= 12.7659574\% \end{aligned}$$

$$\begin{aligned} \text{False Positive} &= 334 / (242 + 334) * 100\% \\ &= 334 / 576 * 100\% \\ &= 57.98611111\% \end{aligned}$$

$$\begin{aligned} \text{True Negative} &= 242 / (242 + 334) * 100\% \\ &= 242 / 576 * 100\% \\ &= 42.01388889\% \end{aligned}$$

$$\begin{aligned} \text{False Negative} &= 41 / (41 + 6) * 100\% \\ &= 41 / 47 * 100\% \\ &= 87.2340426\% \end{aligned}$$

$$\begin{aligned} \text{Presisi} &= 6 / (334 + 6) * 100\% \\ &= 6 / 340 * 100\% \end{aligned}$$

$$= 1.79640719\%$$

$$\begin{aligned} \text{Akurasi} &= (242 + 6) / (242 + 334 + 41 + 6) * 100\% \\ &= 248 / 623 * 100\% \\ &= 39.80738363\% \end{aligned}$$

Dari hasil perhitungan *Confusion Matrix* diketahui bahwa algoritma *K-Means* memiliki nilai akurasi sebesar 51.72413793%, dan algoritma *K-Medoids* memiliki nilai akurasi sebesar 39.80738363%. Dari perbedaan nilai akurasi ini dapat disimpulkan bahwa dengan menggunakan 1411 data dengan 7 *dataset*, algoritma *K-Means* memiliki performansi lebih baik jika dibandingkan dengan algoritma *K-Medoids* untuk permasalahan sarana promosi penerimaan mahasiswa baru di Politeknik TEDC Bandung yang dibahas pada penelitian ini.

V. KESIMPULAN DAN SARAN

Kesimpulan

Dari Penelitian ini dapat disimpulkan bahwa perbedaan nilai akurasi dengan menggunakan 1411 data dengan 7 *dataset*, algoritma *K-Means* memiliki performansi lebih baik jika dibandingkan dengan algoritma *K-Medoids* untuk permasalahan sarana promosi penerimaan mahasiswa baru di Politeknik TEDC Bandung yang dibahas pada penelitian ini.

Saran

Hasil *clustering* data penerimaan mahasiswa baru tahun akademik 2015 sampai dengan tahun akademik 2017 menggunakan sebanyak 1.411 data. Penelitian ini dapat dikembangkan lagi dengan mengangkat topik tentang sistem pendukung keputusan dengan sasaran dapat menghasilkan pertimbangan yang dapat diusulkan kepada Direktur Politeknik TEDC Bandung dan Jajarannya untuk memilih media yang tepat dalam promosi penerimaan mahasiswa baru.

DAFTAR PUSTAKA

- Astroni, Ronald Adrian. 2015. "*Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang*". Semesta Teknika. pp: 76-82.
- C. Zhang dan F. Fang, 2015. "*An Improved K-means Clustering Algorithm*". *Journal of International & Computational Science*. pp.193-199.
- J. Han, M. Kamber, & J. Pei. 2006. *Data Mining: Concept and Techniques*. Second Edition. Waltham: Morgan Kaufmann Publishers.
- Kusrini dan Lutfi, E.T. 2009. "*Algoritma Data Mining*". Yogyakarta: Andi Offset.
- Lamb, Hair, dan McDaniel. 2001. "*Pemasaran*". Buku1. Penerjemah David Octarevia. Jakarta: Penerbit Salemba Empat.
- Larose, D.T. 2005. *Discovering Knowledge in Data: An Introduction to Data mining*. New Jersey: John Willey & Sons. Inc.
- Miswaningsih, N., & Insani, N. 2016. "*Analisis Perilaku Pengguna E-Learning BESMART Melalui Teknik Clustering dengan Algoritma K-Means*".

- Yogyakarta: ISBN 978-602-73403-0-5. pp: 241-246.
- Moekijat. 2000. "*Manajemen Pemasaran*". Bandung: Penerbit Mandar Maju.
- Nugraha, D. D., Naimah, Z., Fahmi, M., & Setiani, N. 2016. "*Klasterisasi Judul Buku dengan Menggunakan Metode K-Means*". Seminar Nasional Aplikasi Teknologi Informasi (SNATI). Yogyakarta: ISSN: 1907 - 5022. pp: G1-G4.
- Ong, J. O. 2013. "*Implementasi Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing President University*". Jurnal Ilmiah Teknik Industri. Bekasi: ISSN 1412-6869. pp: 11-20.
- Santosa, B. 2007. "*Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*". Yogyakarta: Graha Ilmu.