

TEKNIK CLUSTERING DENGAN ALGORITMA K-MEANS UNTUK MENANGANI STRATEGI PROMOSI DI POLITEKNIK TEDC BANDUNG

Novita Lestari Anggreini¹⁾, Indra Budi²⁾
Sistem Informasi, Universitas AMIKOM Yogyakarta¹⁾
Email: vienovie@poltektedc.ac.id¹⁾indra.ib.budi@gmail.com²⁾

Abstrak

Berdasarkan data promosi yang telah dilakukan di tahun-tahun yang lalu, dapat dilihat suatu fakta bahwa penyebaran informasi mengenai Politeknik TEDC Bandung terjadi secara tidak merata dalam kegiatan Penerimaan Mahasiswa Baru. Jangkauan pengenalan kampus ini melalui koran Tribun Jabar hanya menjangkau mahasiswa baru sebesar 10%, sedangkan dari wawancara peneliti dengan Ketua Penerimaan Mahasiswa Baru 2017 dapat memperoleh mahasiswa baru sebesar 40% melalui presentasi ke sekolah-sekolah yang ada di Cimahi, Bandung dan sekitarnya, promosi melalui para alumni dapat menjangkau sebanyak 45% mahasiswa baru, sedangkan promosi dengan media baliho dan *website* dapat menjangkau mahasiswa baru sebesar 5%. Kegiatan promosi-promosi ini mengakibatkan penyebaran jumlah mahasiswa baru di tiap-tiap program studi menjadi tidak merata. Untuk dapat melakukan promosi-promosi yang lebih efektif dan efisien, pada penelitian ini akan mengolah semua data yang telah didapatkan dari calon mahasiswa baru yang telah mendaftar, sehingga dapat mengetahui pola dari data-data tersebut dan akhirnya dapat diambil informasi-informasi yang tersembunyi dari data-data tersebut. Pengolahan data yang dilakukan adalah melalui klusterisasi dengan algoritma *K-Means* terhadap data para pendaftar yaitu tentang Nama Pendaftar, Tempat dan Tanggal lahir, Alamat, Agama, Nomor Telepon, Asal Sekolah, Program Studi Pilihan, dan Pembawa Informasi, yang mana data ini diperoleh dari kegiatan promosi tahun 2015 sampai tahun 2017. Setelah dilakukan pembersihan dan perbaikan data, kemudian dapat diperoleh suatu *dataset* yang harus dibuat lebih detail dan dapat memberikan informasi yang lebih spesifik dalam pengelompokan data. Bilamana diperlukan maka ada juga beberapa data yang dibuang agar lebih efektif dalam melakukan pengujian dalam proses klusterisasi. Hasil akhir dari klusterisasi ini diperoleh adalah informasi dalam bentuk *chart bars* yang menunjukkan bahwa cara promosi yang paling vital adalah dengan kerjasama Politeknik TEDC Bandung dengan sekolah-sekolah lanjutan atas, juga dapat diketahui bahwa para pendaftar terbanyak adalah berasal dari Jawa Barat dan Sumatera Selatan, selain itu dapat diperoleh informasi bahwa program studi Sistem Informasi paling banyak yang diminati para lulusan SMA, dan para lulusan SMK lebih memilih program studi Teknik Informatika.

Kata Kunci: klusterisasi, promosi, mahasiswa, dataset.

Abstract

This study discusses clustering with the K-Means algorithm so that the results can be proposed as material for decision making to handle the promotion strategy of the TEDC Polytechnic Bandung in the New Student Admission activities. The researcher made direct observations in the Promotion of New Student Admissions section of TEDC Polytechnic Bandung to collect the data and information needed. The researcher also conducted an interview directly in the Promotion of New Student Admissions at the TEDC Polytechnic Bandung regarding data collection on new student admissions. Interviews were also conducted directly with 2015-2017 class students regarding how easily they could get information about new student admissions at the TEDC Polytechnic in Bandung. The clustering process requires some data as input data which includes the Registrant's Name, Place of Birth Date, Address, Religion, No. Cellphone / Telephone, School of Origin, Preferred Study Program, and Information Carrier, which was obtained from promotional activities from 2015 to 2017. After cleaning and repairing, a dataset can be obtained that must be made more detailed, such as Address data can be broken down again into Districts, City and Province in order to provide more specific information in grouping data. There are also some data that are discarded in order to be more effective in carrying out tests such as Registrant's Name, Place, Date of Birth, Religion and No. Cellphone / Telephone. In this study the researchers also added attributes such as the type of school can be divided into two, namely high school and vocational school. This attribute is made so that it can be seen more clearly the type of school which is more registered with the TEDC Polytechnic in Bandung.

Keywords: clustering, promotion, student, dataset.

I. PENDAHULUAN

Politeknik TEDC Bandung merupakan perguruan tinggi di Jawa Barat selalu

berkomitmen memberikan pendidikan yang baik kepada calon mahasiswanya. Seiring dengan perkembangan teknologi dan informasi khususnya di lingkup pendidikan tinggi, maka daya saing

untuk mendapatkan mahasiswa baru juga meningkat tajam. Begitu juga yang terjadi di Politeknik TEDC Bandung di mana setiap tahun selalu melakukan promosi-promosi ke seluruh wilayah Indonesia untuk meningkatkan jumlah penerimaan mahasiswa. Dilihat dari promosi-promosi yang telah dilakukan pada tahun-tahun yang lalu, dapat dilihat suatu permasalahan yang terjadi di kegiatan Penerimaan Mahasiswa Baru adalah penyebaran informasi mengenai Politeknik TEDC Bandung terjadi secara tidak merata.

Dari data Penerimaan Mahasiswa Baru didapatkan bahwa promosi yang dilakukan melalui Koran Tribun Jabar hanya menjangkau mahasiswa baru sebesar 10%, karena kemampuan penyebaran informasinya hanya di sekitar Jawa Barat dan wilayah di luar Jawa Barat tidak terjangkau. Peneliti juga melakukan wawancara dengan Ketua Penerimaan Mahasiswa Baru 2017, dan memperoleh informasi bahwa promosi yang dilakukan melalui presentasi ke sekolah-sekolah yang ada di Cimahi, Bandung dan sekitarnya dapat memperoleh mahasiswa baru sebesar 40%. Sedangkan promosi melalui para alumni Politeknik TEDC Bandung dapat menjangkau sebanyak 45% mahasiswa baru. Sedangkan promosi dengan media baliho dan *website* dapat menjangkau mahasiswa baru sebesar 5%. Kegiatan promosi-promosi tersebut diatas mengakibatkan penyebaran jumlah mahasiswa baru di tiap-tiap program studi yang ada di Politeknik TEDC Bandung menjadi tidak merata.

Beberapa penelitian terdahulu yang memberi motivasi dan menjadi dasar dari penelitian ini antara lain:

- a. Pada penelitian Ong (2013) menyebutkan bahwa data yang telah didapatkan - pada tahap pengumpulan data - kemudian diolah untuk mengetahui pola dari data-data tersebut sehingga dapat diambil informasi-informasi yang tersembunyi dari data-data tersebut. Metode pengolahan data seperti ini sering disebut dengan istilah *Data Mining*. Penelitian Ong (2013) ini memotivasi peneliti untuk menjelaskan pengelompokan data / klusterisasi (*clustering*) data dari calon mahasiswa baru yang mendaftar ke Politeknik TEDC Bandung untuk dapat mendukung promosi penerimaan mahasiswa baru yang lebih efektif dan efisien.
- b. Di dalam penelitian Miswaningsih, dkk (2016) menuliskan bahwa tujuan dari *clustering* adalah untuk mempartisi sekumpulan data yang diberikan untuk diatur ke dalam sebuah kelompok berdasarkan ciri tertentu. Penelitian Miswaningsih, dkk (2016) ini juga memberi motivasi untuk penelitian ini yaitu memilih menggunakan *Data Mining* dengan teknik *clustering*, dengan alasan bahwa teknik *clustering* merupakan teknik pengelompokan data yang paling lama dan metode yang paling cocok digunakan terhadap data calon

mahasiswa baru pada institusi tersebut di atas.

- c. Pada penelitian Nugraha, dkk (2016) menyebutkan bahwa dengan adanya data dalam skala besar maka memungkinkan *Data Mining* dengan teknik *clustering* diharapkan dapat mengelompokkan data ke dalam beberapa kelompok yang diinginkan. Salah satu algoritma yang banyak digunakan dalam melakukan *clustering* dengan partisi ini adalah algoritma *K-Means*. Penelitian Nugraha, dkk (2016) ini memberi motivasi mengenai pemakaian algoritma *K-Means* untuk menangani proses klusterisasi atas data pendaftaran mahasiswa baru di Politeknik TEDC Bandung.
- d. Kemudian pada penelitian Asroni (2015) menuliskan bahwa *K-Means* diperlukan karena mampu mengelompokkan strategi-strategi promosi yang dapat menjadi acuan untuk menentukan prediksi penerimaan mahasiswa baru di masa yang datang. Pendapat Asroni (2015) ini memberi motivasi untuk penelitian saat ini, yaitu algoritma *K-Means* diterapkan pada proses klusterisasi data para calon mahasiswa baru, yang diharapkan dapat menghasilkan informasi-informasi yang dapat diusulkan untuk menentukan strategi promosi di masa yang akan datang.

Dari permasalahan di atas peneliti tertarik untuk melakukan penelitian yang membahas mengenai Teknik *Clustering* untuk Menangani Strategi Promosi di Politeknik TEDC Bandung.

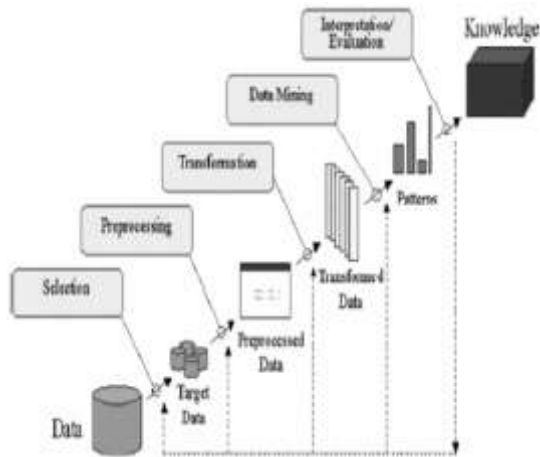
II. LANDASAN TEORI

Data Mining

Terdapat beberapa definisi untuk istilah *Data Mining*, yaitu: 1) Proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam *Data Mining* sangat bervariasi. Salah satu metode yang terdapat dalam *Data Mining* yang digunakan dalam penelitian ini adalah pengelompokan (*clustering*) yang mana metode tersebut mengidentifikasi objek yang memiliki kesamaan karakteristik tertentu, dan kemudian menggunakan karakteristik tersebut sebagai "vektor karakteristik" atau "*centroid*" (Larose, 2005). 2) Suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database*. *Data Mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terakut dari berbagai *database* besar (Kusrini, dkk, 2009). 3) Serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui

secara manual. Yang mana hasil dari proses penggalian tersebut akan membentuk pola-pola dari kumpulan data, yang sering disebut dengan pengenalan pola (*pattern recognition*). Pengenalan pola merupakan bagian dari *data mining*. *Data Mining* sering juga disebut *Knowledge Discovery in Database (KDD)*, yaitu kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar (Santoso, 2007).

Logika proses dalam KDD ditunjukkan pada Gambar 1 di bawah ini.

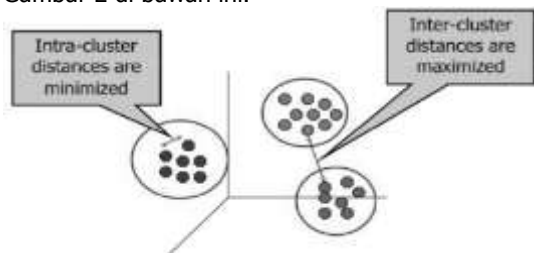


Gambar 1. Aliran Informasi dalam *Data Mining* (Han, 2006)

Data Mining terdiri dari beberapa kelompok berdasarkan tugas dan fungsinya, antara lain: 1) Deskripsi, 2) Estimasi, 3) Prediksi, 4) Klasifikasi, 5) Clustering, dan 6) Asosiasi (Larose, 2005).

Clustering

Clustering merupakan salah satu teknik *data mining* yang digunakan untuk mendapatkan kelompok-kelompok dari obyek-obyek yang mempunyai karakteristik yang umum pada data yang cukup besar. Tujuan utama dari teknik *clustering* adalah pengelompokan sejumlah data/obyek ke dalam grup (*cluster*) sehingga dalam setiap *cluster* akan berisi data yang semirip mungkin. *Clustering* melakukan pengelompokan data yang didasarkan pada kesamaan antar objek, oleh karena itu klasterisasi digolongkan sebagai metode *unsupervised learning* (Santosa, 2007). Logika dari konsep clustering ditunjukkan pada Gambar 2 di bawah ini.



Gambar 2. Konsep *Clustering* (Han, 2006)

K-Means

Algoritma *K-Means* adalah untuk menyelesaikan masalah *clustering* (Fang, 2015). *K-Means* merupakan algoritma untuk mengklasifikasi atau mengelompokkan objek berdasarkan atribut/fitur ke dalam *k* kelompok, dengan *k* adalah bilangan positif (Nugraha, 2016). Pengelompokan ini dilakukan dengan meminimalkan jumlah kuadrat dari jarak antara data dan *cluster centroid* yang bersangkutan. Dengan demikian, tujuan dari *k-means clustering* adalah untuk mengelompokkan data (Nugraha, 2016). Efisiensi keaslian algoritma *K-Means* sangat bergantung pada *centroid* awal. Langkah kerja dari algoritma *K-Means* adalah sebagai berikut: 1) Menentukan jumlah *cluster K* dan *centroid* awal, 2) Menghitung jarak masing-masing objek ke *centroid*, 3) Mengelompokkan objek berdasarkan jarak minimum objek ke *centroid*, 4) Menentukan *centroid* kembali apabila terjadi perpindahan kluster dengan menghitung nilai rata-rata dari seluruh anggota pada masing-masing kluster, 5) Ulangi langkah (1) sampai langkah (4) sampai tidak terjadi perpindahan kluster lagi (Larose, 2005).

Algoritma berakhir ketika semua *record* yang dimiliki oleh masing-masing pusat *cluster* tetap dalam *cluster* itu. Atau, algoritma dapat berhenti ketika beberapa kriteria konvergensi terpenuhi, seperti ada penyusutan yang tidak signifikan dalam jumlah kuadrat *error (sum of squared errors)* (Larose, 2005).

Strategi Promosi

Beberapa definisi mengenai Strategi Promosi, antara lain: 1) Kegiatan perusahaan untuk mendorong penjualan dengan mengarahkan komunikasi-komunikasi yang meyakinkan kepada para pembeli (Moekijat, 2000), 2) Rencana untuk penggunaan yang optimal dari elemen-elemen promosi seperti periklanan, hubungan masyarakat, penjualan pribadi dan promosi penjualan (Lamb dkk, 2001).

Dari beberapa definisi di atas dapat dipahami bahwa strategi promosi merupakan kegiatan yang direncanakan dengan maksud membujuk, merangsang konsumen agar mau membeli produk perusahaan sehingga tujuan untuk meningkatkan penjualan diharapkan dapat tercapai (Lamb dkk, 2001).

III. METODE PENELITIAN

Penelitian ini bersifat deskriptif, yaitu mendeskripsikan kegiatan penelitian dari tahap awal, melakukan analisis dengan suatu tool untuk mendukung pengolahan klasterisasi, sampai membuat suatu kesimpulan berdasarkan hasil keluaran berupa *chart bars*.

Pendekatan penelitian ini bersifat kualitatif, yaitu memahami cara kerja algoritma *K-Means*, kemudian algoritma itu dipakai untuk pengolahan klasterisasi data sampai menghasilkan keluaran

berupa informasi-informasi dalam bentuk *chart bars* yang menunjukkan nilai-nilai dari media promosi, program studi yang diminati sesuai asal sekolah pendaftar, dan domisili asal pendaftar untuk mendapatkan gambaran yang jelas mengenai strategi promosi yang paling efektif dan sesuai untuk meningkatkan Penerimaan Mahasiswa Baru di Politeknik TEDC Bandung.

Pengumpulan data dalam penelitian ini dilakukan dengan cara:

a. Observasi

Merupakan pengamatan langsung di bagian Promosi Penerimaan Mahasiswa Baru Politeknik TEDC Bandung serta melakukan pencatatan secara cermat dan sistematis guna mengumpulkan data pendaftaran mahasiswa tahun 2015 sampai tahun 2017. Peneliti juga mengumpulkan informasi ketercapaian promosi penerimaan mahasiswa untuk rentang waktu tiga tahun tersebut di atas.

b. Wawancara

Peneliti melakukan wawancara langsung di bagian Promosi Penerimaan Mahasiswa Baru Politeknik TEDC Bandung terkait pendataan penerimaan mahasiswa baru. Wawancara juga dilakukan langsung dengan mahasiswa angkatan 2015-2017 mengenai seberapa mudah mereka mendapatkan informasi tentang penerimaan mahasiswa baru di Politeknik TEDC Bandung.

c. Studi Literatur

Peneliti melakukan studi literatur dengan mempelajari buku-buku, jurnal dan paper dengan pembahasan yang dapat mendukung pelaksanaan penelitian ini.

Data yang akan dianalisis disini adalah data mahasiswa baru dari bagian Promosi Penerimaan Mahasiswa Baru Politeknik TEDC Bandung yang selanjutnya data-data tersebut disesuaikan dengan kebutuhan penelitian sehingga menghasilkan atribut yang dibutuhkan seperti berikut ini : **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 tahap pengumpulan data dari hasil observasi di bagian Promosi Penerimaan Mahasiswa Baru Politeknik TEDC Bandung, diperoleh data mahasiswa tahun 2015-2017 sebanyak 1.411 orang. Data yang diperoleh terdiri dari Nama Pendaftar, Tempat Tanggal lahir, Alamat, Agama, No. HP/Telepon, Asal Sekolah, Program Studi Pilihan, dan Pembawa Informasi.

Setelah dilakukan pembersihan dan perbaikan, maka diperoleh *dataset* yang harus dibuat lebih detail, seperti data Alamat dapat diperinci menjadi Kecamatan, Kota dan Propinsi agar dapat memberikan atribut spesifik dalam

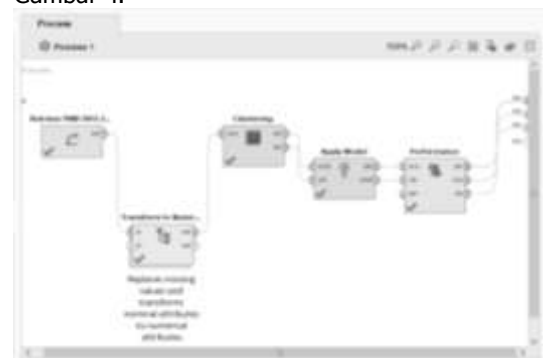
pengelompokan data. Ada juga beberapa data yang dibuang agar lebih efektif dalam melakukan pengujian seperti Nama Pendaftar, Tempat, Tanggal lahir, Agama dan No. HP/Telepon. Peneliti perlu menambahkan atribut Jenis Sekolah yang dibedakan menjadi SMA dan SMK. Atribut ini dibuat agar dapat terlihat lebih jelas jenis sekolah yang mana yang lebih banyak mendaftarkan ke Politeknik TEDC Bandung.

Gambar 3 menunjukkan cara pengaturan *dataset* dalam melakukan impor ke dalam *tool* RapidMiner.



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

Peneliti membuat desain *clustering* dengan algoritma *K-Means* seperti yang ditunjukkan pada Gambar 4.



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

Pada Gambar 4 dapat dilihat bahwa pengujian dimulai dengan *retrieve* 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*. Setelah dari Algoritma *Clustering K-Means* kemudian disambungkan lagi ke *Apply Model*, lalu disambungkan kembali ke *Performance*, dan terakhir disambungkan ke *Result*.

Setelah dilakukan pengujian, diperoleh informasi bahwa lama waktu pemrosesan *clustering* dengan Algoritma *K-Means* dilakukan dalam 35 detik, seperti yang tercatat di dalam *Result History* di Gambar 5.



Gambar 5. Result History Clustering menggunakan Algoritma K-Means

Dapat dilihat juga *Davies Bouldin Index (DBI)* yang ada di dalam *Performance* pada gambar 6, dengan semakin rendah nilai DBI, maka *cluster* tersebut semakin baik.



Gambar 6. Nilai DBI

Gambar 7 menunjukkan deskripsi *Performance* yang menghasilkan *Centroid Distance* pada setiap *cluster*.



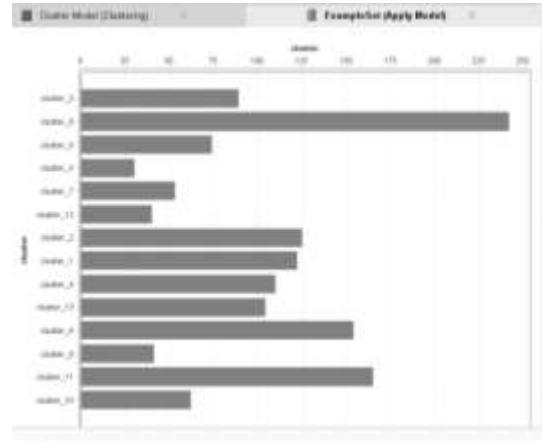
Gambar 7. Centroid Distance K-Means

Pada penelitian ini *cluster* yang digunakan berjumlah 14, dimulai dari *cluster* 0 sampai *cluster* 13, total *dataset* yang di-*clustering* sebanyak 1.411, dengan jumlah *items* pada tiap-tiap *cluster* diinformasikan pada Gambar 8.



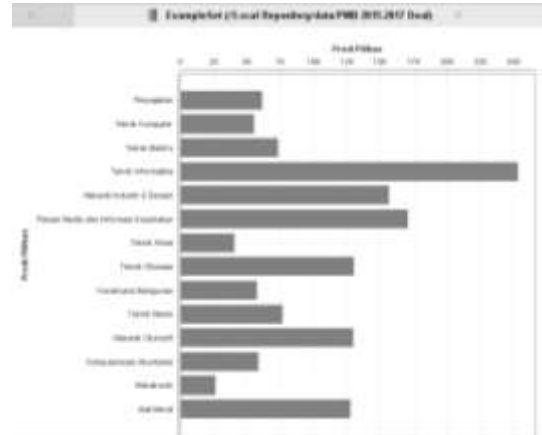
Gambar 8. Hasil Cluster Model K-Means

Hasil dari *cluster model* dapat digambarkan dalam bentuk *Charts Bars* seperti pada Gambar 9, yang menunjukkan *cluster* Kerjasama Perguruan Tinggi memiliki nilai paling tinggi, dan *cluster* Orang Tua memiliki nilai paling rendah dalam *set* data Pembawa Informasi.



Gambar 9. Hasil Cluster Model K-Means untuk Data Set Pembawa Informasi

Charts Bars pada Gambar 10 menunjukkan *cluster* Teknik Informatika memiliki nilai paling tinggi, dan *cluster* Mekatronik memiliki nilai paling rendah dalam *set* data Prodi Pilihan.



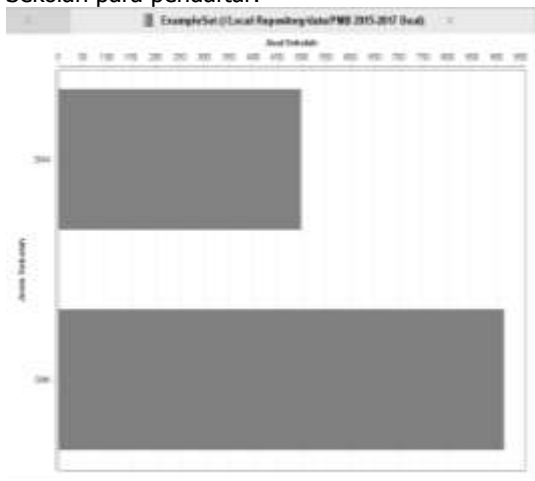
Gambar 10. Hasil Cluster Model K-Means untuk Data Set Prodi Pilihan

Charts Bars pada Gambar 11 menunjukkan *cluster* Jawa Barat memiliki nilai dengan posisi paling tinggi, dan *cluster* Sumatera Selatan memiliki posisi nilai kedua dalam urutan *descending* untuk *set* data Provinsi asal para pendaftar.



Gambar 11. Hasil Cluster Model K-Means untuk Data Set Provinsi

Charts Bars pada Gambar 12 menunjukkan *cluster* SMK memiliki nilai dengan posisi paling tinggi, dan *cluster* SMA memiliki posisi nilai kedua dalam urutan *descending* untuk *set* data Asal Sekolah para pendaftar.



Gambar 12. Hasil *Cluster* Model *K-Means* untuk Data *Set* Asal Sekolah

Dengan dilakukan analisis terhadap data para pendaftar dengan pemrosesan *cluster* data dapat diketahui bahwa dengan penyebaran informasi melalui kerjasama perguruan tinggi diperoleh para pendaftar dengan asal sekolah SMK yang berasal dari provinsi Jawa Barat merupakan para pendaftar dengan jumlah terbanyak, dan para pendaftar dari kelompok SMK lebih meminati program studi Teknik Informatika.

V. KESIMPULAN DAN SARAN

Kesimpulan

Algoritma *K-Means* untuk *clustering* data dapat diterapkan untuk menghasilkan informasi-informasi berupa beberapa *dataset* sebagai bahan pertimbangan yang dapat diusulkan kepada pihak manajemen Politeknik TEDC Bandung untuk mendukung kegiatan promosi penerimaan mahasiswa baru dapat dilakukan dengan lebih efisien dan lebih efektif, yaitu dengan cara penyebaran informasi pendaftaran/penerimaan mahasiswa baru dilakukan dengan kerjasama pihak Politeknik TEDC Bandung dengan sekolah-sekolah SMK di wilayah provinsi Jawa Barat dan Sumatera Selatan.

Saran

Hasil klasterisasi data penerimaan mahasiswa baru untuk tahun 2015-2017 yang dipakai pada penelitian ini memiliki data sebanyak 1.411 data pendaftar. Dari data ini dalam penelitian saat ini menghasilkan empat buah *dataset*, yaitu berdasarkan data pembawa informasi

pendaftaran/penerimaan mahasiswa baru, program studi pilihan yang banyak diminati para pendaftar, provinsi asal para pendaftar, dan sekolah asal para pendaftar. Hasil klasterisasi ini dapat dikembangkan di dalam penelitian yang akan datang dengan mengangkat topik bahasan tentang sistem pendukung keputusan dengan sasaran dapat menghasilkan bahan pertimbangan yang dapat diusulkan kepada pihak manajemen Politeknik TEDC Bandung untuk dapat memilih dengan tepat apa media pembawa informasi yang efektif dan di provinsi mana yang tepat untuk dilakukan kegiatan promosi pendaftaran mahasiswa baru pada tahun mendatang.

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.