

PENERAPAN ALGORITMA K-MEANS CLUSTERING SEBAGAI STRATEGI PROMOSI PENERIMAAN MAHASISWA BARU PADA UNIVERSITAS HASYIM ASY'ARI JOMBANG

Imam Mahmudi

Program Studi S1 Teknik Informatika, Fakultas Teknologi Informasi, Universitas Hasyim Asy'ari
Email : imam.pbb@gmail.com

Aries Dwi Indriyanti

Program Studi S1 Teknik Informatika, Fakultas Teknologi Informasi, Universitas Hasyim Asy'ari
Email : ariesindriyanti@unhasy.ac.id

Indana Lazulfa

Program Studi S1 Teknik Informatika, Fakultas Teknologi Informasi, Universitas Hasyim Asy'ari
Email : indanalazulfa@unhasy.ac.id

Abstrak

Penerimaan mahasiswa baru di Universitas Hasyim Asy'ari Jombang dilaksanakan pada setiap tahun. Untuk mendapatkan mahasiswa baru yang lebih banyak, panitia penerimaan mahasiswa baru melakukan beberapa promosi sebagai kegiatan awal yang sangat penting seperti: *online*, spanduk, brosur, *event-event* sekolah, maupun dengan lisan dengan peran mahasiswa dan alumni. Banyaknya persaingan dalam mencari pendaftar mahasiswa baru, mengharuskan pihak Universitas Hasyim Asy'ari untuk melakukan analisis dari beberapa cara promosi yang sudah dilakukan sehingga dapat dilihat strategi promosi mana yang lebih tepat dan efektif. Penelitian ini akan melakukan pengelompokan/*clustering* kabupaten atau kota berdasarkan atribut tertentu dalam sebuah aplikasi berbasis web. Metode yang digunakan dalam penelitian ini adalah *algoritma k-means clustering* yang dapat mengelompokkan data mahasiswa ke dalam beberapa *cluster* berdasarkan kemiripan atribut yang sama. Atribut yang digunakan adalah kota asal, *online*, lisan, spanduk/baliho, brosur dan *event*. Pada penelitian ini menghasilkan jumlah 5 *cluster* ($k=5$) dengan *cluster* pertama 20 kota asal dengan media promosi paling efektif *online* dan *lisan*, *cluster* kedua 31 kota asal dengan media promosi paling efektif lisan dan *online*, *cluster* ketiga 4 kota asal dengan media promosi paling efektif *event* dan spanduk, *cluster* keempat 15 kota asal dengan media promosi paling efektif brosur dan lisan, *cluster* kelima 2 kota asal dengan media promosi paling efektif lisan dan *event*. Hasil dari penelitian ini digunakan sebagai rekomendasi untuk menentukan strategi promosi berdasarkan media promosi masing-masing *cluster* yang terbentuk.

Kata Kunci: Strategi promosi, Penerimaan Mahasiswa Baru, *Algoritma K-means*, *Clustering*

Abstract

Admission of new students at the Hasyim Asy'ari University in Jombang is held every year. To more of the new students, admissions Committee conducted several promotions as very important early activities such as: online, banners, brochures, school events, and orally with student roles and Alumni. The number of competition in finding new student applicants, requiring the University of Hasyim Asy'ari to conduct analysis of several ways of promotion that have been done so that the promotion strategy can be seen which is more precise and effective. This research will conduct grouping/clustering of districts or cities based on certain attributes in a Web-based application. The method used in this study is a K-means clustering algorithm that can group student data into multiple clusters based on similar attribute agreements. The attributes used are hometown, online, oral, banners/billboards, brochures and events. At this Peletitian generate a total of 5 clusters ($k = 5$) with the first cluster 20 hometown with the most effective promotional media online and oral, the second cluster of 31 Origin cities with the most effective promotional media oral and online, the third cluster of 4 cities originating with media The most effective promotion of events and banners, the fourth cluster of 15 hometown with the most effective promotional media brochures and oral, the fifth cluster 2 hometown with the most effective promotional media oral and event. The results of this study were used as a recommendation to determine a promotional strategy based on the promotional media of each cluster formed.

Keywords: Promotion strategy, Admission of New Students, K-means Algorithm, Clustering

PENDAHULUAN

Penerimaan Mahasiswa Baru (PMB) merupakan suatu agenda yang harus dilakukan di berbagai universitas, seperti halnya pada Universitas Hasyim Asy'ari. Dari banyaknya kampus yang ikut bersaing, berbagai strategi promosi selalu menjadi hal utama dalam menentukan seberapa banyak mahasiswa baru yang didapatkan. Sehingga analisa strategi promosi Universitas Hasyim Asy'ari yang tepat sasaran dan akurat dalam mendapatkan mahasiswa baru sangat dibutuhkan. Dengan diadakannya kegiatan penerimaan mahasiswa baru di Universitas Hasyim Asy'ari tiap tahunnya, hal ini menimbulkan penumpukan data yang besar dalam *database*, hal ini perlu dilakukan analisis data guna mendapatkan informasi-informasi yang bermanfaat dalam membantu menentukan strategi promosi penerimaan mahasiswa baru pada tahun berikutnya. Dalam melakukan promosi yang lebih menasar dan lebih efisien, maka dalam penelitian ini, peneliti melakukan pengolahan data mahasiswa terdahulu guna untuk mengetahui pola data mahasiswa tersebut, sehingga peneliti akan mendapatkan informasi-informasi yang tersembunyi dalam data mahasiswa tersebut misalnya informasi tentang pengelompokan data mahasiswa yang berpotensi berdasarkan asal kota atau wilayahnya.

Pada penelitian ini, akan dilakukan analisa data menggunakan metode *clustering* di *data mining* dengan menggunakan *Algoritma K-Means*. *Algoritma K-Means* sangat populer digunakan oleh para peneliti sebelumnya untuk mendapatkan berbagai informasi, yang digunakan untuk menentukan strategi promosi dengan cara mengelompokkan data.

Clustering merupakan metode data mining yang berfungsi untuk membagi atau mengelompokkan objek ke dalam *cluster* tertentu dengan tingkat kemiripan yang tinggi. Objek akan dikelompokkan ke dalam satu atau lebih *cluster* sehingga objek-objek yang berada dalam satu *cluster* akan mempunyai kesamaan yang tinggi antara satu dengan yang lainnya. Objek-objek dikelompokkan berdasarkan prinsip memaksimalkan kesamaan objek pada *cluster* yang sama dan memaksimalkan ketidaksamaan pada *cluster* yang berbeda. Kesamaan objek biasanya diperoleh dari nilai-nilai atribut yang menjelaskan objek data (Suprawoto, 2016:13).

Algoritma K-means merupakan salah satu metode *clustering non hierarki* yang berfungsi untuk mempartisi atau mengelompokkan data kedalam sebuah *cluster* yang memiliki karakteristik yang sama dan karakteristik berbeda akan dikelompokkan kedalam *cluster* lain (Darmi dan Setiawan, 2016:148).

Penelitian Wirta Agustin dan Erlin dalam penelitiannya tentang penerapan metode *K-means* dalam menentukan media promosi penerimaan mahasiswa baru di STMIK Amik Riau. Atribut yang digunakan adalah media *online*, spanduk, brosur atau *map* dan secara lisan yang di *cluster* menjadi 3 *cluster* yaitu sangat efektif, efektif, dan kurang efektif. Pada penerimaan mahasiswa baru pada tahun 2015 sampai 2017 STMIK Amik Riau diperoleh hasil bahwa promosi dari lisan lebih dominan dibandingkan *online*, brosur dan spanduk (Agustin dan Erlin, 2016:10). Hasil dari penelitian tersebut untuk menentukan strategi promosi yang paling efektif.

Pada bidang iklim, penelitian tentang pengelompokan data menggunakan metode *K-means* bisa diterapkan untuk mengelompokkan curah hujan pada Provinsi Kalimantan Timur. Dimana analisis dan pengelompokan dilakukan berdasarkan data dari data curah hujan tahunan pada 13 stasiun pengamatan di Kaltim. Hasil penelitian tersebut terbentuk 3 *cluster* untuk pengelompokan curah hujan tinggi, curah hujan sedang, dan curah hujan rendah (Puspitasari dan Haviluddin, 2016:20).

Penelitian ini menggunakan data asli atau data primer yaitu data jumlah mahasiswa Unhasy tiap kabupaten/kota berdasarkan media publikasi yang diketahui seperti: brosur, *online*, acara atau *event* sekolah dan lisan. Dengan menggunakan metode *K-Means* mengelompokkan data mahasiswa ke dalam beberapa *cluster* berdasarkan kemiripan atribut yang sama. Dari hasil proses *K-Means* tersebut akan didapatkan 5 klaster atau kelompok dengan kemiripan yang tinggi. Setiap klaster atau kelompok akan diinterpretasikan strategi promosi atau publikasi yang tepat untuk setiap kelompok tersebut.

Dengan adanya salah satu teknik analisis data dengan metode *k-means* akan memberikan kelompok-kelompok data yang diharapkan bisa memberikan rekomendasi bagi Universitas Hasyim Asy'ari dalam menentukan strategi promosi pendaftaran mahasiswa baru yang lebih efektif di setiap wilayah atau kota asal mahasiswa.

METODE

Penelitian ini mengambil objek dari data mahasiswa Universitas Hasyim Asy'ari dari 5 Fakultas mulai angkatan 2016 sampai dengan angkatan 2018. Peneliti mengambil objek data kota asal mahasiswa dari semua jurusan beserta media promosi yang mereka ketahui sebagai atribut penelitian. Dalam teknik pengumpulan data pada penelitian ini yaitu wawancara, penyebaran angket atau kuesioner serta dokumentasi.

Metode K-Means

Algoritma yang paling populer saat ini adalah *K-means*. *K-means* sendiri merupakan salah satu *algoritma clustering* berbasis jarak yang membagi data ke dalam sejumlah *cluster* dengan tingkat kemiripan atribut yang sama dan algoritma ini hanya bekerja pada atribut *numeric*.

Algoritma *K-means* akan menghasilkan kelompok catatan sebanyak jumlah *cluster* yang dibentuk. Algoritma *K-means* pertama kali ditemukan oleh J. MacQueen (Susanto, 2010). Berikut alur proses atau *flowchart* Algoritma *K-means Clustering*:



Gambar 1. Flowchart Algoritma K-Means

Berikut penjelasan tentang *flowchart K-means* diatas:

- a. Menentukan jumlah *cluster* yang akan dibentuk.
- b. Inisialisasi k pusat *cluster* dan memberi nilai awal dengan angka-angka acak/random.
- c. Menghitung jarak data dengan pusat *cluster*. Untuk menghitung jarak semua data dapat menggunakan rumus berikut:

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

Dimana:

- $D(i, j)$ = Jarak objek i ke pusat *cluster* j
- X_{ki} = Data ke i variabel data ke k
- X_{kj} = Titik pusat ke j pada variabel ke k

- d. Kelompokkan data ke dalam *cluster* berdasarkan jarak yang paling dekat.
- e. Mencari pusat *cluster* baru (C_k) dengan cara menghitung rata-rata dari data objek dalam *cluster* yang terbentuk.

$$C_k = \left(\frac{1}{n_k}\right) \sum d_i \tag{2}$$

Dimana n_k adalah jumlah dokumen dalam *cluster* k, d_i adalah dokumen dalam *cluster* k.

- f. Kembali ke langkah c sampai dengan e sampai tidak ada lagi objek yang berpindah ke *cluster* yang lain.

Analisa Performa K-means

Pada metode *K-Means*, pengelompokan data ke dalam sejumlah *cluster* dilakukan berdasarkan tingkat kemiripan data atribut. Kemudian, tingkat kemiripan data kelompok tersebut dapat diukur dengan beberapa metode pengukuran jarak. Pada penelitian ini, untuk mengetahui tingkat kemiripan data atribut peneliti menggunakan metode pengukuran jarak *Sum of Squared Error* (SSE). Dalam metode ini, SSE akan memberikan informasi tingkat *error* jarak data dengan *centroid*. Apabila hasil SSE semakin kecil, maka hasil *clustering* akan semakin akurat. Berikut rumus untuk menghitung SSE:

$$SSE = \sum_{k=1}^k \sum_{p \in C_i} d(p, m_i)^2 \tag{3}$$

Dimana:

k = jumlah *cluster*

$p \in C_i$ = setiap data poin pada *cluster* i

d = jarak terdekat pada masing-masing *cluster* i .

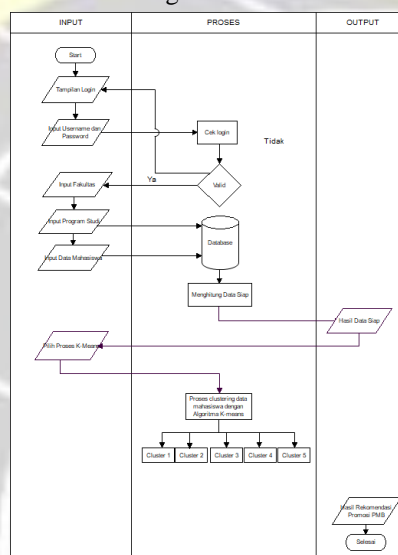
m_i = *centroid* dari *cluster* i

Perancangan Sistem

Perancangan sistem meliputi *flowchart* sistem, *Data Flow Diagram* (DFD). Tujuan dilakukan perancangan untuk mempermudah dalam membuat Aplikasi Penerapan *Algoritma K-Means Clustering* Sebagai Strategi Promosi Penerimaan Mahasiswa Baru Pada Universitas Hasyim Asy'ari Jombang. Berikut rancangan dari sistem.

Flowchart Sistem

Sebagai pendukung dalam membangun sistem rekomendasi promosi penerimaan mahasiswa baru di universitas hasyim asy'ari dengan metode *K-Means*, *flowchart* sistem ini menjelaskan tentang proses dari awal sistem berjalan sampai selesai. Maka dibentuk suatu rancangan *flowchart* sistem sebagai berikut:

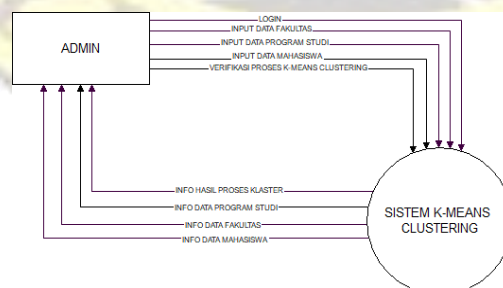


Gambar 2. *Flowchart* Sistem

Rancangan *Data Flow Diagram* (DFD)

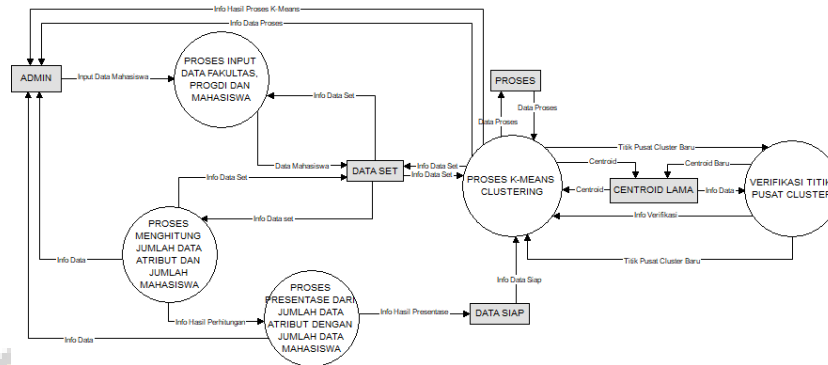
Perancangan sistem menggunakan pemodelan dengan *Data Flow Diagram* (DFD) memiliki beberapa tahapan. Tahap pertama adalah penjelasan secara umum mengenai cara kerja sistem dan dilanjutkan dengan penjelasan yang lebih spesifik dari tahap sebelumnya.

DFD level 0 atau bisa disebut *diagram konteks* merupakan DFD level tertinggi yang menggambarkan keseluruhan dari ruang lingkup suatu sistem dan menggambarkan seluruh input ke dalam sistem dan output yang dihasilkan oleh sistem.



Gambar 3. DFD Level 0 (*Diagram Konteks*)

DFD level 1 merupakan alur sistem lanjutan dari proses sebelumnya yaitu DFD level 0. DFD level 1 menggambarkan aliran data secara kompleks setiap proses sistem yang kemudian membentuk *data store* dari aliran data secara menyeluruh. Rancangan DFD level 1 dapat dilihat pada gambar berikut:



Gambar 4. DFD Level 1

HASIL DAN PEMBAHASAN

Dalam penelitian ini, data mahasiswa yang diolah sebanyak 519 data mahasiswa yang tersebar dalam 72 kota (lingkup pulau jawa) maupun provinsi (lingkup luar pulau jawa) dari 5 fakultas mulai tahun angkatan 2016 sampai 2018 pada Universitas Hasyim Asy’ari. Dalam penelitian ini sebelum dilakukannya proses *Clustering* menggunakan *Algoritma K-Means*, ada beberapa tahap yang harus dilakukan yaitu:

Pre-processing data

Pre-processing data merupakan tahapan atau proses tranformasi data, data mining serta tahap interpretasi dan evaluasi yang diharapkan dapat menghasilkan output yang lebih baik. Pada proses ini, sistem akan menghasilkan data yang sudah siap untuk dilakukannya proses *clustering* dengan *algoritma k-means*. Berikut beberapa tahap *pre-processing data*: Pada *pre-processing data* tahap pertama yaitu perhitungan data awal merupakan proses menjumlahkan data atribut atau media promosi seperti *online*, lisan, spanduk, brosur dan *event* dan menjumlahkan data mahasiswa berdasarkan objek penelitian yaitu per kota atau wilayah. Berikut hasil perhitungan data awal dapat dilihat pada Tabel 1:

Tabel 1. Hasil Perhitungan Data Awal

Kota	Online	Lisan	Spanduk	Brosur	Event	Jmh MHS
Bondowoso	1	0	0	0	0	1
Aceh	0	0	0	1	0	1
Balikpapan	0	2	0	0	0	2
Bandung	1	1	0	0	0	1
Bangka Belitung	2	1	0	0	0	2
Banjarmegara	0	0	0	1	0	1
Banten	0	2	0	0	0	2
Banyuwangi	2	3	1	1	0	4
Batam	1	1	0	0	0	1
Batang	1	2	0	0	0	2

Perhitungan data siap adalah tahap kedua dalam *pre-processing* data yaitu mencari *presentase* data dari jumlah atribut dengan jumlah mahasiswa. Berikut sempel hasil perhitungan data siap dapat dilihat pada Tabel 2:

Tabel 2. Hasil Perhitungan Data Siap

Kota	Online	Lisan	Spanduk	Brosur	Event
Bondowoso	1	0	0	0	0
Aceh	0	0	0	1	0
Balikpapan	0	1	0	0	0
Bandung	1	1	0	0	0
Bangka Belitung	1	0,5	0	0	0
Banjarnegara	0	0	0	1	0
Banten	0	1	0	0	0
Banyuwangi	0,5	0,75	0,25	0,25	0
Batam	1	1	0	0	0
Batang	0,5	1	0	0	0

Proses *Algoritma K-Means*

Setelah didapatkan hasil dari tahap *pre-processing* data, maka proses selanjutnya yaitu menganalisis data siap dengan metode *clustering* menggunakan *algoritma k-means*. Sehingga didapatkan hasil pengelompokan kota dengan strategi promosi apa yang paling efektif. Berikut beberapa tahapan proses *algoritma k-means clustering*:

Pertama menentukan jumlah *cluster*, dalam hal ini peneliti membuat lima *cluster* ($k=5$) yang akan dibentuk, yaitu *cluster 1*, *cluster 2*, *cluster 3*, *cluster 4*, *cluster 5*. Dimana hasil data dari *cluster-cluster* tersebut akan mempresentasikan objek atau wilayah dengan tingkat keefektifan media promosi yang berbeda-beda.

Kedua menentukan nilai *centroid* atau pusat *cluster*, dalam melakukan proses *k-means clustering* ini *centroid* atau pusat *cluster* awal ditentukan secara acak atau *random*. Berikut adalah nilai *centroid* awal dapat dilihat pada Tabel 3:

Tabel 3. Tabel data *centroid* awal

Jumlah K	Online	Lisan	Spanduk	Brosur	Event
<i>Cluster 1</i>	0.8	0.4	0	0	0
<i>Cluster 2</i>	0.3	0.5	0.08	0.1	0.08
<i>Cluster 3</i>	0	0.4	0.8	0.2	0.2
<i>Cluster 4</i>	0.2	0.2	0.2	0.7	0.3
<i>Cluster 5</i>	0.2	0.2	0.2	0.3	0.7

Ketiga menghitung jarak *centroid* ke i dengan data ke j dan mengalokasi objek ke *centroid* terdekat menggunakan teori jarak yang diciptakan dan dipopulerkan oleh *Euclidean*, berikut formula *K-means* yang ditunjukkan oleh persamaan (1):

Perhitungan jarak *centroid 1* dan data 1 sebagai berikut:

$$D(C1,1) = \sqrt{(1 - 0,8)^2 + (0 - 0,4)^2 + \dots (0 - 0)^2} = 0.4472$$

$$D(C2,1) = \sqrt{(1 - 0,3)^2 + (0 - 0,5)^2 + \dots (0 - 0,08)^2} = 0.8734$$

$$D(C3,1) = \sqrt{(1 - 0)^2 + (0 - 0,4)^2 + \dots (0 - 0,2)^2} = 1.3711$$

$$D(C4,1) = \sqrt{(1 - 0,2)^2 + (0 - 0,2)^2 + \dots (0 - 0,3)^2} = 1.1402$$

$$D(C5,1) = \sqrt{(1 - 0,2)^2 + (0 - 0,2)^2 + \dots (0 - 0,7)^2} = 1.1402$$

Hasil dari perhitungan jarak menggunakan rumus di atas dapat dilihat pada Tabel 4:

Tabel 4. Hasil iterasi pertama beberapa kab/kota

Kota Asal	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster Terdekat
Bondowoso	0.4472	0.8734	1.3711	1.1402	1.1402	Cluster 1
Aceh	1.3416	1.0783	1.2166	0.5477	1.0488	Cluster 4
Balikpapan	1.0000	0.6023	1.0392	1.1402	11.402	Cluster 2
Bandung	0.6325	0.8734	1.4422	1.3784	1.3784	Cluster 1
Bangka Belitung	0.2236	0.7161	1.3153	1.1619	1.1619	Cluster 1
Banjarnegara	1.3416	1.0783	1.2166	0.5477	1.0488	Cluster 4
Banten	1.0000	0.6023	1.0392	1.1402	1.1402	Cluster 2
Banyuwangi	0.5809	0.4004	0.8471	0.8292	0.9421	Cluster 2
Batam	0.6325	0.8734	1.4422	1.3784	1.3784	Cluster 1
Batang	0.6708	0.5593	1.1533	1.1619	1.1619	Cluster 2

Keempat menghitung pusat cluster baru, *centroid* atau pusat cluster baru bisa didapatkan dengan mencari rata-rata dari semua data/objek dalam cluster tertentu, selanjutnya ulangi langkah 3 dengan menghitung jarak dan mengalokasikan ke cluster terdekat sampai pusat cluster tidak berubah. Pada penelitian ini perulangan *algorithm k-means clustering* dengan menggunakan jumlah cluster 5 (k=5) sebanyak 5 kali perulangan sampai titik pusat cluster atau *centroid* tidak berubah. Berikut hasil dari *centroid* terakhir dapat dilihat pada Tabel 5:

Tabel 5. *centroid* terakhir

Jumlah K	Online	Lisan	Spanduk	Brosur	Event
Cluster 1	0.919166	0.515	0	0.108333	0.008333
Cluster 2	0.195007	0.839256	0.040242	0.117900	0.078641
Cluster 3	0	0.375	0.625	0	0.75
Cluster 4	0.185374	0.328072	0.098557	0.624037	0.243180
Cluster 5	0	1	0	0	1

Berikut hasil perhitungan jarak dan pengalokasian data ke dalam cluster terdekat dengan *centroid* terakhir dapat dilihat pada Tabel 6:

Tabel 6. Hasil Perhitungan jarak dengan *Centroid* Terakhir

Kota Asal	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster Terdekat
Bondowoso	0.5325	1.1722	1.4470	1.1088	1.7321	Cluster 1
Aceh	1.3803	1.2362	1.4470	0.5935	1.7321	Cluster 4
Balikpapan	1.0449	0.2925	1.1592	0.9717	1.0000	Cluster 2
Bandung	0.5036	0.8340	1.5309	1.2543	1.4142	Cluster 1
Bangka Belitung	0.1363	0.8859	1.4031	1.0731	1.5000	Cluster 1
Banjarnegara	1.3803	1.2362	1.4470	0.5935	1.7321	Cluster 4
Banten	1.0449	0.2925	1.1592	0.9717	1.0000	Cluster 2
Banyuwangi	0.5600	0.4106	1.0753	0.7064	1.1990	Cluster 2
Batam	0.5036	0.8340	1.5309	1.2543	1.4142	Cluster 1
Batang	0.6502	0.3749	1.2624	1.0044	1.1180	Cluster 2

- Cluster 1 memiliki nilai *centroid* (0.919166, 0.515, 0, 0.108333, 0.008333), terdiri dari 20 kota asal yang dapat diartikan *online* dan lisan menjadi media promosi yang paling efektif.
- Cluster 2 memiliki nilai *centroid* (0.195007, 0.839256, 0.040242, 0.117900, 0.078641), terdiri dari 31 kota asal yang dapat diartikan lisan dan *online* menjadi media promosi yang paling efektif di kota tersebut.
- Cluster 3 memiliki nilai *centroid* (0, 0.375, 0.625, 0, 0.75), terdiri dari 4 kota asal yang dapat diartikan *event* dan spanduk menjadi media promosi yang paling efektif di kota tersebut.
- Cluster 4 memiliki nilai *centroid* (0.185374, 0.328072, 0.098557, 0.624037, 0.243180), terdiri dari 15 kota asal yang dapat diartikan brosur dan lisan menjadi media promosi yang paling efektif di kota tersebut.
- Cluster 5 memiliki nilai *centroid* (0, 1, 0, 0, 1), terdiri dari 2 kota yang dapat diartikan lisan dan *event* menjadi media promosi yang paling efektif di kota tersebut.

Analisa Performa *K-Means*

Tahap akhir adalah pengujian menggunakan metode *Sum of Squared Error* (SSE) yang memberikan informasi *error* jarak data ke *centroid*. Pada hasil pengujian menggunakan metode SSE dilakukan pengujian nilai $k = 3$ sampai $k = 5$ agar terlihat jelas hasil untuk analisis jumlah *cluster* terbaik. Setelah dilakukan proses percobaan *clustering* pada dataset dengan jumlah *cluster* berbeda, berikut nilai SSE yang dihasilkan dapat dilihat pada tabel berikut ini:

Tabel 7. Hasil *Sum Squared Error*

Jumlah Cluster	Hasil SSE
K = 3	37,28435263
K = 4	30,37181504
K = 5	27,80007154

PENUTUP

Simpulan

Berdasarkan penelitian dan pembahasan yang telah dilakukan, maka dapat diambil kesimpulan bahwa metode *Clustering* menggunakan *Algoritma K-means* dapat membantu memilih strategi promosi penerimaan mahasiswa baru pada Universitas Hasyim Asy'ari yang paling efektif pada setiap kota atau wilayah yang ada. Berdasarkan analisis penentuan jumlah *cluster* terbaik dengan metode SSE, dihasilkan jumlah 5 *cluster* ($k=5$) merupakan jumlah *cluster* terbaik dengan jumlah tingkat *error* data dengan *centroid* paling kecil, sehingga didapatkan hasil dari proses *algoritma k-means clustering* dengan $k=5$ sebagai berikut *cluster* 1 terdapat 20 kota asal mahasiswa dengan strategi promosi yang paling efektif yaitu media *online* dan lisan, pada *cluster* 2 terdapat 31 kota asal mahasiswa dengan strategi promosi yang paling efektif yaitu lisan dan media *online*, pada *cluster* 3 terdapat 4 kota asal mahasiswa dengan strategi promosi yang paling efektif adalah *event* dan spanduk, pada *cluster* 4 terdapat 15 kota asal mahasiswa dengan strategi promosi yang paling efektif adalah brosur dan lisan, pada *cluster* 5 terdapat 2 kota asal mahasiswa dengan strategi promosi yang paling efektif adalah lisan dan *event*.

Saran

Penelitian penerapan *Algoritma K-Means Clustering* sebagai rekomendasi strategi promosi penerimaan mahasiswa baru pada Universitas Hasyim Asy'ari ini memberikan saran untuk pengembangan penelitian di masa mendatang, yaitu: penambahan jumlah k untuk mendapatkan hasil yang lebih maksimal dan menggunakan metode *Elbow* untuk lebih mendapatkan jumlah k terbaik.

DAFTAR PUSTAKA

- Agustin, W., Erlin. 2016. Implementasi Metode K-Means *Cluster Analysis* untuk Memilih Strategi Promosi Penerimaan Mahasiswa Baru. *Seminar Nasional Ilmu Komputer*, Semarang, 10 Oktober, hal. 9-15.
- Puspitasari, N., Haviluddin. 2016. Penerapan Metode K-Means Dalam Pengelompokan Curah Hujan di Kalimantan Timur, *Seminar Nasional Riset Ilmu Komputer (SNRIK)*, Vol. 1, Makassar, Desember 14, 20-24.
- Suprawoto, T. 2016. "Klasifikasi Data Mahasiswa Menggunakan Metode K-Means Untuk Menunjang Pemilihan Strategi Pemasaran". *Jurnal Informatika dan Komputer (JIKO)*. Vol.1 (1): hal. 13.
- Susanto, S., Suryadi. D. 2010. *Pengantar Data Mining Menggali Pengetahuan Dari Bongkahan Data*. Yogyakarta: Andi.
- Suyanto. 2017. *Data Mining Untuk Klasifikasi dan Klasterisasi Data*. Bandung: Informatika Bandung.
- Darmi, Y., Setiawan, A. 2016. "Penerapan Metode Clustering K-Means Dalam Pengelompokan Penjualan Produk". *Jurnal Media Infotama* .Vol.12 (2): hal.148-150.