
ALGORITHMAMA K-MEANS CLUSTERING STRATEGI PEMASARAN
PENERIMAAN MAHASISWA BARU UNIVERSITAS SATYA
NEGARA INDONESIA

*Algoritma K-Means Clustering Strategy Marketing Admission
Universitas Satya Negara Indonesia*

Riama Sibarani, riama_sarah@yahoo.com¹⁾, Omby, ombyti@yahoo.com²⁾

¹⁾²⁾ Teknik Informatika /Fakultas Teknik, USNI

ABSTRACT

K-Means is one of the non-hierarchical data grouping methods that seeks to partition existing data into two or more groups. This method partitions the existing data into groups so that the same characteristic data is entered into the same group and the different characterized data are grouped into another group. This study aims to produce an analysis that can cluster or categorize data of graduate students in 2016 and 2017 to make the campaign more targeted and targeted. This research mengklaster graduate student data with attribute address, majors and GPA into three Clusters based on distance (Euclidean). The data that is processed is the graduation data from 2016-2017. This study processed population data. Data processing using SPSS 21 IBM software. The process of clustering as much as 17 iterations with minimum distance between clusters is 6.47. The value of significance indicates that there are significant differences between cluster1, cluster and cluster 3 associated with the attributes in the study. Cluster results show that the address (area) of students come from South Jakarta, Tangerang and West Jakarta for Department of Informatics, Information Systems, Accounting, Manajemen. Department of Environmental Engineering, Manajemen comes from several regions and not grouped. The PSP and BDP departments are from Java, Sumater and other regions. GPA > 3.00 comes from South Jakarta.

Keywords: *K-Means, Cluster, Iteration. Euclidean, Significant.*

ABSTRAK

K-Means merupakan salah satu metode pengelompokan data non-hirarki yang berusaha mempartisi data yang ada kedalam bentuk dua atau lebih kelompok. Metode ini mempartisi data yang ada ke dalam kelompok sehingga data berkarakteristik sama dimasukkan kedalam satu kelompok yang sama dan data yang berkarakteristik berbeda dikelompokkan kedalam kelompok yang lain. Penelitian ini bertujuan untuk menghasilkan suatu analisa yang dapat mengcluster atau mengelompokkan data mahasiswa wisudawan tahun 2016 dan 2017 untuk membuat promosi menjadi lebih tertarget dan tepat sasaran. Penelitian ini mengklaster data mahasiswa lulusan dengan atribut alamat, jurusan dan IPK kedalam tiga Cluster berdasarkan jarak (Euclidean). Data yang diolah adalah data wisudawan tahun 2016-2017. Penelitian ini mengolah data populasi. Proses pengolahan data menggunakan software SPSS 21 IBM. Proses pengklasteran sebanyak 17 iterasi dengan jarak minimal antar cluster adalah 6,47. Nilai signifikansi menunjukkan terdapat perbedaanyang berarti antara cluster1, cluster dan cluster 3 yang berhubungan dengan atribut-atribut pada penelitian. Hasil Cluster menunjukkan bahwa alamat (daerah) mahasiswa berasal dari Jakarta Selatan, Tangerang dan Jakarta Barat untuk jurusan Teknik Informatika, Sistem Informasi, Akutansi, Manajemen. Jurusan Teknik Lingkungan, Manajemen berasal dari beberapa daerah dan tidak mengelompok. Jurusan PSP dan BDP berasal dari Jawa, Sumater dan daerah lain. IPK > 3,00 berasal dari Jakarta Selatan.

Kata Kunci: *K-Means, Cluster, Iterasi. Euclidean, Signifikan*

PENDAHULUAN

Universitas Satya Negara Indonesia adalah lembaga pendidikan tinggi swasta berdiri sejak tahun 1985. Visi Universitas adalah menghasilkan lulusan yang mudah mendapatkan pekerjaan dan mampu menciptakan lapangan pekerjaan (usni.ac.id). Saat ini Universitas Satya Negara Indonesia mempunyai 4 fakultas yaitu Fakultas Teknik, Fakultas Ekonomi, Fakultas ISIP dan Fakultas FPIK. Universitas Satya Negara Indonesia sudah melaksanakan wisuda 28 kali (usni.ac.id). Pada tahun 2016 wisudawan ke 27 jumlah wisudawan adalah 474 yang terdiri dari Fakultas teknik berjumlah 140, Fakultas Ekonomi 203, FIPK 18 dan ISIP 98 (Buku wisudawan 2016). Dan tahun 2017 wisudaw ke 28 Jumlah wisudawan adalah 576 terdiri dari Fakultas Teknik berjumlah 138, Fakultas Ekonomi 294, ISIP 23 dan FPIK 23 (Buku wisudawan 2017).

Berdasarkan peraturan Nomor 12 Tahun 2012 Pasal 73 Perguruan Tinggi menjaga keseimbangan antara jumlah maksimum mahasiswa dalam setiap Program Studi dan kapasitas sarana dan prasarana, Dosen dan tenaga kependidikan, serta layanan dan sumber daya pendidikan lainnya (UU RI 2012). Pada tahun 2016 jumlah mahasiswa yang diwisuda adalah 474 dan pada tahun 2017 jumlah mahasiswa yang diwisuda adalah 576. Jumlah mahasiswa baru tahun akademik 2017/2018 sebanyak 1008. Agar jumlah mahasiswa baru dapat dipertahankan dan bahkan bertambah, maka Universitas harus membuat suatu strategi pemasaran yang efektif dan efisien berdasarkan data-data mahasiswa yaitu: alamat asal sekolah, jurusan. Data tersebut bertumpuk selama bertahun-tahun. Data tersebut perlu ditambang dan dikelompokkan berdasarkan alamat asal sekolah dan jurusan. Pengelompokan data dengan menggunakan *K-Means Clustering*. Algoritma K-Means merupakan salah satu metode data clustering non hirarki yang berusaha mempartisi data ke dalam bentuk satu atau

lebih klaster. Data yang memiliki karakteristik yang sama berada dalam satu klaster dan data yang mempunyai karakteristik yang berbeda dikelompokkan dalam kelompokkan yang lain.

Berdasarkan uraian diatas masalah diidentifikasi sebagai berikut: belum adanya penambangan data tentang alamat asal mahasiswa, jurusan dan Indeks Prestasi mahasiswa. Sehingga bagian pemasaran belum mengetahui klaster-klaster alamat asal, jurusan serta IPK para mahasiswa

Berdasarkan indetifikasi masalah yang telah diuraikan, maka masalah dalam penelitian ini dirumuskan sebagai berikut: Bagaimana menerapkan algoritma K-Means untuk menentukan klaster alamat asal, jurusan dan IP mahasiswa, berdasarkan data wisudawan tahun akademik 2016 dan 2017

Batasan pembahasan dalam penelitian ini adalah; Data yang diolah adalah data mahasiswa lulusan Universitas Satya Negara Indonesia tahun 2016 dan 2017. Metode yang digunakan adalah metode K- Means, Data diolah dengan menggunakan SPSS 21 IBM

Adapun tujuan dari penelitian ini adalah mengimplementasikan algoritma K-Means untuk mengetahui klaster alamat asal, jurusan dan IPK mahasiwa berdasarkan data wisuda tahun 2016-2017. Manfaat dari penelitian ini adalah untuk membantu bagian pemasaran Universitas Satya Negara Indonesia dalam melakukan promosi yang efektif dan efisien.

Pemasaran adalah suatu proses sosial yang di dalamnya individu dan kelompok, mendapatkan apa yang mereka butuhkan dan inginkan dengan menciptakan, menawarkan dan secara bebas mempertukarkan produk yang bernilai dengan pihak lain (Kotler, 2008:87). Menurut Jefkins (2004:107). Angipora (2007:75) mengemukakan bahwa pemasaran dapat didefinisikan dalam dua pengertian dasar, yaitu dalam arti kemasyarakatan dan dalam arti bisnis.

Dalam arti kemasyarakatan, pemasaran adalah setiap kegiatan tukar-menukar yang bertujuan untuk memuaskan keinginan manusia, sedangkan dalam arti bisnis, pemasaran adalah sebuah sistem dari kegiatan bisnis yang dirancang untuk merencanakan, memberi harga, mempromosikan dan mendistribusikan jasa serta barang-barang pemuas keinginan pasar

Kata *algorism* berasal dari nama penulis buku Arab yang terkenal, yaitu Abu Ja'far Muhammad ibnu Musa Al-Khuwarizmi (Al-Khuwarizmi dibaca orang Barat menjadi *algorism*). Dalam bahasa Indonesia, kata *algorithm* diserap menjadi algoritma. (Munir, 2005). Algoritma adalah urutan logis langkah – langkah penyelesaian masalah yang disusun secara sistematis.

Data Mining sering disebut sebagai *Knowledge Discovery in Database (KDD)* yaitu suatu kegiatan yang mencakup pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola hubungan dalam set data yang berukuran sangat besar. *Output* dari data mining ini dapat digunakan untuk menentukan kebijakan atau keputusan di masa yang akan datang. Salah satu teknik yang dikenal dalam data mining yaitu: *clustering*. Pengertian keilmuan dalam data mining adalah pengelompokan sejumlah data atau objek dalam suatu cluster atau group, dimana satu objek data adalah anggota hanya satu cluster. Setiap data dalam suatu *cluster* memiliki kemiripan yang hampir sama dan berbeda dengan objek dalam *cluster* yang lain.

Data mining ini juga dikenal dengan istilah *pattern recognition* (Santosa, 2007). *Data mining* merupakan metode pengolahan data berskala besar oleh karena itu data mining ini memiliki peranan penting dalam bidang industri, keuangan, cuaca, ilmu dan teknologi. Secara umum kajian data mining membahas metode-metode seperti, *clustering*, klasifikasi, regresi, seleksi variable, dan market basket analisis (Santosa, 2007). Salah satu teknik

yang dibuat dalam data mining adalah bagaimana menelusuri data yang ada untuk membangun sebuah model, kemudian menggunakan model tersebut agar dapat mengenali pola data yang lain yang tidak berada dalam basis data yang tersimpan. (Prasetyo E, 2012).

Cluster : *Clustering* merupakan suatu metode untuk mencari dan mengelompokkan data yang memiliki kemiripan karakteristik (*similarity*) antara satu data dengan data yang lain. *Clustering* merupakan salah satu metode *data mining* yang bersifat tanpa arahan (*unsupervised*). Dalam data mining ada dua jenis metode *clustering* yang digunakan dalam pengelompokan data, yaitu *hierarchical clustering* dan *non-hierarchical clustering* (Santosa, 2007). **Hierarchical clustering** adalah suatu metode pengelompokan data yang dimulai dengan mengelompokkan dua atau lebih objek yang memiliki kesamaan paling dekat. Kemudian proses diteruskan ke objek lain yang memiliki kedekatan kedua. Demikian seterusnya sehingga cluster akan membentuk semacam pohon dimana ada hierarki (tingkatan) yang jelas antar objek, dari yang paling mirip sampai yang paling tidak mirip. Secara logika semua objek pada akhirnya hanya akan membentuk sebuah *cluster*. Metode *non-hierarchical clustering* justru dimulai dengan menentukan terlebih dahulu jumlah cluster yang diinginkan (dua cluster, tiga cluster, atau lain sebagainya). Setelah jumlah cluster diketahui, baru proses cluster dilakukan tanpa mengikuti proses hierarki. Metode ini biasa disebut dengan *K-Means Clustering* (Santoso, 2010).

Tahapan data mining terdiri dari:
a). Pembersihan Data (*Data Cleaning*); Pembersihan data merupakan proses menghilangkan *noise* dan data yang tidak konsisten atau data tidak relevan.
b). Integrasi Data (*Data Integration*): Integrasi data merupakan penggabungan data dari berbagai *database* ke dalam satu *database* baru.
c). Seleksi Data (*Data Selection*): Data yang ada pada *database* sering kali

tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari *database*. d). Transformasi Data (*Data Transformation*): data diubah atau digabung kedalam format yang sesuai untuk diproses dalam data mining. e). Proses mining: Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data. f). Evaluasi pola (*Pattern Evaluation*): Untuk mengidentifikasi pola-pola kedalam *knowledge based* yang ditemukan.

Data mining dibagi menjadi beberapa kelompokan berdasarkan tugas yang dilakukan yaitu: a).Estimasi; Model dibangun menggunakan record lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi. b).Prediksi; nilai dari hasil akan ada dimasa mendatang. Contoh prediksi dalam bisnis dan penelitian adalah prediksi harga daging dalam 3 bulan yang akan datang contoh algoritmanya: Linear Regression, Neural Network, Support Vector Machine. c).Klasifikasi: Mengklasifikasikan variable kategori dalam tinggi, sedang, rendah. Algoritma yang digunakan Naives Bayes, K-Nearest Neighbor, C4.5. d)Pengkusteran: pengelompokan record, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. Kluster adalah kumpulan record yang memiliki kemiripan satu dengan yang lainnya ketidak miripan dengan record-record dalam kluster lain. Contoh algoritmanya: K-Means, Fuzzy C-Means, K-Medoids. f).Asosiasi :Tugas asosiasi dalam data mining adalah menemukan atribut yang muncul dalam suatu waktu. contoh algoritmanya: Apriori.

K-Means Clustering merupakan salah satu metode data clustering non-hirarki yang mengelompokan data dalam bentuk satu atau lebih *cluster*/kelompok. Data-data yang memiliki karakteristik yang sama dikelompokan dalam satu *cluster*/kelompok dan data yang memiliki karakteristik yang berbeda dikelompokan dengan *cluster*/kelompok yang lain sehingga data yang berada dalam satu

cluster/kelompok memiliki tingkat variasi yang kecil (Agusta, 2007).

METODOLOGI PENELITIAN

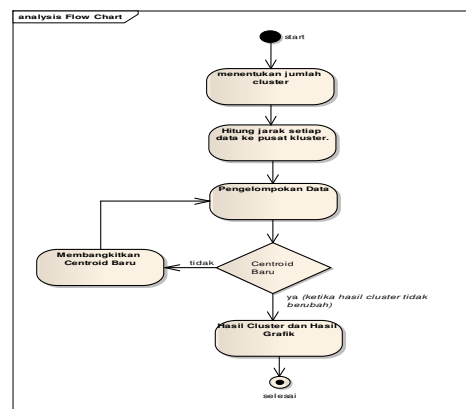
Menurut Santosa (2007), langkah-langkah melakukan clustering dengan metode *K-Means* adalah sebagai berikut:a). Pilih jumlah *cluster k*, pusat cluster diberi nilai awal dengan angka random. b). Alokasikan semua data/objek ke cluster terdekat. Kedekatan dua objek ditentukan berdasarkan jarak kedua objek tersebut. Demikian juga kedekatan suatu data ke *cluster* tertentu ditentukan jarak antara data dengan pusat *cluster*. Untuk menghitung jarak semua data ke setiap titik pusat cluster dapat menggunakan teori jarak *Euclidean* yang dirumuskan sebagai berikut:

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

dimana:

$D(i,j)$ = Jarak data ke i ke pusat cluster j ; X_{ki} = Data ke i pada atribut data ke k ; X_{kj} = Titik pusat ke j pada atribut ke k .

d). Hitung kembali pusat *cluster* dengan keanggotaan *cluster* yang sekarang. Pusat *cluster* adalah rata-rata dari semua data/objek dalam *cluster* tertentu atau menggunakan median dari *cluster* tersebut. e).Tugaskan lagi setiap objek memakai pusat *cluster* yang baru. Jika pusat *cluster* tidak berubah lagi maka proses *clustering* selesai. Atau, kembali ke langkah nomor 3 sampai pusat *cluster* tidak berubah lagi



Gambar 1. Tahapan K-Means

Suatu data akan menjadi anggota dari kluster ke-k apabila jarak data tersebut ke pusat kluster ke-k bernilai paling kecil jika dibandingkan dengan jarak ke pusat kluster lainnya. Gambar 1. Menunjukkan tahapan K-Means

Nilai pusat kluster yang baru dihitung dengan cara mencari nilai rata-rata dari data-data yang menjadi anggota pada kluster tersebut, dengan menggunakan rumus (2)

$$C = \frac{\sum m}{n}; \dots\dots\dots(2)$$

C = Centroid data,
 m=Data yang termasuk dalam centroid data,
 n=jumlah data yang menjadai anggota centroid tertentu.

Objek Penelitian adalah Universitas Satya Negara Indonesia berada di Jalan Arteri Pondok Indah No.11, Kebayoran Lama, RT.4/RW.2, RT.4/RW.2, Kby. Lama Utara, Kby. Lama, Kota Jakarta Selatan, Daerah Khusus Ibukota Jakarta 12240

Data-data yang digunakan pada penelitian data mahasiswa yang telah lulus tahun 2016-2017, sumber data adalah buku wisudawan . Data IPK diperoleh dari bagian BAKPSI. Data-data yang telah didapatkan pada tahap pengumpulan data kemudian dilakukan transformasi pada data-data yang berjenis data string :kota asal dan jurusan diinisialisasikan ke dalam bentuk data ordinal dan nominal. Pengolahan data ini menggunakan teori-teori dari data mining dengan algoritma *K-Means Clustering* atau yang lebih spesifik lagi algoritma *K-Means Clustering*. Dengan metode *K-Means Clustering* ini, data yang diperoleh dikelompokkan berdasar kemiripan karakteristik dari setiap data, sehingga dapat ditemukan informasi yang tersembunyi dari data-data tersebut. Untuk pengolahan data digunakan software SPSS IBM

Untuk melakukan inisialisasi kota asal dilakukan dengan langkah-langkah

sebagai berikut: a). Pada data kota asal terlebih dulu dilakukan pembagian wilayah yang menjadi beberapa bagian wilayah, yaitu: wilayah DKI Jakarta, Wilayah Bogor dan sekitarnya serta wilayah Tangerang dan sekitarnya. b).Kemudian wilayah-wilayah tersebut diurutkan dari yang terbesar berdasarkan frekuensi mahasiswa yang berasal dari wilayah tersebut.

Iterasi dilakukan sampai beberapa kali proses. Iterasi akan berhenti jika tidak ada lagi data yang berpindah dari kluster satu ke kluster lain. Hasil akhirnya ditemukan hasil final kluster *center*.

HASIL DAN PEMBAHASAN

Dengan menggunakan bantuan *software* SPSS 21 IBM, diperoleh *output* sebagai berikut : Tabel 1. initial Cluster menunjukkan pada tahapan awal terbentuk 3 buah cluster. Kemudian metode K-Mean *Cluster* akan menguji dan melakukan iterasi untuk rekolasi data hingga ditemukan tidak ada lagi objek yang berpindah cluster.

Tabel 1. Initial Cluster Centers

	Cluster		
	1	2	3
Zscore(jurusan)	-.14071	1.38530	.24079
Zscore(alamat)	-1.32583	-2.37515	3.92078
Zscore(ipk)	-4.53339	2.10417	1.10298

Hasil dari K-Means adalah final *Cluster Center*, pada *table 2* yang berisi 3 kluster yang membagi data mahasiswa berdasarkan alamat, jurusan dan IPK. Jarak masing-masing cluster terhadap cluster lain.

Table 2 Distances between Final Cluster Centers

Cluster	1	2	3
1		2.518	1.553
2	2.518		2.411
3	1.553	2.411	

Jarak Klaster 1 dan klaster 3 sangat dekat, dan mempunyai kesamaan yang dekat jika dibandingkan dengan sementara klaster 2 Jumlah anggota/ data setiap klaster ditunjukkan pada table 3:

Tabel 3. Number of Cases in each Cluster

Cluster	1	330.000
	2	209.000
	3	475.000
Valid		1014.000
Missing		.000

Pada cluster 1, Jurusan Teknik Informatika, Sistem Informasi, Manajemen Informatika lebih dominan di daerah Jakarta Selatan, Tangerang, dan Jakarta Barat. Dengan IPK tertinggi berada di Jakarta Selatan. Jurusan IKOM dan IHI dominan berasal dari Jakarta Barat dan Tangerang. Pada Cluster 2, Akuntansi dan Manajemen dominan berasal dari Jakarta Barat dan Tangerang. Sementara jurusan yang lain lebih menyebar dari beberapa daerah. IPK tertinggi berasal dari Jakarta Barat. Pada Cluster 3, Jurusan Teknik Informatika dominan berasal dari Jakarta Selatan, Tangerang dan Jakarta Pusat, Jurusan PSP dan BDP berasal dari luar Sumatra dan luar Jawa.

SIMPULAN

Simpulan

Metode K –Means telah berhasil mengelompokkan data alamat, jurusan dan IPK mahasiswa. Hasil *cluster* tersebut adalah sebagai berikut:

Cluster 1, daerah pemasaran yg sangat potensial untuk jurusan teknik informatika, system informasi, akuntansi, manajemen, IHI dan IKOM selain Jakarta Selatan adalah Tangerang dan Jakarta Barat. Jurusan PSP dan BDP bersal dari luar DKI. IPK >3,00 berasal dari Jakarta Selatan .

Cluster 2, daerah pemasaran yang potensial untuk jurusan teknik informatika adalah Jawa dan Sumatera. Untuk jurusan Sistem informasi, AKutansi, Manajemen, IHI dan

IKOM adalah Jakarta Selatan, tangerang, Jakarta Barat dan dan Jakarta Pusat. IPK>3,00 berasal dari Jakarta Selatan.

Cluster 3. Jurusan Teknik Lingkungan, BDP, PSP, MI berasal dari daerah jawa, Sumatera dan lain-lain. Daerah-daerah potensil di Clsuter 3 adalah Jakarta selatan, Tangerang dan Jakarta Barat.

Saran

Berdasarkan hasil penelitian dan kesimpulan penulis menyarankan agar bagian pemasaran dapat mempertimbangkan daerah-daerah atau kota-kota yang menjadi daerah yang potensial untuk dikunjungi.

DAFTAR PUSTAKA

- [1] Agusta, Y. 2007. *K-means - Penerapan, Permasalahan dan Metode Terkait*. Jurnal Sistem dan Informatika Vol. 3 (Februari 2007): 47-60.
- [2] Hidayat, T., dan Istiadah, N. 2011 . *Panduan Lengkap Menguasai SPSS 19 untuk Mengolah Data Statistik Penelitian*. Jakarta: Media Kita.
- [3] Hair, J.F. Jr. Anderson, R.E. Tatham, R.L., dan Black W. C. (1992). *Multivariate Data Analysis, Fifth edition*. New Jersey: Prentice Hall.
- [4] Lattin, J., carrol, D., dan green, P. (2003). *Analying Multivariate Data*. United State of America: Thomson Learning
- [5] Myamoto, S and Agusta, Y (1995). *Agorithm of Hard C-Means Clastein using KernelFunncion inSupport Vectro machines*, Journal of Advnced Computational Intellingence and Inttellingent Informatics, vol 7. No. 1 pp 25-30
- [6] Tahta Alfina, dkk: *Analisa Perbandingan Metode HierarchicalClustering, K-means dan Gabungan Keduanya dalam Cluster Data*(Studi kasus : Problem Kerja Praktek Jurusan Teknik Industri, Fakultas Teknologi Industri, Institut

- Teknologi Sepuluh Nopember (ITS),
JURNAL TEKNIK ITS Vol. 1, (Sept,
2012) ISSN: 2301-9271, Jurusan Teknik
Industri ITS)
- [7] Santoso, S. 2010. *Statistik Multivariat*.
Jakarta: Elex Media Komputindo.
- [8] Santosa, B. 2007. *Data Mining: Teknik
Pemanfaatan Data untuk Keperluan
Bisnis*. Yogyakarta: Graha Ilmu.
- [9] Supranto, J. M.A. 2000. *Statistik: Teori
dan Aplikasi Edisi Keenam*. Jakarta:
Erlangga