

**CLUSTERING SEBARAN ALUMNI PROGRAM STUDI SISTEM INFORMASI
POLITEKNIK NEGERI NUSA UTARA**

*ALUMNUS CLUSTERING SPRED of INFORMATION SYSTEM STUDY PROGRAM
POLYTECHINC STATE of NUSA UTARA*

Noldy Sinsu, Oktavianus Lumasuge, Steve Arthur Sehang

Sistem Informasi, Politeknik Negeri Nusa Utara

Email: sinsu.tekno@gmail.com

Abstrak: Teknologi database saat ini memungkinkan untuk menyimpan sejumlah data dalam jumlah yang sangat besar dan terakumulasi namun disinilah awal timbulnya persoalan dengan semakin banyaknya data, seperti pada Program Studi Sistem Informasi Politeknik Negeri Nusa Utara. Oleh sebab itu sangat penting untuk mengetahui sebaran mahasiswa dengan menggunakan tracer alumni, sehingga data yang ada dapat dipakai guna mengelompokkan sebaran mahasiswa berdasarkan kesamaan ciri dari data menggunakan metode k-means clustering. Penelitian ini bertujuan untuk memudahkan menganalisis pengelompokan sebaran mahasiswa. Data tracer tersebut diperoleh data alumni yang ada pada Program Studi Sistem Informasi, berdasarkan data alumni yang ada informasi yang tersembunyi dapat diketahui dengan cara melakukan pengolahan terhadap data tersebut sehingga berguna bagi pihak Program Studi Sistem Informasi. Penelitian ini menganalisis Tracer Alumni Program Studi Sistem Informasi dari angkatan 2006 sampai dengan angkatan 2015 dengan menggunakan algoritma k-means clustering menggunakan microsoft excel. Atribut yang digunakan adalah domisili, waktu masuk, waktu wisuda dan waktu tunggu kerja. Cluster yang terbentuk adalah dua cluster. Hasil dari penelitian ini digunakan sebagai salah satu dasar pengambilan keputusan untuk menentukan strategi promosi berdasarkan cluster yang terbentuk oleh pihak Program Studi Sistem Informasi Politeknik Negeri Nusa Utara.

Kata kunci: Clustering, K-Means, Sebaran, Alumni.

Abstract: *In this era database technology makes it possible to store a large amount of data and accumulates, but this is where the beginning of the problem arises with the increasing number of data, such as in Polytechnic State of Nusa Information System Study Program. That's way it I was very important to know the distribution of students use alumni tracers, so the available data can be used to classify student distribution based on the similarity of features of the data using the K-means Clustering Method. This study aims make easy analyze distribution of student distribution. Tracer data is obtained alumni data in the Information Systems Study Program, based on existing alumnus, and the hidden data about information can be known by processing it so it is useful for the Information Systems Study Program. This research introduces Information Systems Study Program Alumni Tracer from class of 2006 to the class of 2015 used the K-Means Clustering Algorithm by used Microsoft Excel. The attributes were used domicile, time of entry, graduation time and waiting time for work. Clusters were formed two clusters. The results of this study were used as a basic to made decision to determine promotion strategies based on clusters formed by the Polytechic State of Nusa Utara in Information System Study Program.*

Keyword: *Clustering, K-Means, Distribution, Alumnus.*

PENDAHULUAN

Dalam pemanfaatan dataset disuatu lembaga pendidikan maupun dalam suatu instansi sangat diperlukan sebab pemanfaatan teknologi informasi

disuatu instansi perlu ditingkatkan dalam kajian pengembangan sumber daya manusia dari segi serapan lulusan, oleh karena itu Teknologi informasi bisa mengalami perkembangan menjadi sistem informasi

basis data terintegrasi di segala aspek bidang seperti perdagangan, pemerintahan, organisasi, pendidikan, dunia hiburan dan sekelompok manusia untuk menyampaikan maupun mengolah informasi secara cepat, efektif dan efisien. Berdasarkan Undang-Undang Republik Indonesia Nomor 15 tahun 2018 tentang Organisasi dan Tata Kerja Lembaga Layanan Pendidikan Tinggi pasal 18 ayat 2 menyatakan bahwa, bagian sistem informasi dan kerja sama mempunyai tugas melakukan pengumpulan dan pengolahan data kerja sama, pengelolaan pangkalan data pendidikan tinggi, dan fasilitasi pengembangan kerja sama perguruan tinggi.

Program studi Sistem Informasi merupakan salah satu program studi yang terdapat di Politeknik Negeri Nusa Utara, dimana sudah banyak menghasilkan lulusan Ahli Madya Komputer yang tersebar di berbagai lingkungan pekerjaan dengan begitu Politeknik Negeri Nusa Utara khususnya Program Studi Sistem Informasi telah banyak memproses ratusan data mahasiswa setiap tahunnya. Pihak Program Studi menggunakan sistem manajemen basis data tradisional dalam pengelolaan datanya.

Oleh karena itu penggunaan aplikasi data mining dapat membantu pihak Program Studi dalam melakukan analisis data dan menentukan rencana strategi promosi dari sebaran mahasiswa berdasarkan geografis. Salah satu metode yang terkenal pada data mining untuk mengelompokkan data adalah dengan metode analisis *cluster* atau *clustering*.

K-Means merupakan salah satu metode data clustering non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih cluster / kelompok. Metode ini mempartisi ke dalam cluster / kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama dan yang memiliki karakteristik yang berbeda dikelompokkan pada kelompok yang lain (Giyanto, 2008). Banyak algoritma pengelompokan terpusat telah diperkenalkan dan dianalisis dalam literature (A.K. Jain, M.N dan Murty, P.J. Flynn., 1999). Dibandingkan dengan sejarah panjang penelitian

tentang teknik pengelompokan terpusat, pengelompokan terdistribusi relatif dapat lebih efisien dalam mempermudah suatu pekerjaan (P. Shen dan C. Li., 2014). Berdasarkan uraian permasalahan sebelumnya, metode K-Means Clustering dapat digunakan menganalisis data sebaran lulusan dari Program Studi Sistem Informasi pada Politeknik Negeri Nusa Utara. Hasil dari penelitian ini nanti bukan sekedar dapat membantu Program Studi Sistem Informasi untuk dapat mengetahui sebaran lulusan, tetapi juga dapat menjadi acuan dalam peningkatan kompetensi lulusan serta sumber daya manusia khususnya pengajar di Program Studi Sistem Informasi itu sendiri, dengan tujuan setiap lulusan dapat memiliki kompetensi yang sesuai dengan kebutuhan yang ada dilapangan kerja.

METODE PELAKSANAAN

a. K-Means Clustering

Mengorganisasi data ke dalam kelompok merupakan salah satu pendekatan yang paling mendasar dari pemahaman dan pembelajaran. *Cluster Analysis* adalah studi formal metode dan algoritma untuk pengelompokan, atau *clustering*, objek yang diukur memiliki karakteristik intrinsik atau kesamaan. Sampel untuk *clustering* diwakili sebagai vektor pengukuran dan sebagai titik dalam ruang multidimensi. Sampel dalam *cluster* yang valid lebih mirip satu sama lain dari pada 21 sampel *cluster* yang berbeda. Metodologi *clustering* sangat cocok untuk eksplorasi hubungan timbal balik antara sampel untuk membuat penilaian awal pada struktur sampel. Kemampuan manusia yang kompetitif dengan otomatis pengelompokan prosedur dalam satu, dua, atau tiga dimensi, tetapi kebanyakan masalah nyata melibatkan pengelompokan dalam dimensi yang lebih tinggi. Hal ini sangat sulit bagi manusia untuk intuitif menginterpretasikan data tertanam dalam dimensi ruang yang tinggi (Mehmed Kantardzic, 2011)

Teknik *cluster analysis* berfokus pada mengeksplorasi data set untuk menilai apakah data dapat diringkas bermakna dalam hal yang relatif pada

kelompok kecil atau kelompok objek atau individu yang mirip satu sama lain dan yang berbeda dalam beberapa hal dari individu dalam kelompok lain. analisis *cluster* memeriksa beberapa pendekatan grafis yang dapat membantu dalam mengungkap *cluster* struktur, dan kemudian mempertimbangkan pengukuran kesamaan, perbedaan dan jarak, yang merupakan pusat untuk banyak teknik *clustering* (Brian S. et al, 2011: 13).

1. Menentukan jumlah *cluster*

$$k = \sqrt{\frac{n}{2}} \tag{1}$$

2. Menentukan *centroid*

$$C_i = \frac{\sum_{x_i \in S_i} x_i}{n} \tag{2}$$

3. Menghitung jarak objek ke *centroid*

$$d(x, y) = |x - y| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \tag{3}$$

4. Mengelompokkan objek berdasarkan jarak terdekat
Sebelum pengelompokan data dilakukan, pertama harus dilakukan perhitungan untuk menentukan d (distance) yang bernilai minimum pada setiap data untuk menjadi anggota klaster. Selanjutnya dapat dilakukan pengelompokan data pada setiap *cluster*.

5. Cek apakah kelompok data konvergen
Jika kelompok data yang baru adalah sama dengan kelompok data sebelumnya (konvergen), maka proses clustering selesai. Jika tidak, maka lakukan iterasi dimulai dari penentuan pusat klaster baru (iterasi 2).

Kriteria yang digunakan didefinisikan dalam kode seperti, domisili (C1), tahun masuk (C2), tahun wisuda (C3), waktu tunggu kerja (C4). Data dari empat (4) kriteria yang diambil dari dua puluh tujuh (136) mahasiswa ditunjukkan pada Tabel 1.

Tabel 1. Data Alumni

No	Nama Mahasiswa	Kriteria			
		C1	C2	C3	C4
1	M1	1	2014	2017	12
2	M2	1	2014	2017	1
3	M3	1	2013	2016	6
4	M4	1	2014	2017	12
5	M5	1	2011	2014	1
6	M6	1	2011	2014	-1
....
131	M131	1	2015	2018	0
132	M132	1	2015	2018	0
133	M133	1	2015	2018	1
134	M134	2	2015	2018	6
135	M135	1	2015	2018	0
136	M136	1	2015	2018	0

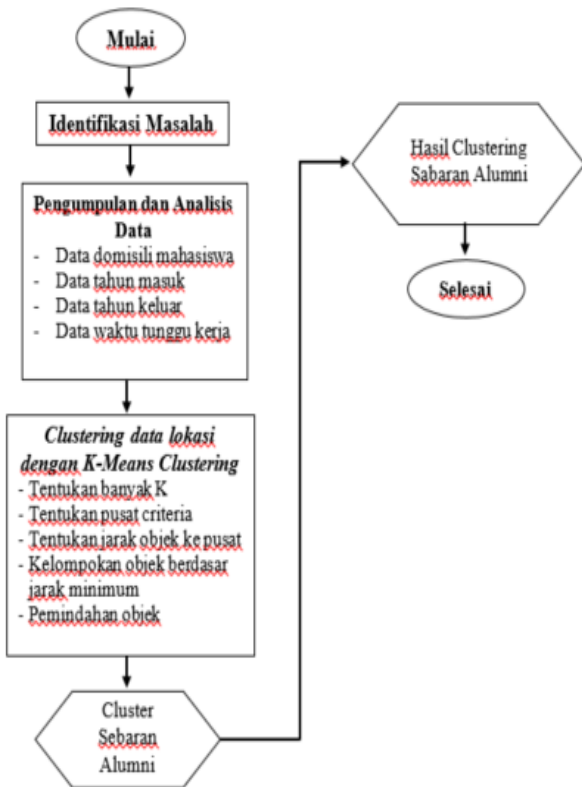
b. Tahapan Penelitian

Tahapan yang akan dilakukan yaitu pengumpulan dan analisa data, pembentukan variabel dan preprocessing data variabel yang digunakan pada proses *K-means clustering*. Pemodelan data atau

analisis data merupakan tahapan yang harus diselesaikan terlebih dahulu sebelum melakukan proses selanjutnya, dimana pada tahapan pemodelan data yang diperhatikan yakni mengklasifikasikan data yang sudah dikumpulkan menjadi variabel yang nantinya akan

digunakan pada proses *K-means clustering* dengan langkah-langkah sebagai berikut :

- a) Menentukan jumlah *cluster*
- b) Menentukan *centroid*
- c) Menghitung jarak objek ke *centroid*
- d) Mengelompokkan objek berdasarkan jarak terdekat
- e) Cek apakah kelompok data konvergen atau tidak



Gambar 1. Tahapan Penelitian

HASIL DAN PEMBAHASAN

Untuk mencapai tujuan penelitian yang telah dibahas sebelumnya, terdapat beberapa tahapan yang harus dilakukan. Tahapan tersebut terbagi atas beberapa proses meliputi, evaluasi parameter sebelum masuk ke proses pre-processing merupakan bagian dari normalisasi data yang digunakan untuk penerapan dari metode *K-Means Clustering* perhitungan penentuan nilai *K*, penentuan jarak terdekat dengan objek, serta membuat *cluster* baru dan melakukan pengelompokan data.

a. Penerapan *K-Means Clustering*

Proses penerapan *K-Means Clustering*, dimana proses ini merupakan tahapan yang meliputi, pre-

processing data yang digunakan untuk penentuan nilai *K*, kemudian dilanjutkan dengan menentukan *centroid* awal, selanjutnya menghitung jarak objek ke *centroid*, mengelompokkan objek serta melakukan pengecekan kelompok data apakah sudah *convergen*.

1. Pre-processing

Pada tahapan proses pre-processing dilakukan karena nilai real yang dimasukkan kedalam sistem memiliki variasi nilai yang cukup jauh berbeda, seperti halnya pada dataset kriteria domisili, tahun masuk dan tahun wisudamemiliki nilai yang tidak menentu dibandingkan dengan nilai pada kriteria lainnya sehingga diperlukan proses pre-processing dengan menggunakan min max sebagai normalisasi data untuk semua kriteria yang ada yaitu sebanyak empat kriteria yang digunakan. Hasil dari proses pre-processing dapat dilihat pada table 2.

Tabel 2. Data Pre-processing

No	Nama Mahasiswa	Kriteria			
		C1	C2	C3	C4
1	M1	0	0.8	0.8	0.351351351
2	M2	0	0.8	0.8	0.054054054
3	M3	0	0.7	0.7	0.189189189
4	M4	0	0.8	0.8	0.351351351
5	M5	0	0.5	0.5	0.054054054
.....
132	M132	1	2015	2018	0
133	M133	1	2015	2018	1
134	M134	2	2015	2018	6
135	M135	1	2015	2018	0
136	M136	1	2015	2018	0

2. Menentukan Jumlah *Cluster*

Proses selanjutnya menentukan jumlah *cluster* yang akan di kelompokkan, dimana tahapan awalnya dengan menentukan nilai *K* dengan menggunakan persamaan dari *Rule Of Thumb*. Dimana *K* merupakan banyaknya jumlah *cluster* yang didapatkan dari akar pangkat *n* jumlah objek yang akan dikelompokkan dibagi dengan nilai dua (2). Maka nilai *K* yang didapatkan dari hasil penerapan

persamaan rule of thumb yakni ($k = 8$). Hasil perhitungan dari rule of thumb dapat dilihat pada persamaan (2.1) seperti dibawah dibawah ini.

$$k = \sqrt{\frac{n}{2}}$$

$$= \sqrt{\frac{136}{2}} = 8$$

$k = 8$

3. Menentukan *Centroid*

Tahapan dalam menentukan centroid awal dilakukan secara acak dengan metode random. Setelah dilakukan random data, maka *centroid* awal didapatkan yakni data dari ID 113, ID 9, ID 126 dan ID 133. Hasil dari penentuan secara random untuk centroid awal dapat dilihat pada table 3.

Tabel 3. Hasil Penentuan *Centroid* Awal

No	Nama Mahasiswa	Kriteria			
		C1	C2	C3	C4
113	M113	0	0.5	0.5	0.108108108
9	M9	1	0.5	0.5	0.081081081
126	M126	0	0.1	0.1	0.135135135
133	M133	0	0.9	0.9	0.054054054

4. Menghitung Jarak Objek ke *Centroid*

Setelah didapatkan hasil random untuk *centroid* awal maka dilakukan proses menghitung jarak objek dengan centroid serta melakukan pelebelan (pengelompokan) dari setiap data mahasiswa. Dimana setiap kelompok yang terbentuk digunakan untuk menyesuaikan dengan kelompok yang akan terbentuk berikutnya. Dalam proses menghitung jarak objek dengan *centroid* digunakan persamaan Euclidian Distance. Hasil dari menghitung jarak objek dengan *centroid* dapat dilihat pada table 4.

Tabel 4. Hasil menghitung jarak objek dan *centroid*

JARAK OBJEK DENGAN TITIK PUSAT				JARAK TERKECIL	KELOMPOK DATA			
JR1	JR2	JR3	R4		KD1	KD2	KD3	KD4
0.190517	1.02264	0.51794	0.675832	0.1905166	1			
0.489047	1.119395	1.013286	0.32922	0.3292198				1
0.427694	1.086614	0.993264	0.141421	0.1414214				1
0.294235	1.044838	0.850248	0.313467	0.2942348	1			
0.489047	1.119395	1.013286	0.32922	0.3292198				1
.....
0.571467	1.150183	1.136524	0.027027	0.027027				1
0.568262	1.14923	1.134273	0	0				1
1.15177	0.575923	1.510934	1.009089	0.5759231		1		
0.571467	1.150183	1.136524	0.027027	0.027027				1
0.571467	1.150183	1.136524	0.027027	0.027027				1

5. Mengelompokkan Objek Berdasarkan Jarak Terdekat

Setelah dilakukan perhitungan untuk menentukan d (distance) yang bernilai minimum pada setiap data untuk menjadi anggota *cluster*. Selanjutnya dapat dilakukan pengelompokan data pada setiap *cluster*. Untuk mengetahui kelompok objek dengan jarak terdekat dapat dilihat pada table

5, angka 1 (satu) pada kelompok data yang baru terbentuk diinisialisasi sebagai nilai yang mendekati pusat *cluster*.

Tabel 5. Hasil Pengelompokan data

JARAK OBJEK DENGAN TITIK PUSAT				JARAK TERKECIL	KELOMPOK DATA			
JR1	JR2	JR3	R4		KD1	KD2	KD3	KD4
0.190517	1.02264	0.51794	0.675832	0.1905166	1			
0.489047	1.119395	1.013286	0.32922	0.3292198				1
0.427694	1.086614	0.993264	0.141421	0.1414214				1
0.294235	1.044838	0.850248	0.313467	0.2942348	1			
0.489047	1.119395	1.013286	0.32922	0.3292198				1
.....
0.571467	1.150183	1.136524	0.027027	0.027027				1
0.568262	1.14923	1.134273	0	0				1
1.15177	0.575923	1.510934	1.009089	0.5759231		1		
0.571467	1.150183	1.136524	0.027027	0.027027				1
0.571467	1.150183	1.136524	0.027027	0.027027				1

Tabel 6. Hasil Pengecekan Data yang Konvergen

JARAK OBJEK DENGAN TITIK PUSAT				JARAK TERKECIL	KELOMPOK DATA LAMA				KELOMPOK DATA BARU				KONVERGEN
JR1	JR2	JR3	R4		KD1	KD2	KD3	KD4	KD1	KD2	KD3	KD4	
0.190517	1.02264	0.51794	0.675832	0.1905166	1				1				Ya
0.489047	1.119395	1.013286	0.32922	0.3292198				1		1			Tidak
0.427694	1.086614	0.993264	0.141421	0.1414214				1		1			Tidak
0.294235	1.044838	0.850248	0.313467	0.2942348	1				1				Tidak
0.489047	1.119395	1.013286	0.32922	0.3292198				1		1			Tidak
.....
0.571467	1.150183	1.136524	0.027027	0.027027				1		1			Tidak
0.568262	1.14923	1.134273	0	0				1		1			Tidak
1.15177	0.575923	1.510934	1.009089	0.5759231		1					1		Tidak
0.571467	1.150183	1.136524	0.027027	0.027027				1		1			Tidak
0.571467	1.150183	1.136524	0.027027	0.027027				1		1			Tidak

Selanjutnya masuk pada tahap iterasi dimana pada tahapan dilakukan agar supaya pusat *cluster* tidak berubah lagi atau stabil. Proses ini dilakukan secara terus menerus sampai pusat *cluster* dinyatakan konvergen. Dimana setiap nilai pada kelompok data lama dijumlahkan kemudian dibagi dengan satu dan banyaknya data dalam satu kelompok serta dikalikan untuk mendapatkan hasil *centroid* baru pada iterasi pertama, sebelum masuk ke penentuan jarak dari objek ke pusat *cluster*.

Tabel 7. Hasil Sebaran Alumni

No	Nama Mahasiswa	Domisili	Tahun Masuk	Tahun Wisuda	Waktu Tunggu Kerja
1	M1	Sangihe	2014	2017	12
2	M2	Sangihe	2014	2017	1
3	M3	Sangihe	2013	2016	6
4	M4	Sangihe	2014	2017	12
5	M5	Sangihe	2011	2014	1
....
132	M132	Sangihe	2015	2018	0
133	M133	Sangihe	2015	2018	1
134	M134	Sangihe	2015	2018	6

135	M135	Sangihe	2015	2018	0
136	M136	Sangihe	2015	2018	0

KESIMPULAN

Penelitian ini berhasil melakukan cluster terhadap data tracer alumni. Dimana hasil yang didapatkan bahwa untuk sebaran alumni dari Program Studi Sistem Informasi Politeknik Negeri Nusa Utara banyak tersebar di Kepulauan Sangihe dan berpotensi mendapatkan pekerjaan dan alumni dari Program Studi Sistem Informasi masih memiliki peluang dilapangan kerja khususnya di Kepulauan Sangihe maupun diluar Kepulauan Sangihe.

DAFTAR RUJUKAN

- Brian S. Everitt, et al. 2011. Cluster Analysis 5th Edition. King's College London, UK.
- Giyanto, dan Heribertus., 2008. Penerapan algoritma Clustering K-Means, K-Medoid, Gath Geva. Tesis Tidak Terpublikasi. Yogyakarta: Universitas Gajah Mada.

- Santosa, B. (2007). Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisni Yogyakarta: Graha Ilmu.
- Jain, A. K., Murthy, M. N. & Flynn, P. J. 1999. Data Clustering: A Review. ACM Computing Surveys, Vol. 31, No. 3.
- Giyanto, dan Heribertus., 2008. Penerapan algoritma Clustering K-Means, K-Medoid, Gath Geva. Tesis Tidak Terpublikasi. Yogyakarta: Universitas Gajah Mada.Darmawan
- Erristhya. 2013. Comparison Of Method Supervised And Unsupervised Through Google Satellite Image Analysis Procedure For Land Use.
- Zhang, L dan Huang Yu., (2014). Optimal Job Selection of University Graduates Based on Grey Clustering.
- P. Shen , C. Li , (2014). Distributed information theoretic clustering, IEEE Trans. Signal Process. 62 3442–3453.
- Han, J., Pei, J., and Yin, Y. (2000). Mining frequent patterns without candidate generation. In SIGMOD '00.
- Han, J., dan Kamber, M., 2001. Data Mining Concepts and Techniques Second Edition. San Francisco: Morgan Kauffman.
- Mehmed, Kantardzic. 2011. Data Mining Concepts, Models, Methods, And Algorithms Second Edition, Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
- Prasetyo, Eko. (2014). Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab, Yogyakarta : ANDI. Peraturan Menteri Riset, Teknologi, Dan Pendidikan Tinggi Republik Indonesia Nomor 15 Tahun 2018.