

# CLUSTERING DATA AKADEMIK MAHASISWA FAKULTAS TEKNIK USNI DENGAN ALGORITMA K-MEANS

Bosar Pandjaitan, S.Si., M.Kom

---

## ABSTRAK

Data akademik fakultas teknik Universitas Satya Negara Indonesia (USNI) perlu di analisis untuk mendapatkan beberapa pola data yang bisa digunakan dalam menentukan strategi manajemen. Peneliti mencoba meneliti data akademik tersebut untuk menentukan pola penerimaan calon mahasiswa fakultas teknik USNI. Adapun metode yang digunakan adalah metode Metode K-Means dalam mengelompokkan data-data tersebut.

Tujuan dari penelitian ini adalah mengelompokkan data dengan instrumen nilai UAN dan IPK terakhir. Hasil dari clustering diproyeksikan ke data-data calon mahasiswa untuk mencari asal sekolah dari mahasiswa tersebut.

Hasil penelitian Menunjukkan bahwa dari 240 orang mahasiswa menunjukkan bahwa ada 58 ada di cluster-1, 92 orang di cluster-2, dan 90 orang di cluster-3. Dari 90 orang yang ada di cluster-3 dan sekaligus menjadi sasaran Penerimaan Mahasiswa Baru (PMB) diperoleh dari SMK Media informatika 4 orang, SMK Budi Mulia 3 orang, SMK Triguna 2 orang, SMK Muhammadiyah 9 2 orang, SMK YUPPEN EK1 3 orang, SMK N 15 2 orang, SMK Pertiwi 2 2 orang, SMK KARTIKA 2 orang, selainnya 1 orang. Hasil ini menunjukkan bahwa tidak ada asal sekolah yang dominan

Kata kunci: Clustering Data Akademik

## ABSTRACT

*The academic data of technical faculty of Satya Negara Indonesia University (USNI) needs to be analyzed some data patterns that can be used in determining management strategy. Researchers try to research this akademik data to determine patterns of acceptance of students of USNI engineering faculty. The method used is K-Means method in grouping the data.*

*The purpose of this research is to group data with UAN instrument and final GPA. The results of clustering are projected onto the data of prospective students to find the origin of the school from the student.*

*The results showed that 240 students showed that there were 58 in cluster-1, 92 in cluster-2, and 90 in cluster-3. The 90 people in cluster-3 and at the same time become the target of New Student Admission (PMB) obtained from SMK Media informatics 4 students, SMK Budi Mulia 3 students, SMK Triguna 2 students, SMK Muhammadiyah-9 2 students, SMK YUPPEN EK-1 3 students, SMK N- 15 2 students, SMK Pertiwi -2 2 students, SMK KARTIKA 2 students and others school 1 students. These results indicate that there is no dominant school origin*

*Keywords: Clustering Academic Data*

---

## PENDAHULUAN

### Latar Belakang Masalah

---

Fakultas Teknik Universitas Satya Negara Indonesia (USNI) sebagai institusi pendidikan telah memiliki data akademik dan administrasi dalam jumlah yang sangat besar, tetapi belum dimanfaatkan secara maksimal. Data akademik mahasiswa merupakan data yang dihimpun dari hasil kegiatan penerimaan mahasiswa baru, dan proses belajar lulus dari suatu perguruan tinggi. Data pribadi mahasiswa, data rencana studi, dan data hasil studi (nilai dan indeks prestasi) merupakan contoh data akademik suatu perguruan tinggi. Disamping itu, bagian admisi pada setiap awal tahun akademik melakukan rekrutmen calon mahasiswa baru. Data mahasiswa baru yang dihimpun pada saat pendaftaran dan/atau registrasi berupa nilai ujian nasional (UN) rata-rata dan/atau nilai hasil tes masuk perguruan tinggi. Jumlah data yang terakumulasi dari tahun ke tahun perlu dilakukan analisis untuk dapat membuka peluang dihasilkannya informasi yang berguna dalam pembuatan alternatif keputusan bagi manajemen perguruan tinggi.

Untuk memperoleh calon mahasiswa yang lebih banyak, diperlukan strategi promosi yang tepat dengan dasar pengolan data terhadap data-data mahasiswa baru yang sudah ada antara lain: asal sekolah, prestasi akademik, lokasi tempat tinggal, tren keminatan calon mahasiswa, kemampuan calon mahasiswa (akademik dan/atau finansial), proses pembelajaran, serta prospek lulusan kedepannya

### **Rumusan Masalah**

Adapun rumusan masalah dalam penelitian ini adalah Bagaimana mengklaster data akademik mahasiswa Fakultas Teknik USNI dengan algoritma K-Means?

### **Tujuan Penelitian**

Adapun tujuan dari penelitian ini adalah meng-Cluster data akademik mahasiswa fakultas teknik Universitas Satya Negara Indonesia

### **Manfaat Penelitian**

Membantu pihak manajemen Fakultas Teknik dalam menyusun strategi penerimaan mahasiswa baru.

## **LANDASAN TEORI**

### **Data Mining**

Teknologi *data mining* merupakan salah satu alat bantu untuk penggalian data pada basis data berukuran besar dengan spesifikasi kerumitan tinggi dan telah banyak digunakan pada lingkungan aplikasi bisnis seperti perbankan, *provider* telekomunikasi, perusahaan pertambangan, perminyakan, dan lain-lain [1]. *Clustering* merupakan salah satu teknik *data mining* yang berfungsi melakukan pengelompokan sejumlah data atau objek ke dalam *cluster* (group) sehingga setiap *cluster* akan berisi data yang semirip mungkin dan berbeda dengan objek dalam *cluster* yang lainnya. Teknik klasifikasi merupakan pendekatan fungsi klasifikasi dalam data mining yang digunakan untuk melakukan prediksi atas informasi yang belum diketahui sebelumnya. Beberapa algoritma yang dapat digunakan antara lain adalah algoritma Decision Tree C.45, Artificial Neural Networks (ANN), K-Nearest Neighbor (KNN), algoritma Naive Bayes, Neural Network serta algoritma lainnya.

### **Clustering**

Klasterisasi merupakan teknik atau metode untuk mengelompokkan sejumlah besar data menjadi suatu bagian-bagian kecil data yang mempunyai atribut kemiripan dalam sifat, letak, ciri atau filter lain yang telah ditentukan. Teknik atau metode klasterisasi dapat dikelompokkan menjadi dua kategori besar, yaitu : (1) Metode Hirarki (Hierarchical

Clustering) dan (2) Metode Non-Hirarki/Partisi (Partitional Hierarchical). Referensi [1] menunjukkan bahwa Algoritma K-Modes merupakan metode pengembangan dari K-Means yang mampu mengelompokkan data kategorikal dan menghasilkan kluster yang lebih stabil dengan waktu komputasi yang lebih singkat daripada metode K-Means.

Teknik klasifikasi merupakan pendekatan untuk menjalankan fungsi klasifikasi dalam data mining yaitu untuk menggolongkan data. Teknik klasifikasi ini dapat pula digunakan untuk melakukan prediksi atas informasi yang belum diketahui sebelumnya. Beberapa algoritma yang dapat digunakan antara lain adalah algoritma Decision Tree C.45, Artificial Neural Networks (ANN), K-Nearest Neighbor (KNN), algoritma Naive Bayes, Neural Network serta algoritma lainnya.

Tujuan pekerjaan pengelompokan (*clustering*) data dapat dibedakan menjadi dua, yaitu pengelompokan untuk pemahaman dan pengelompokan untuk penggunaan. Jika tujuannya untuk pemahaman, kelompok dalam tujuan ini hanya sebagai proses awal untuk kemudian dilanjutkan dengan pekerjaan inti seperti peringkasan atau *summarization* (rata-rata, atau standar deviasi), pelabelan kelas untuk setiap kelompok untuk kemudian digunakan sebagai data latih klasifikasi, dan sebagainya. Sementara jika untuk penggunaan, tujuan utama pengelompokan biasanya adalah mencari prototipe kelompok yang paling representatif terhadap data, memberikan abstraksi dari setiap objek data dalam kelompok dimana sebuah data terletak di dalamnya.

### K-Means

K-Means merupakan salah satu metode pengelompokan data nonhierarki(sekatan) yang berusaha mempartisi data yang ada ke dalam bentuk dua atau lebih kelompok. Metode ini mempartisi data ke dalam kelompok sehingga data yang berkarakteristik berbeda dikelompokkan yang lain. Adapun tujuan pengelompokan data ini adalah untuk meminimalkan fungsi objektif yang diset dalam pengelompokan, yang pada umumnya berusaha meminimalkan variasi di dalam satu kelompok dan memaksimalkan variasi antar kelompok.

#### Algoritma K-Means:

- 1) Menentukan Jumlah kelompok;
- 2) Menentukan koordinat titik tengah setiap *cluster*;
- 3) Menentukan jarak setiap objek terhadap koordinat titik tengah;
- 4) Mengelompokkan objek-objek tersebut berdasarkan pada jarak minimumnya;
- 5) Tentukan pusat *cluster* baru ;
- 6) Apakah ada selisih antara *cluster* lama dengan *cluster* baru?, Jika masih ada kembali ke langkah-1, jika tidak lanjut kelangkah 6.
- 7) Selesai.

Pada langkah 2 lokasi sentroid(titik pusat) setiap kelompok yang diambil dari rata-rata (mean) semua nilai data pada setiap fiturnya harus dihitung kembali. Jika M menyatakan jumlah data dalam sebuah kelompok, i menyatakan fitur ke –i dalam sebuah kelompok, dan P menyatakan dimensi data, untuk menghitung sentroid fitur ke-i digunakan formula:

$$C_i = \frac{1}{M} \sum_{j=1}^M x_j$$

Formula tersebut dilakukan sebanyak P dimensi sehingga i mulai dari 1 sampai P. Sedangkan pada langkah 4 dilakukan perhitungan jarak dengan menggunakan formula Euclidean:

$$D(x_2, x_1) = \sqrt{(\sum_{j=1}^P |x_{2j} - x_{1j}|)}.$$

## METODE PENELITIAN

### Objek Penelitian

Objek penelitian bersumber dari data mahasiswa yang telah melakukan registrasi di Universitas Satia Negara Indonesia. Atribut data yang akan digunakan adalah nama mahasiswa, jurusan SLTA, nilai UAN, kota asal mahasiswa, program studi yang dipilih dan IPK. Hasil yang diperoleh dari penelitian ini diharapkan dapat menampilkan profil mahasiswa, keterkaitan antara nilai UN terhadap IPK mahasiswa, serta sebaran asal sekolahnya. Dari hasil analisis tersebut diharapkan dapat membantu pihak admisi perguruan tinggi dalam menyusun strategi promosi program studi yang menjadi target. Penggalan informasi pada sebuah data yang berukuran besar (mempunyai jumlah *record* dan jumlah atribut yang cukup banyak) tidak dapat dilakukan dengan mudah.

Prosedur pelaksanaan yang dilakukan dalam penelitian ini meliputi:

- 1) Menghimpun data yang bersumber dari bagian akademik dan admisi untuk mahasiswa jenjang D3 dan S1, selanjutnya diintegrasikan untuk dilakukan *data cleaning*.
- 2) Menentukan jumlah *cluster* yang akan diproses lebih lanjut dan menetapkan titik pusat dari masing-masing *cluster*.
- 3) Selanjutnya langkah-langkah yang dilakukan adalah menyusun data-data yang akan diolah termasuk parameter dan alternatif yang akan dirangking, merancang dan memproses perhitungan menggunakan metode K-Means untuk mendapatkan nilai pengelompokan mahasiswa berdasarkan Nila Rata-rata UN dan IPK.
- 4) Hasil proses yang telah dinyatakan valid selanjutnya dilakukan analisis yang dikaitkan dengan kelompok dari asal sekolahnya.

## HASIL DAN PEMBAHASAN

Setelah dilakukan pengumpulan data dari data-data mahasiswa dengan tahun masuk tahun akademik 2012/2013, 2013/2014, dan 2014/2015 maka diperoleh data seperti dalam tabel berikut ini:

Tabel 4.1. Data Inisial

NO	NO Formulir	NIM	...	Asal Sekolah	NEM	IPK
1	011401001	011401503125001		SMK Darus Salam	5,83	2,96
2	011401045	011401503125002		SMKN 1 Geneng	7,73	3,18
3	011401016	011401503125004		SMK Nusantara	8,27	3,05
4	011401011	011401503125005		SMK Jakarta 1	7,22	3,14
5	011401085	011401503125006		SMK Yadika 2	7,38	2,30
6	011401089	011401503125007		SMK Triarga 2	7,30	1,74
7	011401103	011401503125008		SMKN 1 Ketapang	7,35	2,96
8	011401073	011401503125009		SMKN 59	8,10	2,54
9	011401006	011401503125012		SMK K.Bayangkara	5,17	2,84

10	011401137	011401503125014		SMU Al-Fajar	8,62	3,27
11	011401001	011401503125018		SMA N 24 Jakarta	6,73	3,16
12	011401200	011401503125021		SMA N 1 Pondok Aren	8,37	2,83
13	011401150	011401503125022		SMK Yadika 4	4,22	2,70
14	011401219	011401503125023		SMKN 3 Bandar Lam	7,18	3,23
15	011401227	011401503125024		SMK PGRI 25	4,88	2,83
16	011401228	011401503125025		SMK Muhammadiyah 5	5,07	2,27
17	011401274	011401503125026		SMK N 2 Klaten	8,25	2,27
18	011401204	011401503125033		SMK Bina Siswa	7,29	2,66
19	011401214	011401503125034		SMKN Jenawi	8,16	2,99
20	011401246	011401503125035		SMK N 18 Jakarta	7,79	2,85
21	011401316	011401503125037		SMK 4 Yadika 4	6,83	2,10
:	:	:	:	:	:	:
:	:	:	:	:	:	:
240	11202246	11205503215047		SMK Sasmita Jaya 2	6,90	2,31

Setelah mengumpulkan data maka dilanjutkan dengan pemilihan atribut yang dibutuhkan untuk proses Algoritma K-Means.

Tabel 4.2 Data Hasil seleksi

NO	NEM	IPK
1	5,83	2,96
2	7,73	3,18
3	8,27	3,05
4	7,22	3,14
5	7,38	2,30
6	7,30	1,74
7	7,35	2,96
8	8,10	2,54
9	5,17	2,84
10	8,62	3,27
11	6,73	3,16
12	8,37	2,83
13	4,22	2,70
14	7,18	3,23
15	4,88	2,83
16	5,07	2,27
17	8,25	2,27
18	7,29	2,66
19	8,16	2,99
20	7,79	2,85

21	6,83	2,10
⋮	⋮	⋮
240	6,90	2,31

Setelah data hasil seleksi ini diperoleh, maka dilanjutkan dengan pengolahan data. Data diolah dengan menggunakan SPSS versi 21 diperoleh hasil sebagai berikut

**Tabel 4.3 Initial Cluster Centers**

	Cluster		
	1	2	3
NEM	3,37	6,22	9,23
IPK	2,12	3,35	2,21

**Tabel 4.4 Iteration History<sup>a</sup>**

Iteration	Change in Cluster Centers		
	1	2	3
1	,947	,685	1,119
2	,740	,053	,162
3	,278	,099	,029
4	,182	,132	,023
5	,096	,093	,028
6	,045	,058	,028
7	,014	,022	,014
8	,013	,022	,013
9	,000	,007	,007
10	,000	,000	,000

a. Convergence achieved due to no or small change in cluster centers. The maximum absolute coordinate change for any center is ,000. The current iteration is 10. The minimum distance between initial centers is 3,104.

**Tabel 4.6 Final Cluster Centers**

	Cluster		
	1	2	3
NEM	5,52	7,03	8,22
IPK	2,61	2,77	2,93

**Tabel 4.8 Number of Cases  
in each Cluster**

	1	58,000
Cluster	2	92,000
	3	90,000
Valid		240,000
Missing		,000

Dari data-data di atas diperoleh deskripsi data sebagai berikut:

**Tabel 4.9 Deskripsi Data Per Cluster**

	N.MIN	IPK.MIN	N.MAKS	IPK. MAKS	RATA N	RATA2IPK
C1	2,85	0,95	6,27	3,74	5,46	2,61
C2	6,22	0,94	7,67	3,49	7,02	2,77
C3	7,50	1,64	9,40	3,55	8,22	2,92

Sebagai sasaran PMB terbaik tentunya adalah calon mahasiswa yang mempunyai NEM terbaik dan IPK terbaik. Hal ini terdapat di Cluster 3 dengan nilai rata-rata NEM=8.22 dan rata-rata IPK 2,92.

Setelah dilakukan Query ternyata pada kluster 3 tidak ada asal sekolah yang mendominasi hanya beberapa siswa yang berasal dari sekolah yang sama. Dari SMK Media informatika 4 orang, SMK Budi Mulia 3 orang, SMK Triguna 2 orang, SMK Muhammadiyah 9 2orang, SMK YUPPEN EK1 3 orang, SMK N 15 2 orang, SMK Pertiwi 2 2 orang, SMK KARTIKA 2 orang, selainnya 1 orang.

## **KESIMPULAN DAN SARAN**

Dari hasil dan pembahasan pada penelitian ini dapat diperoleh:

1. Dari sampel yang diambil 240 mahasiswa, ada 58 orang ada di Cluster1, 92 orang ada di Cluster2, dan 90 orang ada di Cluster3;
2. Rata-rata NEM dan IPK pada Cluster 3 adalah 8,22 dan 2,92;
3. Dari 90 data Asal sekolah hanya 4 siswa yang berasal dari SMK Meidya Informatika, 3 siswa dari SMK Budi Mulya, 3 siswa dari SMK YUPPEN EK1, dan masing-masing 2 siswa dari SMK Triguna, SMK Muhammadiyah 9, SMK Pertiwi 2, SMK N 15 ,dan SMK KARTIKA; selainnya masing-masing1 siswa dari 70 sekolah;
4. Distribusi Asal sekolah sangat beragam (tidak ada dominasi);

Dari kesimpulan diatas penulis memberikan beberapa saran yang pastinya bermanfaat bagi instansi terkait :

Dari pengalaman penelitian tentang pengumpulan data mengalami hal yang cukup menyulitkan dikarenakan data base PMB kurang lengkap yaitu Nilai UAN(NEM) tidak, sehingga perlu data base PMB dilengkapi supaya untuk pengolahan data sesuai keperluan manajemen bisa lebih cepat. Diharapkan juga untuk penelitian berikut dilakukan untuk seluruh USNI.

## **DAFTAR PUSTAKA**

- [1] Lin, Ching-Teng; Lee, George. 1996. Neural Fuzzy Systems : A Neuro-Fuzzy Synergism to Intelligent Systems. United States of America. Prentice Hall International Inc.
- [2] Kusumadewi, Sri; Hartati, Sri. 2006. Neuro-Fuzzy : Integrasi Sistem Fuzzy dan Jaringan Syaraf. Yogyakarta. Graha Ilmu
- [3] Eko Prasetyo, 2012, Data Mining: Konsep dan Aplikasi menggunakan MATHLAB. Andi Jogjakarta
- [4] Irwan Budiman, Toni Prahasto, and Yuli Christyono, "DATA CLUSTERING MENGGUNAKAN METODOLOGI CRISP-DM UNTUK PENGENALAN POLA PROPORSI PELAKSANAAN TRIDHARMA," in Seminar Nasional Aplikasi Teknologi Informasi 2012 (SNATI 2012), Yogyakarta, 2012.
- [5] Suprihatin, "Klastering K-Means Untuk Penentuan Nilai Ujian," JUSI, vol. 1, no. 1, pp. ISSN: 2087-8737, Februari 2011.
- [6] Larose, Daniel T, Discovering Knowledge in Data: An Introduction to Data Mining.: John Willey & Sons. Inc, 2005.