

# PENGGUNAAN METODE PENGKLASTERAN UNTUK MENENTUKAN BIDANG TUGAS AKHIR MAHASISWA TEKNIK INFORMATIKA PENS BERDASARKAN NILAI

Entin Martiana S.Kom,M.Kom, Nur Rosyid Mubtada'i S. Kom, Edi Purnomo  
Jurusan Teknik Informatika  
Politeknik Elektronika Negeri Surabaya  
Institut Teknologi Sepuluh Nopember  
Kampus ITS Keputih Sukolilo Surabaya 60111, Indonesia  
Tel:+62-31-5947280 Fax:+62-31-5946114  
Email: [entin@eepis-its.edu](mailto:entin@eepis-its.edu), [rosyid@eepis-its.edu](mailto:rosyid@eepis-its.edu), [edi\\_ephor@yahoo.co.id](mailto:edi_ephor@yahoo.co.id)

## Abstrak

*Tugas Akhir merupakan salah satu kewajiban mahasiswa, khususnya di PENS-ITS, yang harus dikerjakan sebagai syarat kelulusan. Namun tidak jarang mahasiswa mengalami kesulitan dalam penyelesaian tugas akhir tersebut. Salah satu penyebabnya adalah faktor kemampuan mahasiswa dalam bidang Tugas Akhir yang dikerjakannya.*

*Penelitian ini membahas penggunaan metode clustering dan inner product untuk menentukan bidang Tugas Akhir mahasiswa Teknik Informatika PENS-ITS berdasarkan nilai yang didapat mulai dari semester pertama sampai dengan semester sebelum penentuan judul TA. Tiap bidang disusun oleh mata kuliah-mata kuliah tertentu. Nilai mata kuliah-mata kuliah tersebutlah yang digunakan sebagai atribut data dalam sistem ini.*

*Metode clustering yang digunakan adalah Single Linkage Hierarchical, Centroid Linkage Hierarchical, dan K-Means. Metode-metode clustering tersebut digunakan untuk melakukan training data sehingga terbentuk cluster-cluster. Cluster-cluster yang terbentuk kemudian dilabelkan dengan Inner Product. Inner Product dilakukan dengan mengalikan centroid tiap cluster dengan nilai minimum (dari data training) untuk atribut centroid (mata kuliah) yang tidak mempengaruhi bidang TA dan mengalikan dengan nilai maximum (dari data training) untuk atribut centroid (mata kuliah) yang mempengaruhi bidang TA yang sedang diproses. Hasil Inner Product yang paling besar menunjukkan bahwa cluster tersebut memiliki label bidang TA yang sedang diproses.*

*Pengujian dilakukan dengan data baru (data uji) yang memiliki atribut (mata kuliah) yang sama dengan data training. Data uji tersebut dihitung jaraknya menggunakan Euclidean Distance dengan masing-masing cluster yang telah berlabel (bidang TA). Jarak yang terdekat menunjukkan data tersebut merupakan anggota cluster yang dimaksud yang berarti data baru tersebut termasuk ke bidang yang diwakili cluster yang berjarak paling dekat tersebut.*

*Hasil percobaan menunjukkan bahwa metode centroid linkage memiliki nilai variance ( $V_w/V_b$ ) paling kecil yang menandakan bahwa metode tersebut menghasilkan cluster paling baik dibanding dengan metode single linkage dan K-Means pada kasus ini.*

**Kata kunci :** *Hierarchical, Clustering, Centroid, Single, Linkage, K-Means, variance, Euclidean Distance, Inner Product, Bidang Tugas Akhir.*

## 1. PENDAHULUAN

Sekarang ini sudah terbukti peran mahasiswa PENS-ITS dalam kemajuan teknologi bangsa sangatlah besar. Banyak mahasiswa PENS-ITS yang telah berhasil menemukan dan membuat teknologi tepat guna yang dapat langsung dipakai di masyarakat maupun di lingkungan yang lain seperti industri, akademis, dan lain-lain. Banyak dari mereka yang membuat teknologi tepat guna tersebut ketika masih menyandang status sebagai mahasiswa, terutama mahasiswa semester akhir. Mereka yang sudah berada pada semester akhir membuat Tugas Akhir baik yang berupa analisa maupun aplikasi yang sangat berguna. Namun sayang sekali masih

banyak mahasiswa yang mengalami perpanjangan masa studi mereka karena kesulitan dalam tugas akhirnya.

Ada beberapa hal yang dapat membuat mahasiswa merasa sulit dalam pengerjaan Tugas Akhir tersebut, diantaranya adalah faktor kemampuan dan minat mahasiswa. Tidak jarang mahasiswa PENS-ITS merasa kebingungan dengan bidang Tugas Akhir yang akan mereka ambil karena merasa ragu untuk mengambil suatu bidang yang ada dengan alasan merasa tidak mampu. Oleh karena itulah perlu dibuat suatu aplikasi yang dapat membantu para mahasiswa PENS-ITS, khususnya mahasiswa jurusan Teknik Informatika untuk menentukan bidang tugas

akhir yang sebaiknya mereka ambil berdasarkan nilai-nilai mata kuliah yang mereka miliki.

Aplikasi yang dibuat dalam penelitian ini menggunakan metode clustering. Hal ini dikarenakan metode clustering telah terbukti dapat digunakan untuk menyelesaikan persoalan seperti ini. Seperti yang pernah dilakukan oleh Mulyoto – jurusan Matematika ITS dalam thesisnya yang telah menggunakan metode clustering Fuzzy C-Means dan JST Kohonen-SOM untuk membuat aplikasi penjurusan siswa SMA.

Permasalahan yang dihadapi dalam menyelesaikan penelitian ini adalah:

1. Bagaimana menentukan nilai-nilai mahasiswa Teknik Informatika yang akan digunakan untuk data input.
2. Bagaimana melakukan clustering terhadap nilai-nilai mahasiswa Teknik Informatika.
3. Bagaimana menentukan label dari cluster yang didapat.
4. Bagaimana menentukan minat bidang tugas akhir yang sebaiknya diambil.
5. Bagaimanakah hasil dari penentuan bidang untuk masing-masing metode clustering.

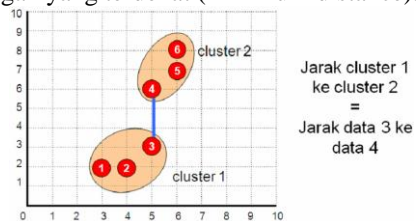
## 2. TEORI PENUNJANG

### 2.1. CLUSTERING

Clustering merupakan proses mengelompokkan suatu set obyek menjadi kelas-kelas yang terdiri dari obyek-obyek yang sama. Dengan kata lain clustering adalah proses mengelompokkan obyek-obyek berdasarkan kesamaan karakteristik diantara obyek-obyek tersebut.

#### 2.1.1. SINGLE LINKAGE

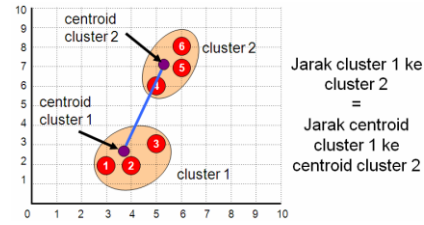
Pada algoritma single linkage, dilakukan pengukuran dan pencarian jarak dengan yang terdekat (minimum distance).



Gambar 1: Ilustrasi Single Linkage

#### 2.1.2. CENTROID LINKAGE

Centroid Linkage adalah proses pengclustering yang didasarkan pada jarak antar centroidnya. Metode ini bagus untuk memperkecil variance within cluster karena melibatkan centroid pada saat penggabungan antar cluster. Metode ini juga baik untuk data yang mengandung outlier.



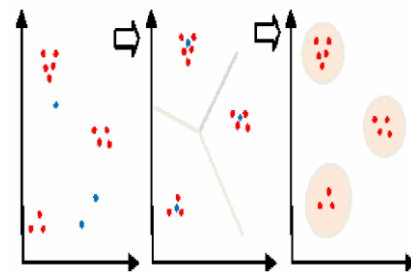
gambar 2: Ilustrasi Centroid Linkage

### 2.1.3. K-MEANS

Algoritma K-Means adalah metode clustering berbasis jarak yang membagi data ke dalam sejumlah cluster dan algoritma ini hanya bekerja pada atribut numerik.

Algoritma ini pada awalnya mengambil sebagian dari banyaknya komponen dari populasi yang akan dijadikan sebagai pusat cluster awal. Pada step ini pusat cluster dipilih secara acak dari sekumpulan populasi data.

Berikutnya K-Means menguji masing-masing komponen di dalam populasi data dan menandai komponen tersebut ke salah satu pusat. Cluster yang telah didefinisikan tergantung dari jarak minimum antar komponen dengan tiap-tiap pusat cluster. Posisi pusat cluster ke dalam tiap-tiap pusat cluster dan terakhir akan terbentuk pusat cluster baru.



Gambar 3: Ilustrasi K-Means

### 2.2. EUCLIDEAN DISTANCE

Euclidean distance merupakan salah satu metode penghitungan jarak yang paling sederhana. Jika terdapat n buah variabel maka perhitungan jarak menggunakan metode Euclidean dinyatakan sebagai berikut:

$$EUCLID(x,y) = \sqrt{\sum_i (x_i - y_i)^2}$$

dimana x dan y merupakan dua objek yang dihitung jaraknya,  $x_1, x_2, \dots, x_n$  dan  $y_1, y_2, \dots, y_n$  merupakan atribut-atribut sebanyak n buah dari objek x dan y.

### 2.3. INNER PRODUCT

Inner product yang merupakan istilah lain dari dot product atau dapat juga disebut sebagai scalar product adalah perkalian antara dua vektor yang dapat berupa vektor baris atau vektor kolom, sehingga hasilnya berupa scalar.

Rumus dasar inner product adalah sebagai berikut:

$$x \cdot y = \sum_{i=1}^n x_i \cdot y_i$$

$$a \cdot b = \sum_{i=1}^n a_i b_i = a_1 b_1 + a_2 b_2 + \dots + a_n b_n$$

Contoh perhitungan dari persamaan di atas dapat dilihat pada contoh berikut:

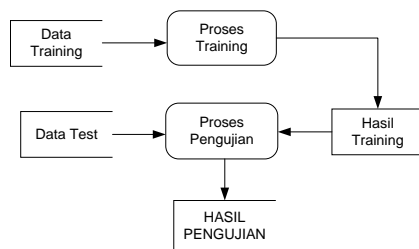
$$[3 \ 1 \ 4] \cdot [2 \ 5 \ 1] = (3)(2) + (1)(5) + (4)(1) = 15$$

$$[1 \ 3 \ 2] \cdot \begin{bmatrix} 3 \\ 5 \\ 2 \end{bmatrix} = (1)(3) + (3)(5) + (2)(2) = 22$$

### 3. PERENCANAAN SISTEM

Dalam sistem ini terdapat tahapan standar, yaitu: input, proses, dan output.

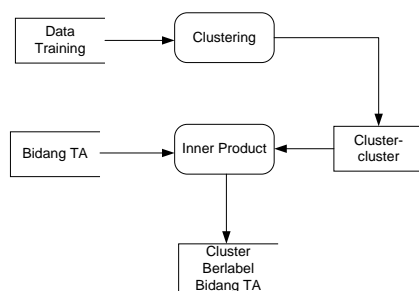
Secara garis besar dapat digambarkan dengan blok diagram sebagai berikut:



Gambar 4: Diagram Sistem

Input yang dimasukkan pada sistem ini adalah nilai-nilai mahasiswa selama kuliah sebelum pengambilan judul Tugas Akhir.

Tahapan proses terdapat dua macam proses yaitu proses training dan proses testing (pengujian).

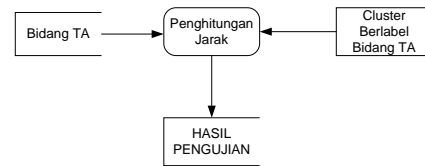


Gambar 5: Proses Training

Proses training dilakukan dengan mencluster data training dengan metode clustering centroid linkage hierarchical, single linkage hierarchical, dan k-means. Cluster yang terbentuk kemudian dilabelkan dengan nama bidang tugas akhir yang ada. Cara pelabelan yang digunakan adalah dengan inner product. Hasil inner product terbesar antara suatu cluster dengan suatu bidang menunjukkan bahwa label cluster tersebut adalah nama bidang yang di-inner product-kan.

Proses testing dilakukan dengan menggunakan inner product antara data test

dengan cluster-cluster yang berlabel. Sama seperti proses training, hasil inner product yang paling besar menunjukkan data test tersebut memasuki cluster (bidang) yang telah diproses.



Gambar 6: Proses Testing

Output yang dihasilkan dari sistem ini adalah bidang tugas akhir yang sesuai dengan data (mahasiswa) yang dijadikan sebagai data test.

### 4. UJI COBA

Uji coba dilakukan dengan menggunakan data nilai mahasiswa D3 angkatan 2005, nilai mahasiswa D4 angkatan 2005, dan nilai mahasiswa D3 angkatan 2006. Langkah pertama uji coba adalah melakukan proses training. Dari proses training tersebut diperoleh hasil (*variance*) sebagai berikut.

Tabel 1: Hasil Training

Data	Metode	Variance(Vw/Vb)
Nilai mahasiswa D3 angkatan 2005	Single Linkage	4.2017580184379966E-18
	Centroid Linkage	3.991511126649698E-18
	K-Means	0.013409520461774616
Nilai mahasiswa D4 angkatan 2005	Single Linkage	6.499494834298336E-18
	Centroid Linkage	5.265828882966592E-18
	K-Means	0.07738628536451325
Nilai mahasiswa D3 angkatan 2006	Single Linkage	5.0296972276279414E-18
	Centroid Linkage	3.720404007614485E-18
	K-Means	0.02686628144403626

Dari hasil training didapatkan nilai *variance* yang paling optimal adalah pada metode *centroid linkage*. Sehingga metode *centroid linkage* merupakan metode yang menghasilkan *cluster* paling ideal pada kasus ini.

Setelah dilakukan proses training, proses selanjutnya adalah pengujian. Proses pengujian dilakukan untuk mengetahui bidang yang seharusnya diambil oleh data uji berdasarkan training yang dilakukan. Berikut hasil pengujian yang telah dilakukan.

- Data uji: nilai mahasiswa D3 angkatan 2005 dengan NRP 7405030041

Tabel 2: Hasil Uji Coba 1

DATA	METODE TRAINING	BIDANG TA	PERSENTASE JARAK
7405030041	SINGLE LINKAGE	PEMROGRAMAN	22.585166942124278%
		KECERDASAN KOMPUTASIONAL	21.531039962562158%
		BASIS DATA DAN SISTEM INFORMASI	20.63466820425035%
		JARINGAN KOMPUTER	18.303084878206125%
		GRAFIS DAN CITRA	16.94604001285709%
		METODE TRAINING	BIDANG TA
	CENTROID LINKAGE	PEMROGRAMAN	22.564046051659123%
		KECERDASAN KOMPUTASIONAL	21.486314757117935%
		BASIS DATA DAN SISTEM INFORMASI	20.890696923369347%
		JARINGAN KOMPUTER	18.21674173748607%
		GRAFIS DAN CITRA	16.842200530367528%
		METODE TRAINING	BIDANG TA
	KMEANS	BASIS DATA DAN SISTEM INFORMASI	43.08528269712252%
		PEMROGRAMAN	30.439229598906074%
		KECERDASAN KOMPUTASIONAL	26.4754877039714%
		JARINGAN KOMPUTER	-
		GRAFIS DAN CITRA	-

- Data uji: nilai mahasiswa D4 angkatan 2005 dengan NRP 7405040046

Tabel 3: Hasil Uji Coba 2

DATA	METODE TRAINING	BIDANG TA	PERSENTASE JARAK
7405040046	SINGLE LINKAGE	GRAFIS AND CITRA	21.28880%
		KECERDASAN KOMPUTASIONAL	20.86355%
		JARINGAN KOMPUTER	19.50891%
		DATABASE AND SISTEM INFORMASI	19.39374%
		PEMROGRAMAN	18.94501%
	METODE TRAINING	BIDANG TA	PERSENTASE JARAK
	CENTROID LINKAGE	KECERDASAN KOMPUTASIONAL	21.27241%
		GRAFIS AND CITRA	20.92956%
		PEMROGRAMAN	20.84420%
		JARINGAN KOMPUTER	19.48323%
		DATABASE AND SISTEM INFORMASI	17.47060%
	METODE TRAINING	BIDANG TA	PERSENTASE JARAK
	KMEANS	KECERDASAN KOMPUTASIONAL	21.25102%
		GRAFIS AND CITRA	20.74937%
		PEMROGRAMAN	20.10002%
		DATABASE AND SISTEM INFORMASI	19.40424%
		JARINGAN KOMPUTER	18.49535%

- Data uji: nilai mahasiswa D3 angkatan 2006 dengan NRP 7406030071

Tabel 4: Hasil Uji Coba 3

DATA	METODE TRAINING	BIDANG TA	PERSENTASE JARAK
7406030071	SINGLE LINKAGE	BASIS DATA DAN SISTEM INFORMASI	21.34397%
		JARINGAN KOMPUTER	20.07383%
		PEMROGRAMAN	19.85669%
		KECERDASAN KOMPUTASIONAL	19.71591%
		GRAFIS DAN CITRA	19.00960%
		METODE TRAINING	BIDANG TA
	CENTROID LINKAGE	BASIS DATA DAN SISTEM INFORMASI	21.33103%
		KECERDASAN KOMPUTASIONAL	20.09147%
		PEMROGRAMAN	19.89190%
		JARINGAN KOMPUTER	19.69720%
		GRAFIS DAN CITRA	18.98839%
	METODE TRAINING	BIDANG TA	PERSENTASE JARAK
	KMEANS	BASIS DATA DAN SISTEM INFORMASI	36.64076%
		JARINGAN KOMPUTER	31.82588%
		PEMROGRAMAN	31.53336%
		KECERDASAN KOMPUTASIONAL	-
		GRAFIS DAN CITRA	-

Dari percobaan di atas, masing-masing metode menghasilkan *output* yang berbeda untuk data uji yang sama. Dalam artian setiap metode akan menghasilkan nilai perentase jarak yang berbeda untuk masing-masing data. Hal ini dipengaruhi oleh letak *centroid cluster* (dalam hal ini bidang TA) hasil training yang berbeda-beda. Untuk metode *hierarchical*, *centroid cluster* didapat dari rata-rata data. *Centroid* didapat dengan hanya menggunakan satu kali percobaan. Karena metode ini memiliki sifat iteratif pada waktu pembentukan *cluster*. Sehingga berapapun jumlah percobaannya jika menggunakan data yang sama akan menghasilkan hasil yang sama juga. Namun untuk mendapatkan *centroid* di metode *k-means* diperlukan beberapa iterasi sampai terbentuk *centroid* akhir.

Setiap metode memiliki perbedaan *variance* optimal yang berbeda. *Variance* pada metode *hierarchical* akan selalu tetap. Metode ini cenderung memiliki *variance* yang lebih kecil dibanding dengan metode *K-Means* pada percobaan-percobaan yang dilakukan. Sedangkan *variance* pada cluster *K-Means* dapat berubah-ubah. Hal ini dikarenakan pembangkitan *centroid* awal yang *random* membuat *cluster K-Means* memiliki hasil *cluster* yang berbeda untuk setiap proses *peng-cluster-an*.

Hasil *cluster* pada percobaan menunjukkan data mengumpul pada suatu *cluster*. Hal ini dapat dikarenakan metode *clustering* yang digunakan tidak dapat mengenali perbedaan atribut untuk menghitung

suatu jarak. Jadi jika suatu jarak minimal didapat dari penghitungan jarak dari data atribut suatu bidang memiliki jarak yang sama dekatnya dengan jarak minimal yang didapat dari perhitungan data dari atribut bidang yang lain, maka kedua data akan dijadikan satu *cluster*. Padahal seharusnya kedua data tersebut berada dalam *cluster* yang berbeda.

## 5. KESIMPULAN DAN SARAN

### 5.1 KESIMPULAN

Setelah dilakukan pengujian dan analisa program, maka dapat diperoleh kesimpulan sebagai berikut :

- Pada kasus ini metode clustering yang dipakai memiliki hasil yang berbeda. Namun jika dilihat dari nilai variance yang didapat, metode centroid linkage hierarchical lah yang memiliki nilai variance yang paling optimal.
- Program yang dibuat berhasil menentukan data uji untuk memilih bidang apa seharusnya dia ambil. Jika dilihat pada hasil training yang dilakukan, penyebaran data pada masing-masing cluster memiliki kemiripan.

### 5.2 SARAN

Dari beberapa kesimpulan yang telah diambil, maka dapat dikemukakan saran-saran yang akan sangat membantu untuk pengembangan aplikasi ini selanjutnya.

- Untuk mendapatkan keputusan dengan pertimbangan yang beragam sebaiknya selain menggunakan metode-metode tersebut digunakan juga metode yang lain.
- Untuk mendapatkan hasil yang optimal pada metode hierarchical diperlukan suatu metode untuk mengenali atribut-atribut yang dihitung jaraknya. Sehingga data yang diproses tidak mengumpul pada cluster tertentu saja.

## DAFTAR PUSTAKA

- [1]. Mulyoto.2007.*Metode Clustering JST Kohonen-SOM dan Fuzzy C-Means: Aplikasinya pada Jurusan Siswa SMA.*
- [2]. Fadli M. Qoyyim. 2007. *Pemilihan warna cat sesuai kepribadian dengan inner product.*
- [3]. Apostol, Tom.M.1967.*CALCULUS Volume 1 Second Edition.* New York: John Wiley & Sons, Inc.
- [4]. Anderberg, M. R. 1973.*Cluster analysis for applications.* New York: Academic Press.
- [5]. Barakbah A.R., *Clustering, In Workshop Data Mining 2006,* Jurusan Teknologi Informasi Politeknik Elektronika Negeri Surabaya, ITS.
- [6]. Barakbah A.R., *Cluster Analysis,* Jurusan Teknologi Informasi Politeknik Elektronika Negeri Surabaya, ITS, 2006.
- [7]. Huang, Z., 2005, *A Fast Clustering Algorithm to Cluster Very Large Categorical Data Sets in Data Mining.* CSIRO Mathematical and Information Sciences – Australia.
- [8]. Hasniawati, Helmi, 2007, *Image Clustering Berdasarkan Warna Untuk Identifikasi Buah Dengan Metode Valley Tracing,* Jurusan Teknologi Informasi Politeknik Elektronika Negeri Surabaya,ITS.
- [9]. Wiry Wardhana, Gede, 2007, *Image Clustering Berdasarkan Warna Untuk Identifikasi Jenis Buah Dengan Metode Hill Climbing.* Jurusan Teknologi Informasi Politeknik Elektronika Negeri Surabaya,ITS.