ARTIKEL

APLIKASI PENENTUAN KONSENTRASI JURUSAN MAHASISWA TEKNIK INFORMATIKA UNIVERSITAS NUSANTARA PGRI KEDIRI MENGGUNAKAN METODE *K-MEANS CLUSTERING*

Oleh:

SITI UMAROH

13.1.03.02.0328

Dibimbing oleh :

1. **Intan Nur Farida, M.Kom.**
2. **Danar Putra Pamungkas, M.Kom.**

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS NUSANTARA PGRI KEDIRI

2017

SURAT PERNYATAAN
ARTIKEL SKRIPSI TAHUN 2017

Yang bertanda tangan di bawah ini:

Nama Lengkap	: Siti Umaroh
NPM	: 13.1.03.02.0328
Telepun/HP	: 085741075861
Alamat Surel (Email)	: sitiumaroh1801@gmail.com
Judul Artikel	: APLIKASI PENENTUAN KONSENTRASI JURUSAN MAHASISWA TEKNIK INFORMATIKA UNIVERSITAS NUSANTARA PGRI KEDIRI MENGGUNAKAN METODE *K-MEANS CLUSTERING*
Fakultas – Program Studi	: Fakultas Teknik – Teknik Informatika
Nama Perguruan Tinggi	: Universitas Nusantara PGRI Kediri
Alamat Perguruan Tinggi	: Jl. KH. Achmad Dahlan 76 Kediri

Dengan ini menyatakan bahwa :

a. artikel yang saya tulis merupakan karya saya pribadi (bersama tim penulis) dan bebas plagiarisme;
b. artikel telah diteliti dan disetujui untuk diterbitkan oleh Dosen Pembimbing I dan II.

Demikian surat pernyataan ini saya buat dengan sesungguhnya. Apabila di kemudian hari ditemukan ketidaksesuaian data dengan pernyataan ini dan atau ada tuntutan dari pihak lain, saya bersedia bertanggungjawab dan diproses sesuai dengan ketentuan yang berlaku.

Mengetahui		Kediri, 9 Agustus 2017
Pembimbing I	Pembimbing II	Penulis,
Intan Nur Farida, M.Kom NIDN. 0704108701	Danar Putra Pamungkas, M.Kom NIDN. 0708028704	Siti Umaroh NPM. 13.1.03.02.0328

APLIKASI PENENTUAN KONSENTRASI JURUSAN MAHASISWA TEKNIK INFORMATIKA UNIVERSITAS NUSANTARA PGRI KEDIRI MENGGUNAKAN METODE *K-MEANS CLUSTERING*

Siti Umaroh
13.1.03.02.0328
Fakultas Teknik – Teknik Informatika
Sitiumaroh1801@gmail.com
Intan Nur Farida, M.Kom. dan Danar Putra Pamungkas, M.Kom
UNIVERSITAS NUSANTARA PGRI KEDIRI

ABSTRAK

Penelitian ini dilatar belakangi oleh tidak adanya konsentrasi jurusan yang ada di program studi Teknik Informatika menyebabkan mahasiswa tidak mengetahui kemampuan yang dimiliki. Hal ini menyulitkan mahasiswa dalam penyerapan mata kuliah dan pengambilan tema untuk judul skripsi.

Permasalahan dalam penelitian ini adalah: bagaimana merancang program aplikasi penentuan konsentrasi jurusan mahasiswa Program Studi Teknik Informatika Universitas Nusantara PGRI Kediri dengan metode *K-Means*. Dan juga bagaimana mengimplementasikan data mining dengan metode *K-Means* dalam program aplikasi sehingga mampu menentukan konsentrasi jurusan mahasiswa.

Penelitian ini menggunakan pendekatan penelitian kuantitatif dikarenakan data nilai mahasiswa, diolah dengan menggunakan metode *K-Means Clustering* untuk mendapatkan hasil pengelompokan konsentrasi jurusan yang akan dijadikan pertimbangan mahasiswa dalam penentuan tema judul skripsi. Data nilai yang digunakan adalah data nilai mahasiswa angkatan 2013 dari semester 1 sampai semester 4. Adapun konsentrasi jurusan yang ada yaitu Sistem Informasi dan Enterprise, Rekayasa Perangkat Lunak dan Animasi, Jaringan Komputer dan Multimedia.

Kesimpulan hasil penelitian ini adalah : Menghasilkan sistem penentuan konsentrasi jurusan dengan menggunakan metode *K-Means Clustering* sesuai dengan kriteria pengelompokan. Menghasilkan aplikasi untuk mengelompokkan mahasiswa berdasarkan konsentrasi jurusannya.

Saran yang diberikan yaitu pengembangan penelitian setelahnya dengan metode *clustering* yang lain. Perbaikan pada desain interface untuk memudahkan user dalam mengoperasikan sistem. Dan juga penambahan angket minat siswa untuk menambah keakuratan data.

Kata kunci : konsentrasi jurusan, *K-Means*, *Clustering*.

I. LATAR BELAKANG

Universitas Nusantara PGRI Kediri merupakan satu dari beberapa instansi perguruan tinggi yang bergerak di bidang pendidikan. Di dalam Universitas Nusantara PGRI Kediri terdapat 5 Fakultas, dan 22 Program Studi. Teknik Informatika merupakan salah satu dari sekian program studi yang ada di Universitas Nusantara PGRI Kediri.

Penentuan pemilihan konsentrasi bukanlah hal yang mudah dan akan sangat berpengaruh terhadap kegiatan akademik mahasiswa. Keputusan penentuan konsentrasi harus dilakukan dengan pertimbangan yang matang agar mahasiswa tidak salah dalam memilih konsentrasi jurusan yang diinginkan. Banyak mahasiswa yang belum mengetahui bakat, minat dan kemampuan yang dimilikinya.

Tidak adanya konsentrasi jurusan yang ada di Prodi Teknik Informatika Universitas Nusantara PGRI Kediri menyulitkan mahasiswa untuk mengetahui kemampuan akademik yang dimiliki. Hal itu menyebabkan mahasiswa merasa kesulitan dalam pengambilan tema untuk judul skripsi.

Pemilihan konsentrasi jurusan tanpa adanya pertimbangan yang matang sesuai dengan kemampuan akademik yang dimiliki akan berdampak pada mahasiswa itu sendiri yang merasa salah mengambil jurusan dan menyebabkan kesulitan dalam penyerapan materi-materi perkuliahan sesuai dengan konsentrasi jurusan yang dipilih.

Dalam penelitian ini konsentrasi pilihan jurusan dibagi menjadi tiga konsentrasi yaitu Sistem Informasi dan Enterprise, Jaringan Komputer dan Multimedia, serta Rekayasa Perangkat Lunak dan Animasi. Dengan adanya konsentrasi penjurusan yang ada mahasiswa diwajibkan untuk memilih salah satu dari konsentrasi jurusan yang ada. Penentuan konsentrasi ini akan dilaksanakan pada akhir semester 4 atau tepatnya masa perkuliahan untuk semester 5.

II. METODE

1. Algoritma *K-Means*

Menurut Larose (2005) definisi dari algoritma *K-Means* adalah :

Algoritma *K-Means* merupakan algoritma sederhana dan efektif untuk menemukan pengelompokan dalam data. Algoritma *K-Means* digunakan untuk mengelompokan data sesuai dengan kemiripan dari data data yang ada. Proses dari *K-Means* adalah sebagai berikut :

a. Langkah pertama : Menentukan banyak *cluster* k yang ada dalam data set yang harus dibagi.
b. Langkah kedua : menetapkan secara acak/random *records* k untuk dijadikan lokasi pusat awal *cluster*.
c. Langkah ketiga : untuk setiap *records*, temukan pusat *cluster* terdekat. Pusat *cluster* diperolah dengan menghitung jarak dengan menggunakan rumus *Euclidean distance*. Dengan demikian pusat *cluster* memiliki *subset* dari *record*, sehingga mewakili partisi dari data set. Oleh karena itu kita memiliki kelompok k, C1,C2,…,Cn.
d. Langkah keempat : untuk masing-masing *cluster* k, temukan *cluster centroid*, dan perbarui lokasi masing-masing pusat *cluster* dengan nilai dari *centroid* baru.
e. Langkah kelima : ulangi langkah 3-5 sampai ditemukan hasil. Hasil diperoleh apabila kelompok *cluster* sesudah dan kelompok *cluster* sebelum sudah sama.

2. Simulasi Perhitungan

Simulasi perhitungan dengan metode *K-Means* dengan 15 data mahasiswa adalah sebagai berikut :

Tabel 1 Dataset mahasiswa

No	NPM	X1	X2	X3
1	13.1.03.02.0097	3.27	3.38	3.6
2	13.1.03.02.0333	3.33	3.05	3.35
3	13.1.03.02.0212	3.16	3.33	3.15
4	13.1.03.02.0025	3.27	3.16	3.4
5	13.1.03.02.0359	3.16	3	3.2
6	13.1.03.02.0370	3.38	3.33	3.6
7	13.1.03.02.0210	3.5	3.5	3.85
8	13.1.03.02.0392	3.33	3.27	3.3
9	13.1.03.02.0350	3.33	3.38	3.2
10	13.1.03.02.0128	3.33	3.16	3.15
11	13.1.03.02.0328	3.33	3.33	3.65
12	13.1.03.02.0389	3.16	3.05	3.25
13	13.1.03.02.0362	3.27	2.94	3.3
14	13.1.03.02.0032	3.22	3.33	3.25
15	13.1.03.02.0084	3.22	3.27	3.1

Keterangan :

X1= rata-rata nilai mata kuliah untuk masuk konsentrasi jurusan sistem informasi dan enterprise.

X2= rata-rata nilai mata kuliah untuk masuk konsentrasi jurusan jaringan komputer dan multimedia.

X3= rata-rata nilai mata kuliah untuk masuk konsentrasi jurusan rekayasa perangkat lunak dan animasi.

Dengan pembagian mata kuliahnya sebagai berikut :

Tabel 2 mata kuliah yang diambil

Vr	Mata Kuliah	Konsentrasi Jurusan
A1	Algoritma Pemrograman 1	Sistem Informasi dan Enterprise (X1)
A2	Praktikum Algoritma Pemrograman 1	
A3	Algoritma Pemrograman 2	
A4	Praktikum Algoritma Pemrograman 2	
A5	Sistem Informasi	
A6	Basis data	
A7	Praktikum Basis data	
A8	Pemrograman Web	
A9	Praktikum Pemrograman Web	
A10	Pengantar Teknologi Informasi	Jaringan Komputer dan Multimedia (X2)
A11	Praktikum Pengantar Teknologi Informasi	
A12	Arsitektur Komputer	
A13	Sistem Operasi	
A14	Pemrograma Berorientasi Objek	
A15	Praktikum Pemrograman Berorientasi Objek	
A16	Dasar Multimedia	
A17	Jaringan Komputer	
A18	Praktikum Jaringan Komputer	
A19	Kalkulus	Rekayasa Perangkat Lunak dan Animasi (X3)
A20	Logika Matematika	
A21	Matematika Diskrit	
A22	Aljabar	
A23	Struktur Data	
A24	Praktikum Struktur Data	
A25	Statistik	
A26	Rekayasa Perangkat Lunak	
A27	Metode Numerik	
A28	Riset Operasi	

Tentukan centroid awal untuk proses perhitungannya. Centroid awal diambil dari dataset yang ada. Penentuan centroid awal ini adalah dengan mengambil dari dataset dengan acak/ random. Centroid awal dapat dilihat pada tabel 3:

Tabel 3 centroid awal

No	NPM	X1	X2	X3
3	13.1.03.02.0212	3.16	3.33	3.15
8	13.1.03.02.0392	3.33	3.27	3.3
13	13.1.03.02.0362	3.27	2.94	3.3

Kemudian dihitung jarak dengan menggunakan rumus *Euclidean Distance*. Setelah jarak ketemu dari 3 *cluste tersebut dicari jarak minimum.* Rumusnya sebagai berikut:

D (x,y) = |x-y| = $\sqrt{\sum_{t-1}^{n}(x1 - y1)2}$

Hasil perhitungan jarak dan pemilihan jarak dilihat pada tabel 4 :

Tabel 4 Perhitungan jarak

No	X1	X2	X3	C1	C2	C3	JT
1	3.27	3.38	3.6	0.46	0.32	0.53	0.32
2	3.33	3.05	3.35	0.38	0.22	0.13	0.13
3	3.16	3.33	3.15	0	0.23	0.43	0
4	3.27	3.16	3.4	0.32	0.15	0.24	0.16
5	3.16	3	3.2	0.33	0.33	0.15	0.16
6	3.38	3.33	3.6	0.5	0.31	0.50	0.31
7	3.5	3.5	3.85	0.79	0.61	0.81	0.66
8	3.33	3.27	3.3	0.23	0	0.33	0
9	3.33	3.38	3.2	0.18	0.14	0.45	0.14
10	3.33	3.16	3.15	0.23	0.18	0.27	0.18
11	3.33	3.33	3.65	0.52	0.35	0.52	0.35
12	3.16	3.05	3.25	0.29	0.28	0.16	0.16
13	3.27	2.94	3.3	0.43	0.33	0	0
14	3.22	3.33	3.25	0.11	0.13	0.39	0.11
15	3.22	3.27	3.1	0.09	0.22	0.39	0.09

Jarak terdekat menunjukkan bahwa data tersebut masuk dalam satu kelompok dengan pusat cluster terdekat. Berikut ini akan ditampilkan data matriks pengelompokan, nilai 1 berarti data tersebut berada dalam kelompok cluster sejenis.

Tabel 5 Kelompok cluster pertama

id	C1	C2	C3
1		1	
2			1
3	1		
4		1	
5			1
6		1	
7		1	
8		1	
9		1	
10		1	
11		1	
12			1
13			1
14	1		
15	1		

Setelah diketahui anggota tiap-tiap cluster, kemudian centroid baru akan dihitung berdasarkan data dari anggota tiap-tiap cluster berdasarkan rumus pusat anggota cluster. Sehingga diperoleh perhitungan sebagai berikut :

- Centroid-1 data ke-3,14,15

  Centroid-1 ke-1 = 3.16 + 3.22 + 3.22 / 3 = 3.21

  Centroid-1 ke-2 = 3.33 + 3.33 + 3.27 / 3 = 3.31

  Centroid-1 ke-3 = 3.15 + 3.25 + 3.1 /3 = 3.16
- Centroid-2 data ke-1,4,6,7,8,9,10,11

  Centroid-2 ke-1 = 3.27 + 3.27 + 3.38 + 3.5 + 3.33 + 3.33 + 3.33 + 3.33 / 8 = 3.34

  Centroid-2 ke-2 = 3.38 + 3.16 + 3.33 + 3.5 + 3.27 + 3.38 + 3.16 + 3.33 / 8 = 3.31

  Centroid-2 ke-3 = 3.6 + 3.4 + 3.6 + 3.85 + 3.3 + 3.2 + 3.15 + 3.65 / 8 = 3.46875
- Centroid-3 data ke-2,5,12,13

  Centroid-3 ke-1 = 3.33 + 3.16 + 3.16+ 3.27 / 4 = 3.23

  Centroid-3 ke-2 = 3.05 + 3 + 3.05 + 2.94 / 4 = 3.01

  Centroid-3 ke-3 = 3.35 + 3.2 + 3.25 + 3.3 / 4 = 3.275

Didapatkan centroid baru yang dapat dilihat pada tabel 6 :

Tabel 6 Centroid baru

No	X1	X2	X3
1	3.21	3.31	3.16
2	3.34	3.31	3.46
3	3.23	3.01	3.27

Hasil dari centroid baru pada tabel 6 akan digunakan untuk menghitung jarak dengan rumus *Euclidean Distance*. Dan mengulang perhitungan dengan algoritma *K-Means* hingga didapatkan kelompok cluster sesudah sama dengan kelompok cluster sebelum sama. Pada perhitungan yang sudah dilakukan cluster sesudah sama dengan cluster sebelum terjadi pada iterasi kelima dengan hasil sebagai berikut :

- Mahasiswa dengan data ke 3,8,9,10,14,15 masuk dalam cluster satu yang berarti mahasiswa tersebut cocok masuk konsentrasi jurusan sistem informasi dan enterprise.
- Mahasiswa dengan data ke 1,6,7,11 masuk dalam cluster dua yang berarti mahasiswa tersebut cocok masuk konsentrai jurusan jaringan komputer dan multimedia.
- Mahasiswa dengan data ke 2,4,5,12.13 masuk dalam cluster 3 yang berarti mahasiswa tersebut cocok masuk konsentrasi jurusan rekayasa perangkat lunak dan animasi.

III. HASIL DAN KESIMPULAN

A. Hasil

Pengujian dalam penelitian ini adalah pengujian *alpha* dimana pengujian *alpha* berdasar pada proses klasterisasi. Data yang digunakan dalam pengujian *K-Means clustering* adalah data nilai mahasiswa dari mata kuliah semester 1 – semester 4 angkatan tahun 2013.

Data yang digunakan ada 150 data. Skenario pengujian yang akan dilakukan ada 6 skenario, dan yang menjadi fokus dalam skenario adalah jumlah data, maksimum iterasi, *cluster* 1, *cluster* 2, *cluster* 3. Hasil skenario pengujian seperti berikut :

Tabel 7 Hasil Skenario Pengujian

No	Jum-lah data	Maksi-mum Iterasi	Jum-lah cen-troid	Clus-ter 1	Clus-ter 2	Clus-ter 3
1.	50	7	3	16	1	33
2.	50	17	4	1	14	35
3.	100	7	3	68	15	17
4.	100	16	4	20	73	7
5.	150	7	3	95	43	12
6.	150	19	4	94	12	44

Analisa pengujian *alpha* sebagai berikut :

Skenario pengujian 1-6 dilakukan untuk mengetahui pengaruh jumlah data, jumlah centroid dan maksimum iterasi terhadap hasil klasterisasi dan waktu yang dibutuhkan untuk mengolah data. Hasil pengujian yang diperoleh, bahwa jumlah data tidak berpengaruh terhadap jumlah iterasi yang dihasilkan namun berpengaruh terhadap waktu yang dibutuhkan dan keakuratan hasil klasterisasi. Semakin banyak data dan semakin banyak iterasi yang dihasilkan maka waktu yang dibutuhkan semakin banyak. Selain itu pengambilan *centroid* juga berpengaruh terhadap hasil dari proses klasterisasi. *Centroid* yang diambil secara random menyebabkan hasil selalu berubah ubah. Hal ini dikarenakan *centroid* yang diambil secara random tersebut tidak diketahui memiliki seberapa besar kemiripan jarak dengan cluster yang dipilih.

Dalam pengujian yang dihasilkan dengan jumlah data yang berbeda beda dan juga

jumlah centroid yang berbeda beda menunjukkan bahwa semakin banyak data dan semakin banyak centroid menghasilkan tingkat perbedaan dari hasil yang diperoleh semakin menipis. Hal ini dikarenakan semakin banyak data membuat data semakin stabil. Dari data pada tabel 5.22 menjelaskan bahwa semakin banyak data dan semakin banyak centroid yang dipilih menyebabkan hasil yang lebih optimum dengan sedikit perbedaan pada hasilnya dibandingkan dengan jumlah data yang lebih sedikit dan jumlah centroid yang lebih sedikit.

Dari 1-6 skenario pengujian yang sudah dilakukan hasil terbaik dari skenario tersebut didapatkan pada skenario keenam dengan prosentase 62% untuk cluster pertama, 8% untuk cluster kedua dan 30% untuk cluster ketiga. Dengan tingkat akurasi sebesar 62%. Dengan perhitungan sebagai berikut :

Akurasi $= \frac{x}{n}\ x\ 100$……..……. (1)

Keterangan :

x = banyak data yang tidak berubah

n = banyak data

Keterangan dari hasil akurasi adalah sebagai berikut :

Banyak data yang tidak berubah merupakan data yang pada saat dilakukan beberapa kali pengujian hasilnya tidak berubah. Hal ini terdapat pada skenario 6 dengan jumlah data 150 dan banyak data yang tidak berubah ada 94 data. Dan nilai n merupakan banyak data dengan skenario terbaik yang mana pada pengujian ini skenario terbaik didapatkan pada skenario 6 dengan banyak data sejumlah 150 data. Jadi hasil akurasi 62% tersebut didapatkan dari banyaknya data yang tidak berubah yaitu 94 data dibagi dengan banyak data yaitu 150 data dikali dengan prosentase 100%.

B. Kesimpulan

Berdasarkan pada analisis dan pembahasan di bab sebelumnya, maka dapat diambil kesimpulan bahwa :

1. Telah didapatkan hasil akhir dari perhitungan dengan menggunakan *algoritma K-Means Clustering* yaitu mengelompokkan konsentrasi jurusan mahasiswa berdasarkan nilai-nilai mata kuliah dengan centroid yang sudah ditentukan. Jumlah konsentrasi jurusan ada 3 yaitu Sistem Informasi dan Enterprise, Jaringan Komputer dan Multimedia, Rekayasa Perangkat Lunak dan Animasi. Tiap nilai mahasiswa yang masuk akan dihitung kemiripannya dengan centroid sesuai dengan kelompoknya. Setelah itu akan dihasilkan data mahasiswa yang

masuk kedalam 3 konsentrasi jurusan yang sudah ditentukan.

2. Telah dihasilkan aplikasi untuk mengelompokkan konsentrasi jurusan pada mahasiswa Teknik Informatika menggunakan metode *K-Means Clustering*. Hasil yang didapatkan adalah 62% masuk Sistem Informasi dan Enterprise, 8% masuk Jaringan Komputer dan Multimedia, 0.30% masuk Rekayasa Perangkat Lunak dan Animasi.

IV. DAFTAR PUSTAKA

Anita, Budi.H, Ema, U., dan Emha, T.H, 2013. Implementasi Metode Fuzzy C- Means dan Topsis dalam Membangun Sistem Pendukung Keputusan Penentuan Jurusan SMA. Jurnal DASI, 14 (2). (Online), tersedia: https://ojs.amikom.ac.id, diunduh 20 November 2016.

Asroni, Ronald, A., 2015. Penerapan Metode K-Means untuk Clustering Mahasiswa berdasarkan Nilai Akademik dengan Weka Interface Studi Kasus pada Jurusan Teknik Informatika UMM Magelang. Jurnal Ilmiah Semesta Teknika, 18 (1). (Online), tersedia: https://journal.umy.ac.id, diunduh 11 November 2016.

Fina, N., Surya, D., 2015. Penerapan Metode K-Means Clustering pada Data Penerimaan Mahasiswa Baru (Universitas Potensi Utama). Makalah diseminarkan pada Seminar Nasional Teknologi Informasi dan Multimedia, 2015. (Online), tersedia: https://ojs.amikom.ac.id, diunduh 11 November 2016.

Larose, T.Daniel., 2005. Discovering Knowledge in Data : an Introduction to Data Mining. Jhon Wiley & Sons Inc.

Liliana, S., 2013. Penerapan Algoritma C.45 untuk Penentuan Jurusan Mahasiswa. Jurnal GEMA AKTUALITA, 2 (1). (Online), tersedia: https://documentslide.com, diunduh 20 November 2016.

Setiawan, Dian. 2015. Perancangan Aplikasi *K-Means* Sebagai Penentu Konsentrasi Bagi Mahasiswa Informatika UMS. (Online), tersedia: eprints.ums.ac.id, diunduh 5 November 2016.

Sreedevi, K., dan Chandra, S.P., 2014. Analyzing the Student's Academic Performance by Using Clustering Methods in Data Mining. International Journal of Scientific & Engineering Research. 5 (6), tersedia : https://www.ijser.org, diunduh 11 November 2016.