

## ANALISIS POLA MASA STUDI MAHASISWA FAKULTAS TEKNIK UNIVERSITAS DARMA PERSADA MENGGUNAKAN METODE *CLUSTERING* K-MEANS

Ayu Hardianti\*<sup>1</sup>, Dewi Agushinta R.<sup>2</sup>

<sup>1</sup>Program Studi Magister Sistem Informasi, Program Pasca Sarjana, Universitas Gunadarma

<sup>2</sup>Jurusan Sistem Informasi, Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Gunadarma

E-mail : <sup>1</sup>hardianti.ayu1988@gmail.com, <sup>2</sup>dewiar@staff.gunadarma.ac.id

\*Penulis Korespondensi

(Naskah masuk: 18 Agustus 2018, diterima untuk diterbitkan: 28 Februari 2019)

### Abstrak

Lama studi mahasiswa menjadi salah satu faktor yang memengaruhi ketepatan waktu mahasiswa menyelesaikan kuliah. Berdasarkan kebijakan dari Badan Akreditasi Nasional Perguruan Tinggi (BAN PT) pada peraturan No 4 tahun 2017 mengenai Kebijakan Penyusunan Instrumen Akreditasi, lama studi merupakan salah satu tolak ukur dan elemen penilaian dalam akreditasi program studi. Dari data akademik Fakultas Teknik Universitas Darma Persada, banyak mahasiswa menempuh lama studi lebih dari empat tahun. Lama studi menjadi salah satu masalah pengelola program studi dalam hal kinerja akademik. Penelitian ini bertujuan menganalisis pola lama studi mahasiswa Fakultas Teknik Universitas Darma Persada dari data akademik. Teknik *clustering* algoritma K-Means digunakan dengan variabel-variabelnya adalah jurusan, daerah asal, umur, jenis kelamin, Indeks Prestasi Kumulatif (IPK), Satuan Kredit Semester (SKS), tahun masuk dan lama studi. Perangkat lunak *Waikato Environment for Knowledge Analysis* (WEKA) dipakai sebagai tool penganalisis. Tahap awal penelitian melalui pengumpulan data dari arsip atau *database* bagian Akademik yaitu data akademik mahasiswa Fakultas Teknik Universitas Darma Persada angkatan 2009 sampai 2014. Tahapan selanjutnya adalah *preprocessing* data yang dilakukan melalui analisis *clustering* algoritma K-Means dengan terlebih dahulu menghitung banyak *cluster* menggunakan metode Elbow dan interpretasi hasil. Dari hasil metode Elbow, jumlah *cluster* yang dipergunakan sebanyak 4 (empat) *cluster*. Berdasarkan hasil proses *clustering* K-Means, pembagian data pada masing-masing *cluster* adalah *cluster* 1 (satu) berjumlah 556 data (26%), *cluster* 2 (dua) berjumlah 414 data (19%), *cluster* 3 (tiga) berjumlah 189 data (9%) dan *cluster* 4 (empat) berjumlah 1010 data (46%). Selanjutnya, yang memiliki lama studi lebih dari 4 tahun terdapat pada *cluster* 2, *cluster* 3, *cluster* 4 dan mahasiswa yang memiliki masa studi 4 tahun ada pada *cluster* 1.

**Kata kunci:** algoritma K-Means, *clustering*, *data mining*, lama studi, WEKA

## PATTERN ANALYSIS OF THE STUDENT STUDY PERIOD IN THE FACULTY OF ENGINEERING AT DARMA PERSADA UNIVERSITY USING K-MEANS CLUSTERING METHOD

### Abstract

The duration of student study is one of the factors that influence the completing students' timeliness. Based on the policy of the National Accreditation Board of Higher Education (BAN-PT) in Regulation No. 4 of 2017 concerning the Policy for Preparing Accreditation Instruments, the duration of study is one of the benchmarks and evaluation elements in accreditation of study programs. From the Faculty of Engineering academic data, Darma Persada University, many students take more than four years of study. The duration of study is one of the problems of the study program manager in terms of academic performance. This study aims to analyze the old patterns of study by students of the Faculty of Engineering, Darma Persada University from academic data. K-Means algorithm clustering technique is used with the variables are majors, the area of origin, age, gender, Grade Point Average (GPA), Semester Credit Unit (SKS), year of entry and study duration. The *Waikato Environment for Knowledge Analysis* (WEKA) software is used as an analytic tool. The initial stage of research is through collecting data from archives or Academic sections, namely academic data from students of the Faculty of Engineering, Darma Persada University, 2009 to 2014. The next stage is preprocessing data through K-Means algorithm clustering analysis by first calculating many clusters using the Elbow method and result interpretation. From the Elbow method result, the number of clusters used is 4 (four) clusters. Based on the results of the K-Means clustering process, the data

*sharing in each cluster is cluster 1 (one) totaling 556 data (26%), cluster 2 (two) totaling 414 data (19%), cluster 3 (three) totaling 189 data (9%) and cluster 4 (four) totaling 1010 data (46%). Furthermore, those who have more than 4 years of study are in cluster 2, cluster 3, cluster 4 and students who have a 4-year study period are in cluster 1.*

**Keywords:** *clustering, data mining, K-Means algorithm, study period, WEKA.*

## 1. PENDAHULUAN

Pendidikan adalah salah satu kunci kemajuan bangsa Indonesia, sampai dengan tahun 2015 terdapat 3.225 perguruan tinggi aktif di Indonesia dengan jumlah mahasiswa lebih dari lima juta orang yang tersebar di 34 provinsi (bps.go.id). Universitas Darma Persada adalah perguruan tinggi swasta di Indonesia, dengan jumlah mahasiswa sebanyak 9.865 orang pada tahun ajaran 2016/ 2017, yang tersebar dalam 15 program studi (forlap DIKTI, 2017).

Agar terwujud efisiensi perguruan tinggi maka pengelola perguruan tinggi dituntut untuk mengelola secara profesional. Salah satu faktor yang memengaruhi efisiensi adalah lama studi mahasiswa. Berdasarkan kebijakan dari Badan Akreditasi Nasional Perguruan Tinggi (BAN PT) yang terdapat pada peraturan No. 4 tahun 2017 mengenai Kebijakan Penyusunan Instrumen Akreditasi, lama studi merupakan salah satu tolak ukur dan elemen penilaian dalam akreditasi program studi (BAN PT, 2017). Berdasarkan data akademik dari program studi Fakultas Teknik Universitas Darma Persada yang terdiri dari tujuh program studi (Teknik Elektro, Teknik Industri, Teknik Informatika, Sistem Informasi, Teknik Mesin, Teknik Perkapalan dan Sistem Perkapalan), banyak mahasiswa yang lama studinya mencapai lebih dari empat tahun. Lama studi menjadi salah satu faktor masalah pengelola program studi sehingga dapat memengaruhi kinerja akademik.

Untuk mengatasi permasalahan mengenai lama studi, dibutuhkan sebuah sistem informasi yang dapat mengumpulkan data dalam jumlah besar. Salah satu teknik yang dapat digunakan adalah teknik *data mining*. *Data Mining* adalah teknik untuk menelusuri data kemudian membuat model yang digunakan untuk mencari pola data lain yang tidak berada dalam *database* yang disimpan (Prasetyo, 2012).

Analisis terhadap pola lama studi merupakan salah satu bagian dari teknik *data mining*. Metode-metode *data mining*, antara lain metode klasifikasi, asosiasi dan *clustering*. Penelitian yang menggunakan metode *clustering* untuk analisis pola lama studi mahasiswa masih sedikit dilakukan dibandingkan menggunakan metode lainnya. *Cluster* merupakan analisis yang mengelompokkan mahasiswa dengan perilaku atau ciri sama pada *database* mahasiswa yang mengandung informasi pada masa lalu (Hervianti, 2018). Salah satu algoritma untuk analisis *cluster* adalah K-Means *clustering*. K-Means *clustering* merupakan bagian

pembelajaran tanpa pengawasan (*unsupervised learning*) (Prasetyo, 2012).

Penelitian sebelumnya yang dilakukan mengenai analisis pola lama studi mahasiswa dilakukan oleh Khyati Manvar dan Madhuri Rao yang berjudul “*Predicting students’ performance in higher education: A Data Mining Approach*”. Mereka menggunakan 500 data siswa pada tahun akademik 2009-2012 untuk menemukan pola memprediksi perilaku siswa. Teknik yang digunakan adalah algoritma *clustering* K-Means dan teknik *Asosiasi Rule Mining* algoritma Apriori. Hasil dari penelitian dengan menggunakan algoritma K-Means, terdapat 3 (tiga) kelompok (*cluster*) siswa yaitu *poor*, *good*, dan *excellent*. Sebanyak 32 siswa berada di *cluster* pertama (*poor*), 101 siswa di *cluster* kedua (*good*) dan 130 siswa di *cluster* ketiga (*excellent*). Hasil yang didapat dengan menggunakan algoritma Apriori untuk memprediksi kinerja mahasiswa berdasarkan nilai minimum *support* 20% dan minimum *confidence* 80% (Manvar dan Rao, 2014).

Penelitian yang dilakukan oleh Arief Jananto untuk memperkirakan waktu studi mahasiswa yang menggunakan teknik *data mining* klasifikasi algoritma Naïve Bayes, melakukan prediksi ketepatan waktu studi mahasiswa Teknik Informasi Universitas Stikubank. Variabel yang digunakan adalah kota lahir, jenis sekolah, jenis kelamin, kota sekolah dan data akademik selama kuliah sampai dengan semester 4. Hasil dari penelitian menunjukkan sebanyak 254 dari 258 mahasiswa diprediksi lulus tepat waktu dan sebanyak 4 mahasiswa diprediksi lulus tidak tepat waktu (Jananto, 2013).

Penelitian oleh Riska Haryati, Aji Sudarsono dan Eko Suryana yang berjudul “*Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5*”, berbasiskan pada pohon keputusan untuk menentukan mahasiswa lulus tepat waktu atau tidak. Tahap penelitian yang dilakukan, mengumpulkan data mahasiswa Fakultas Komputer Universitas Dehasen Bengkulu dengan variabel yang digunakan adalah jenis kelamin, IPK dan jumlah SKS menggunakan aplikasi Rapid Miner. Hasil dari proses C4.5 memberikan nilai *accuracy* 95% dari hasil matriks *confusion* (Haryati, Sudarsono dan Suryana, 2015).

Penelitian oleh Ryan Dwi Pambudi, Ahmad Afif Supianto dan Nanang Yudi Setiawan yang berjudul “*Prediksi Kelulusan Mahasiswa Berdasarkan Kinerja Akademik Menggunakan Pendekatan Data Mining Pada Program Studi Sistem Informasi Fakultas Ilmu Komputer Universitas Brawijaya*”, menggunakan

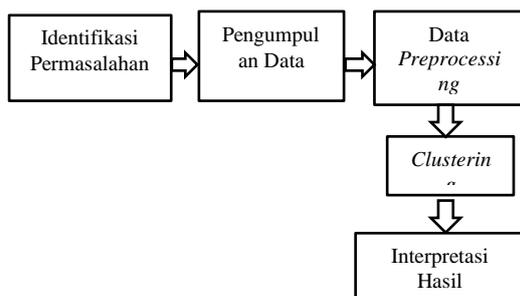
teknik klasifikasi algoritma Naïve Bayes dengan melakukan prediksi ketepatan kelulusan mahasiswa prodi sistem informasi. Ada sebanyak 1.354 data mahasiswa angkatan 2011-2016 diuji. Hasil pengujian *usability* dengan *System Usability Scale* (SUS) menunjukkan bahwa 57.5 termasuk *adjective rating Good* (Pambudi, Supianto dan Setiawan, 2019).

Berdasarkan penelitian-penelitian yang pernah dilakukan mengenai lama studi mahasiswa, analisis pola lama studi mahasiswa yang akan dilakukan Penulis menggunakan teknik *clustering* algoritma K-Means. Penelitian ini diharapkan dapat membantu pengelola program studi Fakultas Teknik Universitas Darma Persada melakukan analisis data kelulusan mahasiswa untuk meningkatkan kinerja akademik.

## 2. METODE PENELITIAN

### 2.1 Tahapan Penelitian

Tahapan penelitian yang dilakukan dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

Pada gambar 1 ada 5 (lima) tahapan penelitian yang dilakukan, dimulai dari identifikasi permasalahan, mengumpulkan data, data *preprocessing*, *clustering* serta interpretasi hasil.

### 2.2 Identifikasi Masalah

Pada tahap identifikasi masalah, informasi yang dibutuhkan adalah lama studi mahasiswa. Pengumpulan informasi yang dilakukan dengan observasi pustaka dan studi literatur dari berbagai jurnal dan penelitian yang terkait lama studi mahasiswa.

### 2.3 Pengumpulan Data

Pada tahap pengumpulan data, data yang dibutuhkan dikumpulkan Penulis sehingga dapat diolah menjadi informasi yang diharapkan. Tahap mengumpulkan data dilakukan dengan mendokumentasikan data akademik mahasiswa Fakultas Teknik Universitas Darma Persada angkatan 2009 hingga 2014. Adapun deskripsi data ditunjukkan pada Tabel 1, sedangkan penjelasan tiap atributnya pada Tabel 2.

Tabel 1. Deskripsi Data

Jenis Data	Item Data
Data akademik	Daerah asal, umur, jenis kelamin Jurusan, IPK, SKS, tahun masuk dan lama studi

Tabel 1 merupakan tabel data akademik yang terdiri dari 8 (delapan) atribut yaitu jurusan, daerah asal, umur, jenis kelamin, IPK (Indeks Prestasi Kumulatif), SKS (Satuan Kredit Semester), tahun masuk dan lama studi.

Tabel 2. Deskripsi Variabel Data

Atribut	Jenis Data	Keterangan
Daerah asal	Nominal	Daerah tempat tinggal sebelum kuliah
Umur	Numerik	Usia mahasiswa saat masuk universitas
Jenis kelamin	Nominal	Jenis kelamin mahasiswa
Jurusan	Nominal	Program studi yang diambil mahasiswa
IPK	Numerik	Indeks Prestasi Kumulatif (IPK) merupakan tolak ukur kemampuan mahasiswa
SKS	Numerik	Total SKS yang telah ditempuh mahasiswa
Tahun Masuk	Numerik	Tahun masuk mahasiswa
Lama Studi	Nominal	Waktu studi yang ditempuh mahasiswa selama masa perkuliahan sampai lulus

### 2.4 Data Preprocessing

Pada tahap ini, dilakukan data *preprocessing* untuk mempersiapkan data. Pada penelitian ini, data *preprocessing* dilakukan melalui langkah pembersihan data (*cleansing*) dan inialisasi atau pemberian label data. Pada tahap inialisasi atau pemberian label, data daerah asal, jenis kelamin, jurusan dan lama studi yang bernilai nominal, terlebih dahulu dikonversikan ke dalam bentuk numerik.

#### a. Inialisasi nilai atribut daerah asal

Pada atribut daerah asal dari data mahasiswa dilakukan inialisasi nilai atribut menjadi kode numerik seperti ditunjukkan pada Tabel 3.

Tabel 3. Inialisasi Daerah Asal

Provinsi	Inialisasi
Aceh	101
Sumatera Utara	102
Sumatera Barat	103
Riau	104
Jambi	105
Sumatera Selatan	106
Bengkulu	107
Lampung	108
DKI Jakarta	109
Jawa Barat	110
Jawa Tengah	111
Banten	112
Jawa Timur	113
Yogyakarta	114
Bali	115

Provinsi	Inisialisasi
NTB	116
NTT	117
Kalimantan Barat	118
Kalimantan Tengah	119
Kalimantan Selatan	120
Kalimantan Timur	121
Kalimantan Utara	122
Sulawesi Utara	123
Sulawesi Tengah	124
Sulawesi Selatan	125
Sulawesi Tenggara	126
Gorontalo	127
Sulawesi Barat	128
Maluku	129
Maluku Utara	130
Papua	131
Papua Barat	132

b. Inisialisasi atribut jenis kelamin  
 Pada atribut jenis kelamin dari data mahasiswa dilakukan inisialisasi nilai atribut menjadi kode numerik seperti ditunjukkan pada Tabel 4.

Tabel 4. Inisialisasi Jenis Kelamin

Jenis Kelamin	Inisialisasi
Laki-laki	11
Perempuan	12

c. Inisialisasi atribut jurusan  
 Pada atribut jurusan dari data mahasiswa dilakukan inisialisasi nilai atribut menjadi kode numerik seperti ditunjukkan pada Tabel 5.

Tabel 5. Inisialisasi Jurusan

Jurusan	Inisialisasi
Teknik Elektro	1
Teknik Industri	2
Teknik Informatika	3
Sistem Informasi	4
Teknik Mesin	5
Teknik Perkapalan	6
Sistem Perkapalan	7

d. Inisialisasi atribut lama studi  
 Pada atribut lama studi dari data mahasiswa dilakukan inisialisasi nilai atribut menjadi kode numerik seperti ditunjukkan pada Tabel 6.

Tabel 6. Inisialisasi Data Lama Studi

Lama Studi	Inisialisasi
= 4 tahun	21
> 4 tahun	22

## 2.5 Clustering

Berdasarkan pembahasan yang telah disebutkan di awal, penelitian dengan metode *clustering* ini menggunakan metode *Elbow* untuk menentukan jumlah *cluster* (*k*) dan algoritma K-Means.

### 2.5.1 Metode Elbow

Metode *Elbow* adalah metode untuk menetapkan jumlah *cluster* terbaik di dalam proses *clustering*

dengan cara melihat perbandingan hasil setiap *cluster* yang membentuk siku (*Elbow*) (Merliana, Ernawati dan Santoso, 2015). langkah-langkah metode *Elbow* untuk menentukan nilai *k* pada algoritma K-Means adalah (Bholowalia dan Kumar, 2014):

1. Mulai
2. Pemberian label nilai awal *k*
3. Tambah nilai *k*
4. Lihat nilai *Sum of Square Error* dari tiap nilai *k*, dalam percobaan ini nilai *Sum of Error* didapat dari nilai *Sum of Square Error* yang ada pada *tools* WEKA
5. Lihat hasil selisih *Sum of Square Error* dengan nilai *k* yang yang paling besar penurunannya
6. Nilai *k* yang selisih penurunannya paling besar ditetapkan sebagai nilai *k* yang berbentuk siku dan ditetapkan sebagai jumlah *cluster*
7. Selesai.

Rumus *Sum of Square Error* (SSE) pada K-Means (Kordinariya dan Makwana, 2013):

$$W(S, C) = \sum_{k=1}^k \sum_{i \in S_k} \|y_i - C_k\|^2 \quad (1)$$

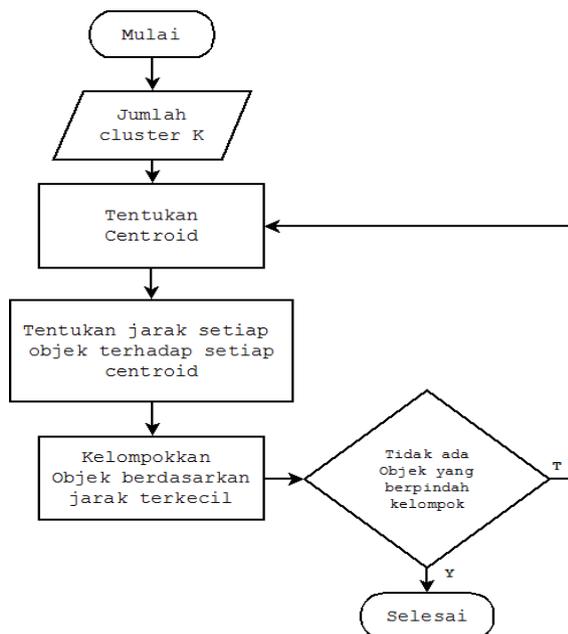
Dengan:

- S : K – *cluster* yang telah terbentuk
- $y_i$  : Data pada *y* pada indeks ke-*i*
- $C_k$  : Rerata K-*cluster* pada nilai *k* (1, 2, ..., *k*).

### 2.5.2 Algoritma K-Means

K-Means adalah algoritma yang mampu mengelola data dalam jumlah banyak, namun sangat mudah dalam proses *clustering* (Rahadian, Kurnianingtyas, Mahardika, Maghfira dan Cholissodin, 2017).

Berikut merupakan tahap-tahap pada algoritma K-Means pada Gambar 2.



Gambar 2. Flowchart K-Means

Penentuan *centroid* awal dilakukan secara *random* atau acak pada  $k$  *cluster*. Persamaan (2) untuk mencari nilai *centroid* (titik pusat) ke- $i$  selanjutnya, penentuan *centroid* awal.

$$v = \frac{\sum_{i=1}^n x_i}{n}; n = 1, 2, 3, \dots, n \quad (2)$$

Dengan

$V$  : *Centroid* pada *cluster*

$x_i$  : Objek ke- $i$

$n$  : Jumlah objek

Setelah menghitung *centroid* per *cluster* selanjutnya menghitung jarak antara objek dengan *centroid* dengan persamaan (3) *Euclidian Distance*.

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

Dengan

$x_i$  : objek x ke- $i$

$y_i$  : sumber y ke- $i$

$n$  : jumlah objek

Pada penelitian ini delapan atribut, yaitu daerah asal, umur, jenis kelamin, jurusan, IPK, SKS, tahun masuk dan lama studi dimasukkan sebagai variabel  $x$  dan  $y$  ke persamaan (3).

## 2.6 Interpretasi Hasil

Hasil luaran (output) dari *data mining* yang menerapkan metode analisis *cluster* atau *clustering* algoritma K-Means berupa deskripsi mahasiswa berdasarkan ke delapan atribut tersebut.

## 3. HASIL DAN PEMBAHASAN

Data yang digunakan penelitian ini adalah data akademik mahasiswa Universitas Darma Persada Fakultas Teknik angkatan 2009 sampai tahun 2014 sebanyak 2.182. Tahapan awal penelitian ini adalah tahapan *preprocessing* data yaitu *cleansing* data dan inisialisasi/ label data.

### 3.1. Menentukan Banyak Cluster Metode Elbow

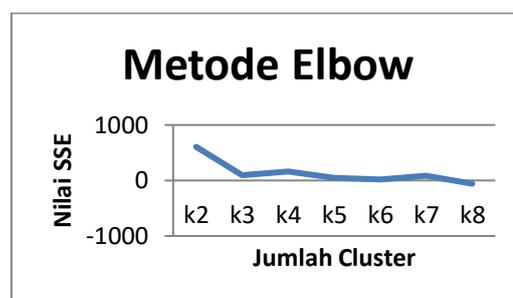
Setelah melakukan tahap *preprocessing* data, tahapan selanjutnya adalah menentukan banyak *cluster* dengan metode *Elbow*. Percobaan dilakukan tiga (3) kali, yaitu 2.182 data (100% dari data akademik), 1.746 data (80% dari data akademik) dan 1.310 data (60% dari data akademik). Banyak *cluster* yang akan diuji adalah dari  $k = 2$  hingga  $k = 8$  (Merliana, Ernawati dan Santoso, 2015).

#### 3.1.1 Percobaan I : 2.182 Data Akademik

Nilai *Sum of Square Error* (SSE) yang ada di dalam aplikasi WEKA 3.8.2 terhadap 2.182 data akademik mahasiswa memberikan hasil yang mengalami penurunan selisih paling besar pada nilai *cluster* ( $k$ ) = 4, selisih nilai *Sum of Square Error* (SSE) dapat dilihat pada tabel 7 dan gambar 3.

Tabel 7. Hasil *Sum of Square Error* 2.182 Data

Cluster	Hasil Sum of Square Error (SSE)	Selisih
2	605.8204	605.8204
3	514.0231	91.79736
4	354.7425	159.28061
5	303.3396	51.40284
6	282.9092	20.430426
7	198.0862	84.822974
8	256.8127	-58.7265



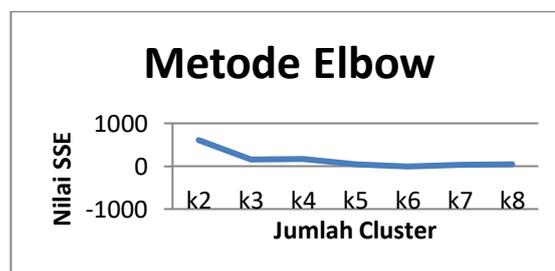
Gambar 3. Grafik *Elbow* 2.182 Data

#### 3.1.2 Percobaan II : 1.746 Data Akademik

Nilai *Sum of Square Error* (SSE) yang ada di dalam aplikasi WEKA 3.8.2 terhadap 1.746 data akademik mahasiswa yang penurunan selisih paling besar adalah pada *cluster* ( $k$ ) = 4, selisih nilai *Sum of Square Error* (SSE) dapat dilihat pada tabel 8 dan gambar 4.

Tabel 8. Hasil *Sum of Square Error* 1.746 Data

Cluster	Hasil Sum of Square Error (SSE)	Selisih
2	648.6235	648.6235
3	591.0298	57.5937
4	443.0051	148.0247
5	346.9833	96.0218
6	322.2287	24.7546
7	294.6317	27.597
8	267.5738	27.0579



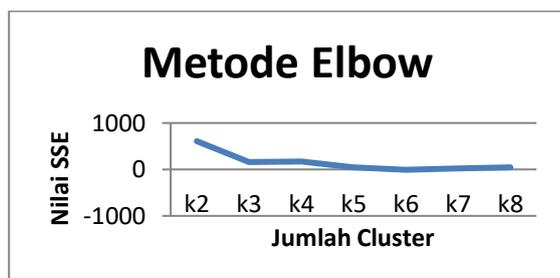
Gambar 4. Grafik *Elbow* 1.746 Data

#### 3.1.3 Percobaan III : 1.310 Data Akademik

Nilai *Sum of Square Error* (SSE) yang ada di dalam aplikasi WEKA 3.8.2 terhadap 1.310 data akademik mahasiswa yang mengalami penurunan selisih paling besar adalah juga pada *cluster* ( $k$ ) = 4, selisih nilai *Sum of Square Error* (SSE) dapat dilihat pada Tabel 9 dan Gambar 5.

Tabel 9. Hasil *Sum of Square* 1.310 Data

Cluster	Hasil <i>Sum of Square Error</i> (SSE)	Selisih
2	607.9075	607.9075
3	448.8437	159.0638
4	278.4295	170.4142
5	229.1670	49.2625
6	236.8686	-7.7016
7	210.4037	26.4649
8	163.4165	46.9872

Gambar 5. Grafik *Elbow* dengan 1.310 Data

Dari ketiga percobaan berdasarkan data akademik mahasiswa Fakultas Teknik, terjadi penurunan selisih nilai *Sum of Square Error* (SSE) yang paling besar semua pada *cluster* ( $k$ ) = 4, sehingga jumlah *cluster* yang digunakan dalam penelitian ini adalah nilai *cluster* ( $k$ ) = 4. Nilai  $k = 4$  akan digunakan sebagai jumlah *cluster* untuk tahapan selanjutnya.

### 3.2 Analisis Hasil *Clustering*

Analisis ini menggunakan data akademik mahasiswa Fakultas Teknik angkatan tahun 2009 sampai tahun 2014 yang telah lulus dan telah melalui tahap *preprocessing* sebanyak 2.182 data. Perhitungan pada algoritma K-Means dengan bantuan tool WEKA 3.8.2 diawali dengan menentukan jumlah *cluster*. Penelitian ini menggunakan 4 *cluster* berdasarkan hasil uji coba dengan metode *Elbow* sebelumnya. Data yang digunakan pada penelitian ini harus dikonversi dahulu menjadi file yang dapat diproses oleh WEKA 3.8.2 yaitu file .csv dengan nama file *data\_akademik\_mahasiswa.csv*.

Berdasarkan data hasil *clustering* dengan total 2.182 data dengan 8 (delapan) atribut (jurusan, provinsi, umur, jenis kelamin, IPK, SKS, tahun masuk dan lama studi) didapatkan hasil *clustering* sampai iterasi ke-11, dengan titik pusat tidak berubah. Nilai *Sum of Square Errors* (SSE) adalah 517.8904605474909. Titik pusat awal pada setiap *cluster* dapat dilihat pada tabel 10.

Tabel 10. Titik Pusat Awal Tiap *Cluster*

Titik Pusat Awal	Jur	Prop	U	JK	IPK	SKS	Msk	Lama Studi
Cluster 1	5	109	18	11	2.64	134	2013	22
Cluster 2	1	109	18	11	1.03	62	2014	22
Cluster 3	3	110	45	11	2.92	143	2014	22
Cluster 4	3	110	18	11	1.95	65	2014	22

data (19%), *cluster* 3 berjumlah 189 data (9%) dan *cluster* 4 berjumlah 1010 data (46%). Pembagian data pada masing-masing *cluster* dilihat pada tabel 11.

Tabel 11. Jumlah Data *Cluster*

Cluster	Banyak Data	Data dalam Cluster (%)
Cluster 1	569	26
Cluster 2	414	19
Cluster 3	189	9
Cluster 4	1010	46

Dari hasil *clustering* K-Means yang telah diperoleh dari tool WEKA, beberapa gambaran terhadap pola lama studi mahasiswa antara lain:

1. Data yang memiliki lama studi lebih dari 4 tahun ada pada *cluster* 2, *cluster* 3, *cluster* 4; mahasiswa yang memiliki masa studi 4 tahun di *cluster* 1.
2. Pengelompokan berdasarkan SKS dan IPK memiliki kemiripan dengan pengelompokan berdasarkan lama studi, *cluster* 1 memiliki IPK 3.1936 dengan SKS 146.1371, sedangkan pada *cluster* 2, *cluster* 3 dan *cluster* 4 memiliki IPK kurang dari 3 dan SKS kurang dari 146 yaitu masing-masing IPK 2.0114 dan SKS 112.756, IPK 2.4043 dan SKS 114.2381, IPK 2.1563 dan SKS 97.096. Mahasiswa yang lulus tepat waktu (4 tahun) adalah yang memiliki IPK lebih besar atau sama dengan 3 dengan jumlah SKS yang ditempuh 146 SKS.
3. Berdasarkan tahun masuk, mahasiswa tahun 2012 pada *cluster* 1 (lulus 4 tahun) diperkirakan lulus tahun 2016, mahasiswa 2009 pada *cluster* 2 (lulus lebih dari 4 tahun) ini diperkirakan lulus setelah tahun 2012, mahasiswa 2009 pada *cluster* 3 (lulus lebih dari 4 tahun) diperkirakan lulus setelah tahun 2016, mahasiswa tahun 2013 pada *cluster* 4 (lulus lebih dari 4 tahun) diperkirakan lulus setelah tahun 2018.
4. Berdasarkan atribut jenis kelamin, jenis kelamin laki-laki memiliki peluang sama besar antara lulus tepat waktu dan tidak tepat waktu, sedangkan untuk jenis kelamin perempuan, peluang lulusnya menjadi tidak tepat waktu.

## 4. KESIMPULAN DAN SARAN

Penelitian ini memberikan gambaran mengenai analisis pola lama studi mahasiswa di Fakultas Teknik Universitas Darma Persada menggunakan teknik *clustering* algoritma K-Means. Kelompok mahasiswa yang memiliki lama studi tepat 4 tahun berada pada *cluster* 1 dan kelompok mahasiswa dengan lama studi lebih dari 4 tahun berada pada *cluster* 2, *cluster* 3 dan *cluster* 4. Hasil analisis proses

K-Means pada *tools* WEKA, ada 2 (dua) dari 8 (delapan) atribut atau variabel yang digunakan pada penelitian ini mempengaruhi lama studi mahasiswa tepat 4 tahun yaitu jumlah SKS dan IPK.

Tabel 12 merupakan tabel hasil *clustering* dengan algoritma K-Means.

Tabel 12. Hasil *Clustering* K-Means

Atribut	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Jurusan	3.5554	3.5459	3.418	3.6644
Provinsi	109.7715	109.6739	109.8519	109.8584
Umur	20.4763	20.8285	20.4127	20.1941
Jenis Kelamin	11.1986	11	12	11
IPK	3.1936	2.0114	2.4043	2.1563
SKS	146.137	112.756	114.238	97.096
Tahun Masuk	1	1	1	1
Lama Studi	2012.64	2009.87	2012.28	2013.25
	32	92	04	45
	21	22	22	22
Jumlah data	569.0	414.0	189.0	1010.0

SKS lebih dari atau sama dengan 146 dan IPK lebih dari atau sama dengan 3 menyebabkan lama studi tepat waktu, begitu sebaliknya. Pada penelitian selanjutnya, teknik *data mining clustering* lainnya atau penggabungan dua teknik *data mining* dan penambahan metode untuk mengukur kualitas dan kekuatan *cluster* dapat dipertimbangkan.

#### DAFTAR PUSTAKA

- B. A. N. Indonesia, "Peraturan Badan Akreditasi Nasional PT No 4/2017," 2017.
- BHOLOWALI, P., dan KUMAR, A., 2014. A Clustering Technique Based on Elbow Method and K-Means in WSN. *International Journal of Computer Applications*. Vol. 105, No. 9. pp. 17-24.
- BPS. 2015. Jumlah perguruan tinggi mahasiswa dan tenaga edukatif negeri dan swasta di bawah kementerian pendidikan dan kebudayaan, [Online] Tersedia di: <<https://www.bps.go.id/>> [Diakses 2 Januari 2018]
- FORLAP DIKTI. 2008. Pangkalan Data Pendidikan Tinggi Kementerian Riset, Teknologi dan Pendidikan Tinggi, [Online] Tersedia di: <<https://forlap.ristekdikti.go.id/>> [Diakses 2 Januari 2018]
- HARYATI, S., SUDARSONO, A., & SURYANA, E., 2015. Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan C4.5 (Studi Kasus: Universitas Dehasen Bengkulu). *Jurnal Media Infotama*. Vol. 11, No 2. pp. 130-138.
- HERVIANTI, S.P., 2018. Analisis Pengelompokan Penyebaran Lulusan Mahasiswa Universitas Gunadarma Menggunakan Metode Clustering. *IKRAITH-Informatika*. Vol. 2, No 2. pp. 42-48.
- JANANTO, A., 2013. Algoritma Naïve Bayes Untuk Mencari Perkiraan Waktu Studi Mahasiswa. *Jurnal Teknologi Informasi Dinamik*. Vol. 18, No 1. pp. 9-16.
- KODINARIYA, T., dan MAKWANA, P., 2013. Review on determining number of Cluster in K-Means Clustering. *International Journal of Advance Research in Computer Science and Management Studies*. Vol. 1, Issue 6. pp. 90-95.
- MANVAR, K., dan RAO, M., 2014. Predicting student's performance in higher education: A Data Mining Approach. *International Journal of Scientific & Engineering Research*. Vol. 5, Issue 2. pp. 1024-1027.
- MERLIANA, E., ERNAWATI, & SANTOSO, J., 2015. Analisis Penentuan Jumlah Cluster Terbaik Pada Metode K-Means Clustering. *Prosiding Seminar Nasional Multi Ilmu & Call For Papers Unisbank*, pp. 17-24.
- PAMBUDI, R., SUPIANO, A., & SETIAWAN, N., 2019. Prediksi Kelulusan Mahasiswa Berdasarkan Kinerja Akademik Menggunakan Pendekatan Data Mining Pada Program Studi Sistem Informasi Fakultas Ilmu Komputer Universitas Brawijaya. *Jurnal Pengembangan Teknologi dan Ilmu Komputer*. Vol. 3, No. 3. pp. 2194-2200.
- PRASETYO, E., 2012. Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab, Yogyakarta: Andi.
- RAHADIAN, B., KURNIANINGTYAS, D., MAHARDIKA, D., MAGHFIRA, T., dan CHOLISSODIN, I., 2017. Analisis Judul Majalah Kawanku Menggunakan Clustering K-Means Dengan Konsep Simulasi Big Data Pada Hadoop Multi Mote Cluster. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*. Vol. 4, No. 2. pp. 75-80.

*Halaman ini sengaja dikosongkan*