

ANALISIS HASIL BELAJAR MAHASISWA DENGAN CLUSTERING MENGGUNAKAN METODE K-MEANS

Berlian Juliarta Martin Putra¹⁾, Dwi Ariani Finda Yuniarti²⁾

berlian@aknpacitan.ac.id ¹⁾, finda@aknpacitan.ac.id ²⁾

^{1, 2)} Prodi Pemeliharaan Komputer dan Jaringan, Akademi Komunitas Negeri Pacitan

Abstrak

Hasil belajar mahasiswa dapat menjadi sebuah informasi yang sangat penting terhadap kemajuan pembelajaran di sebuah Perguruan tinggi. Evaluasi hasil belajar diperlukan untuk menghindari mahasiswa yang lulus tidak tepat waktu, mahasiswa mengundurkan diri, meningkatkan prestasi mahasiswa, dan bahan evaluasi perguruan tinggi untuk meningkatkan kualitasnya. Evaluasi mahasiswa dilakukan secara terus menerus minimal sekali setiap semesternya, apabila dilakukan secara manual maka membutuhkan upaya dan waktu tersendiri sehingga tidak efektif dan efisien. Penelitian ini mengajukan analisis hasil belajar mahasiswa dengan teknik *clustering*. *Clustering* merupakan metode data mining untuk pengelompokan data, salah satunya adalah metode K-Means. Tahapan dalam penelitian ini yaitu: *data collection*, *preprocessing*, seleksi fitur, *clustering* menggunakan metode *simple K-Means* dengan tool *Waikato Environment for Knowledge Analysis* (WEKA) dan perhitungan jarak menggunakan *Euclidean Distance* dan analisa *cluster*. Adapun hasil dari penelitian ini adalah pengelompokan hasil belajar mahasiswa menggunakan variabel IPK memberikan *Sum of Squared Errors* (SSE) lebih kecil dibandingkan pengelompokan dengan variabel nilai di setiap mata kuliah. Rata-rata SSE yang diperoleh dalam penelitian ini sebesar 0,1135 dari 4 kali percobaan. Dari penelitian ini dapat disimpulkan bahwa metode *clustering* menggunakan K-Means bisa digunakan untuk mengelompokkan nilai mahasiswa dan menganalisis hasil belajar mahasiswa setiap semesternya sehingga memudahkan dalam evaluasi mahasiswa.

Kata Kunci : *Clustering*, Hasil Belajar, Indeks Prestasi Kumulatif, K-Means, WEKA

1. PENDAHULUAN

Pendidikan merupakan usaha sadar yang dilakukan untuk menggali potensi yang ada di dalam diri seorang siswa atau mahasiswa untuk menjadikan dirinya menjadi seorang yang profesional. Pada pembelajaran di tingkat perguruan tinggi, setiap akhir semester mahasiswa memperoleh hasil belajar pada setiap mata kuliah yang diambilnya. Tentunya hal ini akan terjadi secara berulang-ulang setiap akhir semester, dimana mahasiswa akan memperoleh hasil belajar atau Indeks Prestasi Kumulatif (IPK).

Teknologi data mining merupakan salah satu cara untuk penggalian data pada basis data

berukuran besar. Dengan data mining, kita dapat melakukan pengklasifikasian, memprediksi, memperkirakan dan mendapatkan informasi lain yang bermanfaat dari kumpulan data dalam jumlah yang besar [1].

K-Means merupakan salah satu metode pengelompokan. Teknik pengelompokan datanya sederhana dan cepat. Ada banyak pendekatan untuk membuat *cluster*, diantaranya adalah membuat aturan yang mendikte keanggotaan dalam grup yang sama berdasarkan tingkat persamaan diantara anggota-anggotanya [2].

Metode K-Means telah banyak digunakan, diantaranya digunakan untuk pengelompokan

nilai akademik mahasiswa, untuk memprediksi masa studi mahasiswa [2], pertimbangan dosen dalam membentuk kelompok belajar mahasiswa [3], pengelompokan nilai UN dan IPK mahasiswa [4], pengelompokan data mahasiswa layak atau tidak layak sebagai penerima beasiswa bidikmisi [5], pengelompokan data kesehatan [6], strategi pemasaran sekolah menengah kejuruan [7], dan pengelompokan provinsi di Indonesia [8].

Penelitian ini menganalisis pengelompokan hasil belajar mahasiswa yang memiliki nilai tinggi, nilai sedang dan nilai rendah dengan menggunakan metode K-Means. Adapun manfaat dari penelitian ini adalah menganalisis hasil belajar mahasiswa, sehingga dari hasil analisis tersebut dosen dapat mengetahui kiat-kiat yang perlu dilakukan dalam proses pembelajaran di dalam kelas, dosen dapat lebih memahami karakteristik masing-masing kelompok belajar mahasiswa serta dapat memposisikan dirinya sebagai fasilitator pembelajaran terhadap kelompok mahasiswa dengan nilai tinggi, sedang dan rendah.

2. METODE PENELITIAN

2.1 Studi Literatur

Data mining adalah proses mempekerjakan teknik pembelajaran untuk menganalisis dan mengekstraksi pengetahuan secara otomatis [9]. Tahapan *data mining* adalah sebagai berikut: *data selection*, *preprocessing*, *transformation*, *data mining*, dan *evaluation* [10].

Tahap *preprocessing* dilakukan supaya bersih dari noise, memiliki dimensi yang lebih kecil, lebih terstruktur sehingga bisa diolah lebih lanjut [11]. Pada *preprocessing* juga dilakukan normalisasi data.

Transformation adalah mengubah format data yang disimpan menjadi standar format data sesuai dengan aplikasi yang digunakan

mengelompokan nilai UTS dari dua kelas pembelajaran yang bertujuan sebagai [12]. Penelitian ini mengubah format data excel menjadi .csv supaya bisa diproses menggunakan WEKA.

Seleksi fitur bisa dilakukan dengan melakukan proses data mining menggunakan semua kemungkinan kombinasi elemen vektor fitur yang ada, fitur yang dipilih adalah fitur yang memberikan akurasi tertinggi [13].

Beberapa teknik yang ada dalam data mining yaitu *classification*, *clustering*, *association rule discovery*, *sequential pattern*, *regression*, dan *deviation detection* [9].

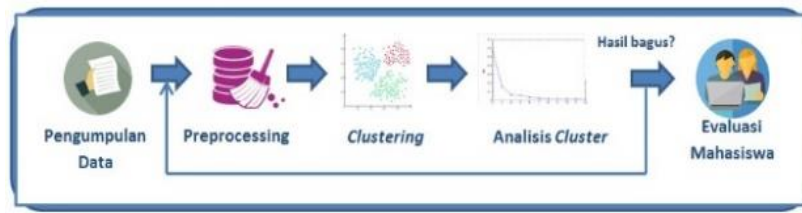
Clustering mempartisi data set menjadi beberapa kelompok, dimana anggota dalam satu kelompok memiliki similaritas yang tinggi dan similaritas antar kelompok rendah [9]. Salah satu metode *clustering* adalah K-Means. *Cluster* yang baik apabila anggota satu kelompok memiliki similaritas yang tinggi dan antar kelompok memiliki similaritas yang rendah. Salah satu cara untuk menganalisis *cluster* menggunakan metode SSE. Semakin kecil nilai SSE maka semakin tinggi tingkat similaritas data yang ada di dalam masing-masing *cluster* atau semakin baik *cluster* yang dihasilkan [14].

2.2 Metode Penelitian

Pada Gambar 1 adalah blok diagram sistem. Data yang digunakan adalah data mahasiswa Akademi Komunitas Negeri Pacitan kelas TI A selama 4 semester. Pengumpulan data dimulai dari pengumpulan nilai mahasiswa setiap mata kuliah dan IPK mahasiswa pada setiap semester.

Salah satu teknik *preprocessing* adalah pengisian *missing data*. Nilai yang hilang akan ditinjau terlebih dahulu, apakah mahasiswa cuti/mengundurkan diri atau memang *missing value*.

Apabila mahasiswa cuti/mengundurkan diri maka nilai dianggap 0 jika tidak maka



Gambar 1. Blok Diagram Sistem

akan dilakukan inputation dengan metode *mean/mode* yang ada pada WEKA.

Pada seleksi fitur dibandingkan antara fitur nilai mahasiswa pada setiap mata kuliah dengan fitur IPK menggunakan teknik *clustering*. Hasil *clustering* yang memiliki SSE terkecil maka fitur yang digunakan untuk *clustering* tersebut yang akan dipakai.

Selanjutnya dilakukan *clustering* menggunakan metode simple K-Means, $k=3$ dengan tool WEKA. K merupakan jumlah *cluster*.

Penelitian ini mengelompokkan mahasiswa kedalam 3 *cluster* untuk mengevaluasi hasil belajar mahasiswa, yaitu mahasiswa dengan nilai tinggi, mahasiswa dengan nilai sedang dan mahasiswa dengan nilai rendah.

Berikut adalah algoritma dari metode K-Means [15] :

- Bangkitkan nilai *centroid* sebanyak k secara *random*.
- Hitung jarak data ke masing-masing *centroid*, penelitian ini menggunakan *Euclidean* dengan rumus (1) sehingga ditemukan jarak terdekat dari setiap data dengan *centroid*.

$$d(x_i, c_j) = \sqrt{\sum (x_i - c_j)^2} \quad (1)$$

Keterangan :

x_i = data

c_j = *centroid* pada *cluster* ke- j

- Kelompokan data berdasar *centroid* terdekat.
- Hitung kembali nilai *centroid* baru dengan rumus (2).

$$c_i = \frac{1}{n_{sj}} \sum_{j \in S_j} x_j \quad (2)$$

Keterangan :

n_{sj} : banyak data pada *cluster* S_j

- Ulangi langkah ke-b hingga ke-d, sampai anggota setiap *cluster* tidak berubah atau sampai *centroid* lama = *centroid* baru.

K-Means bertujuan untuk meminimalkan jumlah *squared error* dari seluruh k *cluster* [14]. Berikut rumus (3) adalah perhitungan SSE.

$$J(c_k) = \sum_{k=1}^k \sum_{x_i \in c_k} (x_i - \mu_k)^2 \quad (3)$$

3. HASIL DAN PEMBAHASAN

3.1 Hasil

3.1.1. Data

Data yang digunakan adalah data nilai mahasiswa TI A angkatan 2017 – 2018 Akademi Komunitas Negeri Pacitan. *Clustering* menggunakan WEKA.

3.1.2. Data Preprocessing

Tahap ini menggunakan data kelas TI A selama 4 semester, mahasiswa yang cuti atau keluar di tengah semester dilakukan penghapusan data. Jika ditemukan nilai yang *missing value* dilakukan pengisian menggunakan metode *mean/mode* pada WEKA.

3.1.3. Seleksi Fitur

Berikut percobaan pemilihan fitur pada poin a dan b.

a. Percobaan I Seleksi Fitur Menggunakan 12 Variabel Nilai

Hasil *clustering* menggunakan 12 variabel nilai setiap mata kuliah dengan metode K-Means, $k=3$ bisa dilihat pada Gambar 2 dan Gambar 3 dihasilkan SSE sebesar 7,46.

```
kMeans
=====

Number of iterations: 4
Within cluster sum of squared errors: 7.460192785950814

Initial starting points (random):

Cluster 0: 83.5,80,67.25,82.1,64,72.5,79.5,78,81.6,70.5,71,66.3
Cluster 1: 87.1,80,71.25,62.8,67.5,83.5,74,65.1,68.7,76.5,67,71.2
Cluster 2: 89.1,80,67.25,77.7,64,54,71.5,66.9,77.1,75.5,60,70.2
```

Missing values globally replaced with mean/mode

Gambar 2. Sum of Squared Errors Clustering dengan 12 variabel

Satu mata kuliah memiliki 3 macam *centroid*. Seorang mahasiswa memiliki nilai dari dua belas mata kuliah tersebut, dengan rumus *Euclidean Distance* (1), nilai masing-masing mata kuliah, dan *centroid-centroid* yang terbentuk akan diketahui jarak mahasiswa terhadap masing-masing *cluster*, mahasiswa akan tergabung ke *cluster* dengan jarak terdekat.

81.0	80.0	55.8	67.2	60.0	68.0	70.0	61.4	74.0	66.5	61.5	66.3	cluster1
80.9	80.0	68.0	70.8	74.0	56.0	65.5	78.6	75.1	80.9	60.0	75.0	cluster2
86.3	80.0	57.25	75.7	71.0	66.5	79.5	70.2	81.1	78.2	72.0	76.2	cluster2
87.1	80.0	71.25	62.8	67.5	83.5	74.0	65.1	68.7	76.5	67.0	71.2	cluster1
80.9	81.25	57.5	64.4	73.5	58.1	79.5	70.2	68.6	75.5	60.0	70.2	cluster2
90.9	82.25	70.5	74.4	65.0	72.0	63.0	78.4	77.2	79.5	68.5	71.2	cluster2
80.6	80.0	56.1	74.9	70.5	62.0	68.5	67.5	77.2	78.0	68.5	66.3	cluster2
80.7	80.0	56.25	69.8	67.5	66.5	56.0	61.4	71.4	72.5	63.0	66.3	cluster1
90.8	86.25	81.05	82.0	83.0	82.5	88.5	78.6	82.1	81.5	74.5	71.2	cluster0
83.5	80.0	67.25	82.1	64.0	72.5	79.5	78.0	81.6	70.5	71.0	66.3	cluster0
89.1	80.0	67.25	77.7	64.0	54.0	71.5	66.9	77.1	75.5	60.0	70.2	cluster2
69.1	80.0	67.0	62.9	60.0	67.5	62.5	70.4	65.8	66.5	56.0	70.2	cluster1

Gambar 3. Hasil clustering 12 variabel

Tabel 1 adalah data hasil clustering pada Gambar 3 dibandingkan dengan IPK mahasiswa yang sama pada semester tersebut.

Tabel 1. Perbandingan Hasil Clustering dengan IPK mahasiswa

No	Cluster	IPK
1	Cluster 1	2,93
2	Cluster 2	3,18
3	Cluster 2	3,33
4	Cluster 1	3,3
5	Cluster 2	2,95
6	Cluster 2	3,4
7	Cluster 2	3,13
8	Cluster 1	2,9
9	Cluster 0	3,88
10	Cluster 0	3,5
11	Cluster 2	3,08
12	Cluster 1	2,8

Dari Gambar 4, *cluster 0* bukan berarti terdiri dari mahasiswa dengan nilai terbaik, *cluster 1* bukan berarti terdiri dari mahasiswa dengan nilai sedang. *Cluster 0* dari mata kuliah

agama hingga praktikum bas memang memiliki nilai *centroid* tertinggi dibandingkan lainnya namun ketika pada mata kuliah praktikum ket, nilai *centroid* tertinggi ada pada *cluster* ke-2 dan pada mata kuliah praktikum sis, nilai *centroid* tertinggi ada pada *cluster* ke-2. Hal ini terjadi karena suatu mahasiswa terkadang jelek di mata kuliah A namun bagus di mata kuliah B, belum berarti mahasiswa jelek di mata kuliah A disemua mata kuliah memiliki nilai jelek.

Karena hal tersebut, hasil pengelompokan tidak maksimal, seperti pada Gambar 3 dan Tabel 1. Sehingga dengan menggunakan variabel semua nilai mata kuliah mengakibatkan ketidak maksimalan hasil clustering.

Attribute	Cluster#		
	0 (2.0)	1 (4.0)	2 (6.0)
Agama	87.15	79.475	84.7833
Bahasa Inggri	83.125	80	80.5833
Matematika 1	74.15	62.575	62.7667
Basis Data	82.05	65.675	72.9833
Keterampilan	73.5	63.75	69.6667
Konsep Pemrog	77.5	71.375	61.4333
Konsep Teknol	84	65.625	71.25
Sistem Operas	78.3	64.575	71.9667
Praktikum Bas	81.85	69.975	76.05
Praktikum Ket	76	70.5	77.9333
Praktikum Kon	72.75	61.875	64.8333
Praktikum Sis	68.75	68.5	71.5167

Gambar 4. Centroid Tiap Cluster dari 12 mata kuliah

b. Percobaan II Seleksi Fitur Menggunakan Variabel IPK

Selanjutnya dicoba menggunakan variabel IPK dengan data yang sama. Hasil akhir pembangkitan *centroid* akhir pada Gambar 5.

Final cluster centroids:			
Attribute	Cluster#		
	0 (2.0)	1 (4.0)	2 (6.0)
Semester 1	3.69	3.3025	2.965

Gambar 5. Hasil akhir pembangkitan *centroid* akhir K-Means

Dari Gambar 5 dapat kita bisa mengelompokan dengan mudah bahwa *cluster* ke-0 adalah mahasiswa dengan IPK bagus, *cluster* ke-1 mahasiswa dengan IPK sedang

dan *cluster* ke-2 adalah mahasiswa dengan IPK rendah karena *centroid* akhir yang terbentuk sudah sesuai sebagai *centroid* dari *cluster* yang akan kita bentuk.

```
kMeans
=====
Number of iterations: 3
Within cluster sum of squared errors: 0.146454
```

Gambar 6. Sum of Squared Errors *Clustering* dengan variabel IPK

Dari Gambar 6 terlihat SSE dengan variabel IPK lebih kecil dibandingkan dengan SSE dengan 12 variabel. SSE didapatkan dari rumus(3) yang menandakan semakin kecil nilai SSE maka hasil *clustering* semakin bagus karena semakin *similar* anggota dari *cluster* tersebut dan semakin rendah tingkat similaritasnya terhadap antar *cluster*. Sehingga pada penelitian ini fitur yang digunakan adalah IPK.

3.1.4. Clustering

Pada tahap ini, penelitian dilakukan dengan mengelompokkan IPK mahasiswa selama 4 semester. Apabila hasil *clustering* selalu bagus maka teknik ini bisa digunakan oleh AKN Pacitan dalam mengevaluasi nilai mahasiswa setiap semesternya. *Clustering* dilakukan dengan metode K-Means dengan perhitungan jarak menggunakan Euclidean Distance(1), dengan nilai $k=3$.

a. Clustering Data Semester I

Pada Gambar 7 adalah hasil *clustering* semester 1. Nilai SSE yang didapatkan adalah 0,1465.

No: Semester 1 4: Cluster		
	Numeric	Nominal
1	2.93	cluster2
2	3.18	cluster1
3	3.33	cluster1
4	3.3	cluster1
5	2.95	cluster2
6	3.4	cluster1
7	3.13	cluster2
8	2.9	cluster2
9	3.88	cluster0
10	3.5	cluster0
11	3.08	cluster2
12	2.8	cluster2

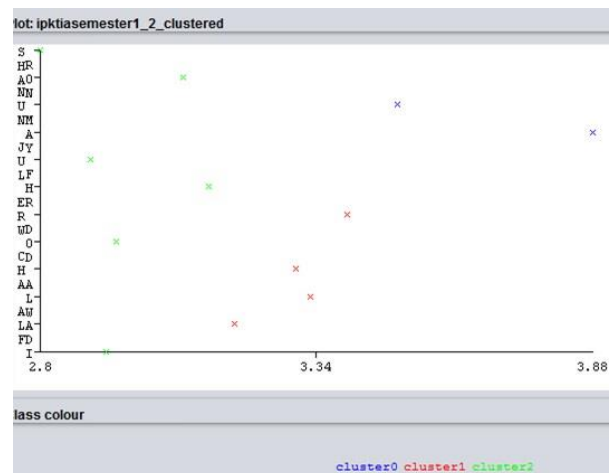
Gambar 7. Hasil *Clustering* Semester I

Tabel 2 adalah evaluasi hasil *clustering* dari Gambar 7. Sedangkan nilai final *centroid* ada pada Gambar 5.

Tabel 2. Evaluasi *Clustering* Semester I

No	Cluster	IPK
1	Cluster 2	2,93
2	Cluster 1	3,18
3	Cluster 1	3,33
4	Cluster 1	3,3
5	Cluster 2	2,95
6	Cluster 1	3,4
7	Cluster 2	3,13
8	Cluster 2	2,9
9	Cluster 0	3,88
10	Cluster 0	3,5
11	Cluster 2	3,08
12	Cluster 2	2,8

Pada Tabel 2 terlihat bahwa *cluster* 0 adalah mahasiswa dengan nilai tinggi, *cluster* 1 adalah mahasiswa dengan nilai sedang dan *cluster* 2 adalah mahasiswa dengan nilai rendah.



Gambar 8. Grafik Sebaran Nilai IPK mahasiswa

Pada Gambar 8 *cluster* 2 berwarna hijau sebanyak 6 mahasiswa, *cluster* 1 sebanyak 4 mahasiswa dan *cluster* 0 sebanyak 2 mahasiswa.

b. Clustering Data Semester II

Pada Gambar 9 adalah hasil *clustering* semester II. Nilai SSE adalah 0.1054. Nilai akhir *centroid* yang dibangkitkan bisa dilihat pada Gambar 10.

Pada Gb (10) terlihat peningkatan nilai *centroid* setiap *cluster*, hal ini menandakan bahwa nilai mahasiswa kelas TI A mengalami kenaikan. Hal ini juga terlihat pada grafik Gambar 10.

No	2: Nama	3: Semester 2	4: Cluster
	Nominal	Numeric	Nominal
1	ADITY...	2.93	cluster2
2	ALFI...	3.85	cluster0
3	ALWA...	3.45	cluster1
4	CHA...	3.58	cluster1
5	DODI...	3.4	cluster1
6	ERW...	3.6	cluster1
7	FHRE...	3.48	cluster1
8	JULIA...	3.1	cluster2
9	MAYA...	3.78	cluster0
10	NUN...	3.63	cluster0
11	RONAL	3.13	cluster2
12	SHAF...	2.93	cluster2

Gambar 9. Hasil *Clustering* Semester II

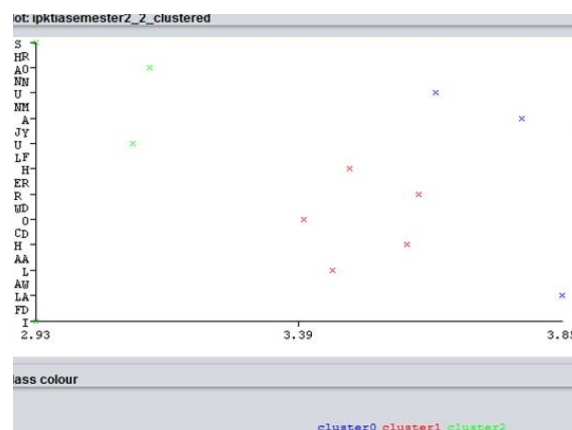
Final cluster centroids:			
Attribute	Cluster#	1	2
	(3.0)	(5.0)	(4.0)
Semester 2	3.7533	3.502	3.0225

Gambar 10 Hasil akhir pembakitan *centroid* akhir Semester II

Tabel 3 adalah evaluasi hasil *clustering* data semester II dari data pada Gambar 10.

Tabel 3. Evaluasi *Clustering* Semester II

No	Cluster	IPK
1	Cluster 2	2,93
2	Cluster 0	3,85
3	Cluster 1	3,45
4	Cluster 1	3,58
5	Cluster 1	3,4
6	Cluster 1	3,6
7	Cluster 1	3,48
8	Cluster 2	3,1
9	Cluster 0	3,78
10	Cluster 0	3,63
11	Cluster 2	3,13
12	Cluster 2	2,93



Gambar 11. Grafik Hasil *Clustering* II

Pada Gambar 11 *cluster* 2 berwarna hijau sebanyak 4 mahasiswa, *cluster* 1 sebanyak 5 mahasiswa dan *cluster* 0 sebanyak 3 mahasiswa.

c. *Clustering* Data Semester III

Pada Gambar 12 adalah hasil *clustering* semester 1. Nilai SSE yang didapatkan adalah 0,094.

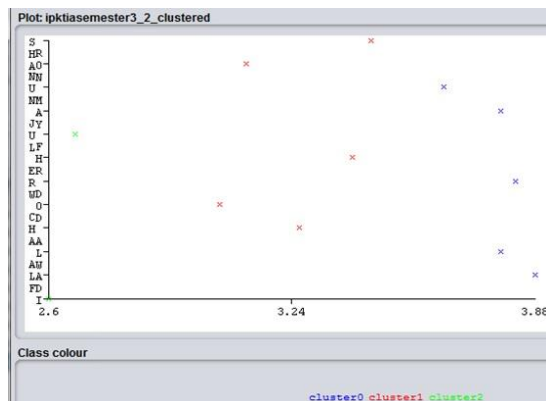
No	2: Nama	3: Semester 3	4: Cluster
	Nominal	Numeric	Nominal
1	ADITY...	2.6	cluster2
2	ALFI...	3.88	cluster0
3	ALWA...	3.79	cluster0
4	CHA...	3.26	cluster1
5	DODI...	3.05	cluster1
6	ERW...	3.83	cluster0
7	FHRE...	3.4	cluster1
8	JULIA...	2.67	cluster2
9	MAYA...	3.79	cluster0
10	NUN...	3.64	cluster0
11	RONAL	3.12	cluster1
12	SHAF...	3.45	cluster1

Gambar 12. Hasil *Clustering* Semester III

Tabel 4 adalah evaluasi hasil *clustering* semester III dari data pada Gambar 12.

Tabel 4. Evaluasi *Clustering* Semester III

No	Cluster	IPK
1	Cluster 2	2.6
2	Cluster 0	3.88
3	Cluster 0	3.79
4	Cluster 1	3.26
5	Cluster 1	3.05
6	Cluster 0	3.83
7	Cluster 1	3.4
8	Cluster 2	2.67
9	Cluster 0	3.79
10	Cluster 0	3.64
11	Cluster 1	3.12
12	Cluster 1	3.45

Gambar 13. Grafik *Clustering* Semester III

Pada Gambar 13 *cluster* 2 berwarna hijau sebanyak 2 mahasiswa, *cluster* 1 sebanyak 5 mahasiswa dan *cluster* 0 sebanyak 5 mahasiswa. Terjadi peningkatan nilai mahasiswa. Dari grafik terlihat bahwa nilai mahasiswa pada semester III mengalami kenaikan daripada semester II.

d. *Clustering* Data Semester IV

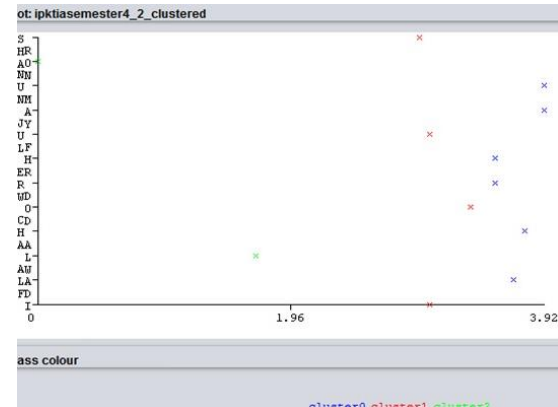
2: Nama 3: Semester 4 4: Cluster		
Nominal	Numeric	Nominal
ADITY...	3.04	cluster1
ALFI ...	3.69	cluster0
ALWA...	1.69	cluster2
CHA...	3.77	cluster0
DODI ...	3.35	cluster1
ERW...	3.54	cluster0
FHRE...	3.54	cluster0
JULIA...	3.04	cluster1
MAYA...	3.92	cluster0
NUN...	3.92	cluster0
RONAL	0.0	cluster2
SHAF...	2.96	cluster1

Gambar 14. Hasil *Clustering* Semester IV

Pada Gambar 14 adalah hasil *clustering* semester 1. Nilai SSE yang didapatkan adalah 0,108. Tabel 5 adalah evaluasi hasil *clustering* dari data Gambar 14.

Tabel 5. Evaluasi Hasil *Clustering*

No	Cluster	IPK
1	Cluster 1	3,04
2	Cluster 0	3,69
3	Cluster 2	1,69
4	Cluster 0	3,77
5	Cluster 1	3,35
6	Cluster 0	3,54
7	Cluster 0	3,54
8	Cluster 1	3,04
9	Cluster 0	3,92
10	Cluster 0	3,92
11	Cluster 2	0
12	Cluster 1	2,96

Gambar 15. Grafik *Clustering* Semester IV

Pada Gambar 15 *cluster* 2 berwarna hijau sebanyak 2 mahasiswa, *cluster* 1 sebanyak 4 mahasiswa dan *cluster* 0 sebanyak 6 mahasiswa. Pada semester IV juga terjadi peningkatan nilai mahasiswa daripada semester sebelumnya, dimana salah satu anggota *cluster* 1 berpindah ke *cluster* 0 walaupun ada mahasiswa yang memiliki nilai 0 dikarenakan mahasiswa yang cuti di semester IV.

Dari data kelas TI A angkatan 2017- 2018, metode *clustering* menggunakan K-Means bisa digunakan untuk mengelompokan mahasiswa.

3.2 Pembahasan

Dari penelitian yang dilakukan dapat disimpulkan bahwa variabel yang menghasilkan hasil *clustering* yang bagus adalah menggunakan jumlah variabel yang lebih sedikit. Penggunaan satu variabel, yaitu variabel IPK menghasilkan hasil *clustering* yang bagus dibandingkan variabel nilai-nilai dari setiap mata kuliah. Hasil SSE yang didapatkan dapat dilihat pada Gambar 16.

Tabel 6. Perbandingan SSE pada Seleksi Fitur

Fitur	SSE
Nilai setiap mata kuliah(12 variabel)	7,46
IPK	0,147



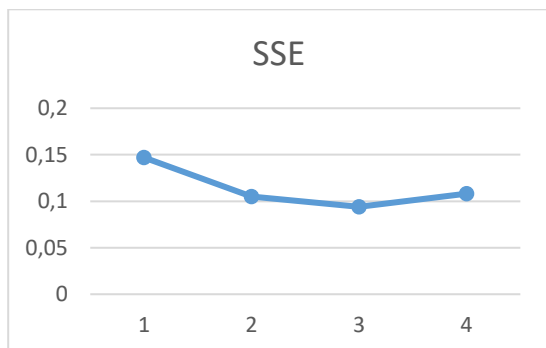
Gambar 16. Grafik Perbandingan SSE pada Fitur yang Digunakan

Dari penelitian dengan menggunakan data mahasiswa selama 4 semester kelas TI A, rata-rata SSE yang didapatkan sebesar 0.1135 dari 4 kali percobaan. Dengan rincian pada Tabel 7. SSE yang didapatkan pada 4 percobaan dapat dilihat pada Gambar 17, dimana hasil SSE yang didapatkan di setiap percobaan tidak jauh berbeda.

Tabel 7. Nilai SSE pada 4 kali Percobaan

No	Semester	SSE
1	1	0,147
2	2	0,105
3	3	0,094
4	4	0,108
Rata-rata		0,1135

Berikut Tabel 8 adalah analisa nilai mahasiswa hasil *clustering* semester I hingga semester 4.



Gambar 17. SSE pada 4 Percobaan *Clustering*

Tabel 8. Analisa Nilai Mahasiswa berdasar Hasil *Clustering*

Mhs ke-	S1	S2	S3	S4
1	2,93	2,93	2,6	3,04
2	3,18	3,85	3,88	3,69
3	3,33	3,45	3,79	1,69
4	3,3	3,58	3,26	3,77
5	2,95	3,4	3,05	3,35
6	3,4	3,6	3,83	3,54
7	3,13	3,48	3,4	3,54
8	2,9	3,1	2,67	3,04
9	3,88	3,78	3,79	3,92
10	3,5	3,63	3,64	3,92
11	3,08	3,13	3,12	0
12	2,8	2,93	3,45	2,96

Keterangan :

Hijau : Mahasiswa dengan nilai rendah

Merah : Mahasiswa dengan nilai sedang

Biru : Mahasiswa dengan nilai tinggi

Dapat disimpulkan bahwa nilai rata-rata mahasiswa kelas tersebut mengalami peningkatan. Namun apabila dievaluasi permahasiswa ada juga mahasiswa yang mengalami penurunan nilai. Mahasiswa dengan kenaikan *cluster* adalah : mahasiswa 1, mahasiswa 2, mahasiswa 4, mahasiswa 5, mahasiswa 6, mahasiswa 7, mahasiswa 8, mahasiswa 12. Sedangkan mahasiswa yang mengalami penurunan nilai adalah mahasiswa ke-3 dan mahasiswa ke-11. Mahasiswa yang memiliki nilai konstan selalu dalam *cluster* tinggi adalah mahasiswa ke-9 dan mahasiswa ke-10.

Dari hasil analisa melalui teknik *clustering* perguruan tinggi bisa melakukan evaluasi hasil belajar mahasiswa, sehingga dapat menentukan strategi-strategi untuk mendorong prestasi mahasiswa.

Mahasiswa dengan penurunan nilai bisa dilakukan bimbingan sehingga di semester berikutnya mahasiswa bisa meningkatkan prestasinya dan tetap termotivasi untuk melanjutkan kuliahnya. Begitu pula dengan mahasiswa dengan hasil prestasi yang

meningkat perlu terus dimotivasi supaya tetap meningkatkan prestasinya.

4. KESIMPULAN

Tahapan dalam penelitian ini terdiri dari *data collection*, *preprocessing*, seleksi fitur, *clustering*, dan analisis hasil *cluster*. Mahasiswa dikelompokkan menjadi 3 *cluster*, mahasiswa dengan nilai tinggi, mahasiswa dengan nilai sedang dan mahasiswa dengan nilai rendah. Pengelompokan menggunakan variabel IPK menghasilkan hasil *clustering* yang bagus dibandingkan pengelompokan dengan variabel nilai di setiap mata kuliah. Rata-rata SSE yang diperoleh dalam penelitian ini sebesar 0,1135 dari 4 kali percobaan.

Dari penelitian yang dilakukan dapat disimpulkan bahwa hasil belajar mahasiswa AKN Pacitan rata-rata setiap semesternya mengalami peningkatan. Metode *clustering* menggunakan metode K-Means bisa digunakan untuk mengelompokkan nilai mahasiswa dan mengevaluasi hasil belajar mahasiswa setiap semesternya sehingga memungkinkan apabila suatu sekolah atau perguruan tinggi untuk mengembangkan sistem informasi evaluasi mahasiswa dengan teknik *clustering* menggunakan metode K-Means.

5. DAFTAR PUSTAKA

- [1] Mardi, Yuli. "Data Mining : Klasifikasi Menggunakan Algoritma C4.5," Jurnal Edik Informatika, Vol.2, No.2, hal.213-219, 2017.
- [2] Rosmini, Abdul, F dan Sunardi. "Implementasi Metode K-Means Dalam Pemetaan Kelompok Mahasiswa Melalui Data Aktivitas Kuliah," IT Journal Research and Development, Vol.3, No.1, hal.22-31, 2018
- [3] Abdurrahman, G. "*Clustering* Data Ujian Tengah Semester (UTS) Data Mining Menggunakan Algoritma K-Means," Jurnal Sistem dan Teknologi Informasi Indonesia, Vol.1, No.2, hal.71-79, 2016.
- [4] Suprawoto, Totok. "Klasifikasi Data Mahasiswa Menggunakan Metode K-Means untuk Menunjang Pemilihan Strategi Pemasaran," Jurnal Informatika dan Komputer, Vol.1, No.1, Hal.12-18, 2016.
- [5] A.E. Rahayu, K. Hikmah, N.Y. Ningsih, A.C. Fauzan. "Penerapan K-Means *Clustering* Untuk Penentuan Klasterisasi Beasiswa Bidikmisi Mahasiswa". ILKOMNIKA: Journal of Computer Science and Applied Informatics, Vol.1, No.2, Hal.82-86, Desember 2019.
- [6] R.A. Haraty, M. Dimishkieh, M. Masud. "An Enhanced K-Means *Clustering* Algorithm for Pattern Discovery in Healthcare Data". SAGE Journal, Vol. 11, Issue. 6, pp.1-11, Juni 2015.
- [7] S. Abadi, dkk. "Application Model Of K-Means *Clustering*: Insights Into Promotion Strategy Of Vocational High School". International Journal of Engineering & Technology, Vol.7(2.27), Hal.182-187, 2018.
- [8] A.S. Ahmar, D. Napitupulu, R. Rahim, R. Hidayat, Y. Sonatha dan M. Azmi. "Using K-Means *Clustering* to Cluster Provinces in Indonesia". 2nd International Conference on Statistics, Mathematics, Teaching, and Research, 2017, 9-10 October 2017.
- [9] A.H. Fajar, "Pendahuluan," dalam *Data Mining*, edisi ke-I, Yogyakarta, Indonesia, 2013, bab I, bagian 1.1 – 1.5., hal. 3-16.
- [10] M.M. Benri dan L.S. Herlina "Analisis *Clustering* Menggunakan Metode K-Means dalam Pengelompokan Penjualan Produk pada Swalayan Fadhila," Jurnal Media Infotama, vol.11, No.2, hal.110-118, September, 2015.
- [11] S. J. Fitri, "Penerapan Support Vector Machine (SVM) untuk Pengkategorian Penelitian," Jurnal Resti, vol.1, hal.19-25. 2017.

- [12] M. Popy, "Penerapan Data Mining dengan Metode Klasifikasi Menggunakan Decision Tree dan Regresi," *Jurnal Teknologi*, vol.7, No.1, Januari, 2015.
- [13] A. Luh, "Seleksi Fitur dalam Klasifikasi Genre Music," *Jurnal Ilmiah Ilmu Komputer*, vol.X, No.1, hal 19-26, April, 2017.
- [14] A.P. I Putu dan H. Agus, "Penerapan Algoritma Invasive Weed Optimnization untuk Penentuan Titik Pusat Klaster pada K-Means," *IJCCS*, vol.9, No.1, hal 65-76, Januari, 2015.
- [15] Jaroji, Danuri dan P.P. Fajri, "K-Means untuk Menentukan Calon Penerima Beasiswa Bidik Misi di POLBENG," *Jurnal Inovtek POLBENG- Seri Informatika*, vol.1, No.1, hal 87-94, Juni, 2016.