



Kampus  
Merdeka  
INDONESIA JAYA



# SERTIFIKAT

Nomor: B/626/UN16.19/PT/2023

DIBERIKAN KEPADA:

*Rahmi Eka Putri*

SEBAGAI

Peserta dalam kegiatan Konferensi Nasional Klaster dan Hilirisasi Riset Berkelanjutan (KNKHRB) IX, dengan judul Penerapan Data Mining Dengan Algoritma K-Means Clustering Untuk Monitoring Studi Mahasiswa

Padang, 19-24 Desember 2023

Ketua LPPM Universitas Andalas



Dr.-Ing. Ir. Uyung Gatot S. Dinata, M.T.

NIP. 196607091992031003



**LAPORAN AKHIR  
SKIM RISET TERAPAN UNIVERSITAS ANDALAS  
TAHUN 2023**

**SUB TEMA PENELITIAN : INOVASI TEKNOLOGI DAN INDUSTRI**

**Sub Topik Penelitian : Produksi Teknologi Informatika Pendukung Tema Tema Utama  
UNAND**

**JUDUL PENELITIAN :**

**Penerapan Data Mining Dengan Algoritma K-Means Clustering  
Untuk Monitoring Studi Mahasiswa**

**TIM PENGUSUL**

**Ketua : Rahmi Eka Putri, MT (NIDN : 0023078402)**

**Anggota : Dwi Welly Sukma Nirad, MT (NIDN : 1012089101)**

**Mahasiswa : 1. Muhammad Zaki Al Hafiz (Nobp : 2211533004 )**

**Mahasiswa 2. Khoiri Putra Mujiza (Nobp : 2211537001)**

**FAKULTAS TEKNOLOGI INFORMASI  
UNIVERSITAS ANDALAS  
PADANG  
2023**

**HALAMAN PENGESAHAN**  
**LAPORAN AKHIR RISET TERAPAN UNIVERSITAS ANDALAS**

---

Judul Penelitian : Penerapan Data Mining Dengan Algoritma K-Means Clustering Untuk Monitoring Studi Mahasiswa

Skim : Riset Terapan  
Sub Tema Penelitian : Inovasi Teknologi Dan Industri  
Sub Topik Penelitian : Produksi Teknologi Informatika Pendukung Tema Tema Utama UNAND

Ketua Peneliti  
a. Nama Lengkap : Rahmi Eka Putri, MT  
b. NIDN : 0023078402  
c. ID Sinta : 6021347  
d. ID Google Scholar : 18xzSDcAAAAJ  
e. ID Scopus : 57189262997  
f. Jabatan Fungsional : Lektor  
g. Prodi, Fak/PPs : Departemen Informatika Fakultas Tekonologi Informasi  
h. Nomor HP : 081266655700  
i. Alamat surel (e-mail) : rahmi@it.unand.ac.id

Anggota Peneliti  
a. Nama Lengkap : Dwi Welly Sukma Nirad, MT  
b. NIDN : 0019119102  
c. Prodi, Fak/PPs : Departemen Sistem Informasi Fakultas Teknologi Informasi

Anggota Mahasiswa 1  
a. Nama Lengkap : Muhammad Zaki Al Hafiz  
b. No BP : 2211533004  
c. Prodi, Fak/PPs : Departemen Informatika Fakultas Tekonologi Informasi

Anggota Mahasiswa 2  
a. Nama Lengkap : Khoiri Putra Mujiza  
b. No BP : 2211537001  
c. Prodi, Fak/PPs : Departemen Informatika Fakultas Tekonologi Informasi

Biaya Penelitian Keseluruhan : Rp 20.000.000,-

Biaya Luaran Tambahan : -

Padang, 18 Desember 2023  
Ketua Peneliti,



Dekan Fakultas Teknologi Informasi

Ahmad Syafruddin Indrapriyatna, MT  
NIP 196307071991031003

Rahmi Eka Putri, MT  
NIP 198407232008122001

## IDENTITAS DAN URAIAN UMUM

1. Judul Penelitian : Penerapan Data Mining Dengan Algoritma K-Means Clustering Untuk Monitoring Studi Mahasiswa

2. Tim Peneliti :

No	Nama	Jabatan	Bidang Keahlian	Fak/PPs	Alokasi Waktu (jam/minggu)
1	Rahmi Eka Putri, MT	Ketua	Informatika	FTI Unand	40 jam
2	Dwi Welly Sukma Nirad, MT	Anggota	Sistem Informasi	FTI Unand	30 jam
3	Muhammad Zaki Al Hafiz	Mahasiswa 1	Informatika	FTI Unand	20 jam
4	Khoiri Putra Mujiza	Mahasiswa 2	Informatika	FTI Unand	20 jam

3. Objek Penelitian (jenis objek/material yang akan diteliti dan tema/topik penelitian) : Mahasiswa Fakultas Teknologi Informasi

4. Masa Pelaksanaan:

Mulai : bulan September tahun 2023

Berakhir : bulan Desember tahun 2023

5. Usulan Biaya ke Fakultas Teknologi Informasi : Rp. 20.000.000,-

6. Lokasi Penelitian (lab/studio/lapangan) : Labor Informatika Fakultas Teknologi Informasi

7. Instansi lain yang terlibat (jika ada, dan uraikan apa kontribusinya) : -

8. Produk temuan yang ditargetkan (produk, model, metode, teori, produk, atau kebijakan) : kebijakan

9. Kontribusi mendasar pada suatu bidang ilmu (uraikan tidak lebih dari 50 kata, tekankan pada gagasan fundamental dan orisinal yang akan mendukung pengembangan IPTEKS) : Penelitian ini akan memberikan kontribusi untuk memonitoring studi mahasiswa melalui pengelompokkan hasil studi mahasiswa berdasarkan Indeks Prestasi Mahasiswa sehingga dapat dibuat kebijakan

untuk membantu mahasiswa untuk dapat menyelesaikan studinya tepat waktu.

10. Kontribusi pada pencapaian RIP dan *roadmap* sub tema penelitian Unand (uraian sedikitnya 2 paragraf)

Kontribusi penelitian ini pada pencapaian RIP Unand selaras dengan sub topik pengembangan kebijakan / regulasi, manajemen, teknologi, bisnis, dan sosial budaya penerapan produk teknologi informatika pendukung tema utama Unand. Penelitian ini fokus untuk mendukung topik penelitian produksi dan penerapan teknologi informatika pendukung tema-tema Unand. Luaran tahap akhir dari *roadmap* penelitian akan menghasilkan kelompok mahasiswa berdasarkan Indeks Prestasi mahasiswa sehingga dapat dibuat kebijakan untuk membantu mempercepat mahasiswa untuk dapat menyelesaikan studi tepat waktu di Fakultas Teknologi Informasi Unand.

11. Jurnal ilmiah atau prosiding seminar yang menjadi sasaran (tuliskan nama jurnal ilmiah atau seminar internasional bereputasi terindeks Scopus dan tahun rencana publikasi)

Rencana Luaran penelitian : *International Symposium on Information Technology and Digital Innovation (ISITDII)* tahun 2024 (terindeks Scopus)

12. Rencana luaran *draft* HKI, *draft* buku, prototipe, rekayasa sosial atau luaran lainnya yang ditargetkan, tahun rencana perolehan atau penyelesaiannya : *draft* buku tahun 2023

## DAFTAR ISI

HALAMAN PENGESAHAN.....	ii
IDENTITAS DAN URAIAN UMUM.....	iii
DAFTAR ISI.....	v
BAB I PENDAHULUAN.....	1
BAB 2. RENCANA INDUK DAN PETA JALAN PENELITIAN UNIVERSITAS ANDALAS ....	3
BAB 3. TINJAUAN PUSTAKA .....	4
BAB 4. METODE PENELITIAN.....	6
BAB 5. HASIL DAN PEMBAHASAN.....	9
5.1. HASIL .....	9
5.1.1 Departemen Sistem Informasi.....	9
5.1.2 Departemen Teknik Komputer .....	14
5.2. ANALISA.....	19
BAB 6. KESIMPULAN.....	20
DAFTAR PUSTAKA .....	21

## BAB I PENDAHULUAN

Proses perkuliahan yang dilaksanakan oleh mahasiswa setiap semester nya harus dapat dimonitor dengan baik untuk memastikan mahasiswa dapat menyelesaikan studinya tepat waktu. Fakultas maupun Departemen / Program Studi memerlukan informasi untuk dapat membuat kebijakan yang tepat guna mendorong percepatan mahasiswa menyelesaikan studinya. Salah satu parameter yang dapat digunakan adalah Indeks Prestasi mahasiswa pada setiap semester. Indeks Prestasi mahasiswa akan dikelompokkan berdasarkan kriteria tertentu, sehingga akan diperoleh kelompok mahasiswa yang dapat menyelesaikan studi tepat waktu maupun yang terlambat studinya.

Metode untuk pengelompokkan Indeks Prestasi mahasiswa adalah dengan penerapan data mining. Teknik data mining dapat mengolah data yang berlimpah menjadi informasi yang penting biasanya disebut knowledge discovery database (KDD). Adapun metode yang digunakan dalam pengelompokan hasil studi mahasiswa adalah metode clustering.

Menurut Narwati (2010) clustering merupakan cara untuk menemukan kelompok objek yang memiliki kemiripan dan dapat menemukan pola penyebaran dan pola hubungan dalam kumpulan data yang besar. Dalam proses clustering yang terpenting adalah mengumpulkan pola ke kelompok yang sesuai untuk menemukan persamaan dan perbedaan agar menghasilkan kesimpulan yang berharga. Metode clustering menggunakan algoritma kmeans dalam pengelompokan data hasil studi mahasiswa.

Metode k-means cluster analysis bisa menjadi solusi untuk pengklasifikasian karakteristik dari objek. Algoritma k-means memiliki ketelitian yang cukup tinggi terhadap ukuran objek, sehingga algoritma ini relatif lebih terukur dan efisien untuk pengolahan objek dalam jumlah besar. Selain itu algoritma k-means tidak terpengaruh dengan adanya urutan objek. (Aranda , 2016).

Pada penelitian ini akan diterapkan *data mining* terhadap Indeks Prestasi mahasiswa dengan menggunakan algoritma K-Means Clustering. Dengan pengelompokkan hasil studi mahasiswa yang didapatkan, diharapkan Fakultas maupun Departemen dapat membuat kebijakan untuk dapat membantu mahasiswa menyelesaikan studinya tepat waktu.

Adapun rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana penerapan data mining untuk pengelompokkan Indeks Prestasi mahasiswa dengan menggunakan algoritma K-Means Clustering
2. Bagaimana melakukan evaluasi terhadap model kluster yang dihasilkan menggunakan metode Davis Bouldin Index (DBI)

Rencana capaian penelitian ini dapat dilihat pada Tabel 1.1 sesuai dengan luaran yang ditargetkan serta lamanya penelitian yang telah dan akan dilakukan.

Tabel 1.1 Luaran penelitian

No	Jenis Luaran				Indikator Capaian		
	Kategori	Sub Kategori	Wajib	Tambahan	TS <sup>1)</sup>	TS+1	TS+2
1.	Artikel ilmiah dimuat di jurnal	Internasional bereputasi					
		Nasional bereputasi					
2.	Artikel ilmiah dimuat di prosiding	Internasional terindeks	Submitted			√	
		Nasional					
3.	Invited speaker dalam temu ilmiah	Internasional					
		Nasional					
4.	Visiting Lecturer	Internasional					
5.	Hak Kekayaan Intelektual (HKI)	Paten					
		Paten sederhana					
		Hak Cipta					
		Merek Dagang					
		Rahasia Dagang					
		Desain produk industri					
		Indikasi Geografis					
		Perlindungan Varietas Tanaman					
6.	Teknologi Tepat Guna	Perlindungan Topografi					
		Sirkuit terpadu					
7.	Model/Purwarupa/Desain/Karya Seni/ Rekayasa Sosial						
8.	Bahan Ajar		Draft		√		
9.	Tingkat Kesiapan Teknologi (TKT)						

## BAB 2. RENCANA INDUK DAN PETA JALAN PENELITIAN UNIVERSITAS ANDALAS

Salah satu tema utama dari Rencana Induk Penelitian (RIP) Unand adalah bidang inovasi sains, teknologi, dan industri sebagaimana terlihat pada Gambar 2.1.

No.	Tema Utama Penelitian	Sub-Tema Penelitian	Topik Penelitian	Sub-Topik Penelitian	Tahapan (Pokok Bahasan Penelitian)	Luaran Topik Penelitian	Luaran Tema Utama Penelitian	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	
2	INOVASI SAINS, TEKNOLOGI, DAN INDUSTRI	1. Inovasi sains	1. Diversitas dan ekologi sumber daya hayati tropika serta kelestarian lingkungan	Diversitas sumber daya hayati, ekologi sumber daya hayati, kelestarian lingkungan, dan dimensi sosial kelestarian lingkungan	Penelitian, kajian, evaluasi, penerapan, pengembangan, inovasi, operasi dan/atau produksi SDH dan lingkungan	Database dan kebijakan/regulasi keanekaragaman hayati tropika dan menjaga kelestarian lingkungan	Kontribusi Unand pada pembangunan nasional dan daerah serta IPTEK melalui inovasi sains untuk pengelolaan sumber daya hayati dan lingkungan serta ilmu-ilmu terapan pendukung tiga tema utama Unand	
				Pengembangan matematika terapan, fisika terapan, kimia terapan, dan biologi terapan untuk mendukung tiga tema utama Unand	Penelitian, kajian, evaluasi, penerapan, inovasi, operasi dan/atau produksi ilmu-ilmu terapan pendukung tiga tema utama Unand	Berbagai perangkat ilmu terapan untuk mendukung tiga tema utama Unand		
			2. Inovasi mitigasi bencana	Mitigasi bencana (pm, ssat dan pascabencana)	Pengembangan kebijakan/regulasi, ekonomi, manajemen, teknologi dan bisnis, dan sosial budaya untuk mendukung penerapan mitigasi bencana alam	Penelitian, kajian, evaluasi, penerapan, pengembangan, inovasi, operasi dan/atau produksi pendukung mitigasi bencana	Kebijakan/regulasi, sistem, manajemen dan teknologi serta pengetahuan sosial budaya terkait kebencanaan yang mendukung ekonomi daerah dan menjadi rujukan internasional	Kebijakan/regulasi, sistem, manajemen dan teknologi untuk pengembangan dan peningkatan pemanfaatan energi baru dan terbarukan secara nasional berorientasi ekonomi untuk ketahanan energi dan lingkungan
					Pengembangan kebijakan/regulasi, ekonomi, manajemen, teknologi dan bisnis, dan sosial budaya untuk mendukung penerapan mitigasi bencana alam	Penelitian, kajian, evaluasi, penerapan, pengembangan, inovasi, operasi dan/atau produksi pendukung mitigasi bencana	Kebijakan/regulasi, sistem, manajemen dan teknologi untuk pengembangan dan peningkatan pemanfaatan energi baru dan terbarukan secara nasional berorientasi ekonomi untuk ketahanan energi dan lingkungan	
			3. Inovasi teknologi dan industri	1. Konservasi energi, serta konservasi dan produksi energi baru dan energi terbarukan (air, angin, surya, laut, bioenergi, panas bumi, dsb.)	Pengembangan kebijakan/regulasi, ekonomi, manajemen, teknologi dan bisnis, dan sosial budaya penerapan konservasi energi, serta konservasi dan produksi energi baru dan energi terbarukan	Penelitian, kajian, evaluasi, penerapan, pengembangan, inovasi, operasi dan/atau produksi pendukung energi baru dan energi terbarukan	Kebijakan/regulasi, sistem, manajemen dan teknologi untuk produksi dan pemanfaatan bahan alam (berbasis gambir, karet, bambu, dsb.) dan siku esdang industri	Produk-produk teknologi informasi terapan yang mendukung tema utama Unand
					2. Produksi dan penerapan bahan majo alam (berbasis gambir, sawit, karet, bambu, dsb.) dan siku esdang industri	Pengembangan kebijakan/regulasi, ekonomi, manajemen, teknologi, bisnis, dan sosial budaya pendukung produksi dan penerapan bahan majo alam dan siku esdang industri		
		3. Produksi dan penerapan teknologi informatika pendukung tema-tema Unand	4. Infrastruktur dan teknologi, produksi dan penerapan produk berbasis keilmuan	Pengembangan kebijakan/regulasi, manajemen, teknologi, bisnis, dan sosial budaya penerapan produk-produk teknologi informatika pendukung tema-tema Unand	Penelitian, kajian, evaluasi, penerapan, pengembangan, inovasi, operasi dan/atau produksi pendukung tema-tema Unand	Kebijakan/regulasi, sistem, manajemen dan teknologi untuk mendukung infrastruktur, teknologi dan produksi berbasis keilmuan	Kebijakan/regulasi, sistem, manajemen dan teknologi untuk mendukung infrastruktur, teknologi dan produksi berbasis keilmuan	
				Pengembangan kebijakan/regulasi, manajemen, teknologi, bisnis, dan sosial budaya pendukung infrastruktur, produksi dan penerapan produk-produk berbasis keilmuan	Penelitian, kajian, evaluasi, penerapan, pengembangan, inovasi, operasi dan/atau produksi pendukung produk berbasis keilmuan			

Gambar 2.1 Rencana Induk Penelitian Unand

Berdasarkan RIP tersebut, penelitian ini fokus untuk mendukung topik penelitian produksi dan penerapan teknologi informatika pendukung tema-tema Unand. Luaran tahap akhir dari roadmap penelitian akan menghasilkan kelompok mahasiswa berdasarkan Indeks Prestasi mahasiswa sehingga dapat dibuat kebijakan untuk membantu mempercepat mahasiswa untuk dapat menyelesaikan studi tepat waktu di Fakultas Teknologi Informasi Unand.

## **BAB 3. TINJAUAN PUSTAKA**

### **3.1. Data Mining**

Data mining adalah suatu metode pengolahan data untuk menemukan pola yang tersembunyi dari data tersebut. Hasil dari pengolahan data dengan metode data mining ini dapat digunakan untuk mengambil keputusan di masa depan. Data mining ini juga dikenal dengan istilah pattern recognition (Santosa, 2007).

Data mining merupakan metode pengolahan data berskala besar oleh karena itu data mining ini memiliki peranan penting dalam bidang industri, keuangan, cuaca, ilmu dan teknologi. Secara umum kajian data mining membahas metode-metode seperti, clustering, klasifikasi, regresi, seleksi variable, dan market basket analisis (Santosa, 2007).

### **3.2. Clustering**

Pada dasarnya clustering merupakan suatu metode untuk mencari dan mengelompokkan data yang memiliki kemiripan karakteristik (similarity) antara satu data dengan data yang lain. Clustering merupakan salah satu metode data mining yang bersifat tanpa arahan (unsupervised), maksudnya metode ini diterapkan tanpa adanya latihan (training) dan tanpa ada guru (teacher) serta tidak memerlukan target output. Dalam data mining ada dua jenis metode clustering yang digunakan dalam pengelompokan data, yaitu hierarchical clustering dan non-hierarchical clustering (Santosa, 2007).

### **3.3. K-means Clustering**

K-means clustering merupakan salah satu metode data clustering non-hirarki yang mengelompokkan data dalam bentuk satu atau lebih cluster/kelompok. Data-data yang memiliki karakteristik yang sama dikelompokkan dalam satu cluster/kelompok dan data yang memiliki karakteristik yang berbeda dikelompokkan dengan cluster/kelompok yang lain sehingga data yang berada dalam satu cluster/kelompok memiliki tingkat variasi yang kecil (Agusta, 2007).

Menurut Santosa (2007), langkah-langkah melakukan clustering dengan metode K-Means adalah sebagai berikut :

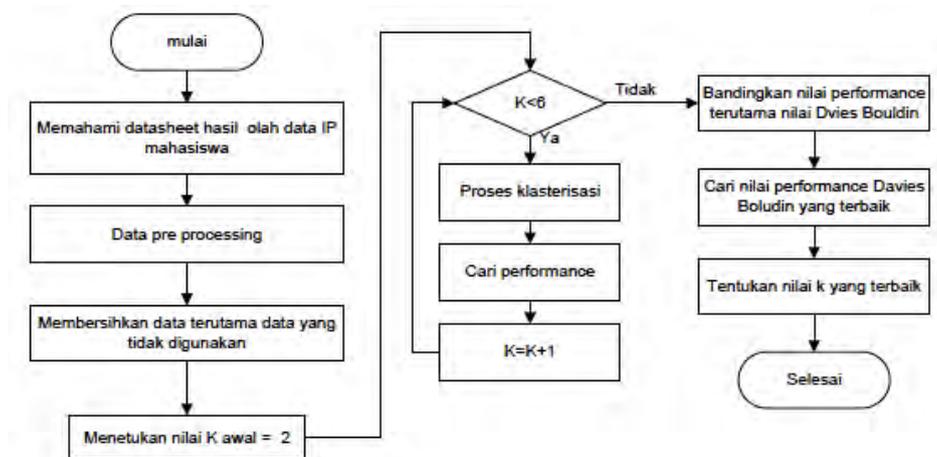
1. Pilih jumlah cluster K.
2. Inisialisasi k pusat cluster ini bisa dilakukan dengan berbagai cara. Namun yang paling

sering dilakukan adalah dengan cara random. Pusat-pusat cluster diberi nilai awal dengan angka-angka random,

3. Alokasikan semua data/ objek ke cluster terdekat. Kedekatan dua objek ditentukan berdasarkan jarak kedua objek tersebut. Demikian juga kedekatan suatu data ke cluster tertentu ditentukan jarak antara data dengan pusat cluster. Dalam tahap ini perlu dihitung jarak tiap data ke tiap pusat cluster. Jarak paling antara satu data dengan satu cluster tertentu akan menentukan suatu data masuk dalam cluster mana.
4. Hitung kembali pusat cluster dengan keanggotaan cluster yang sekarang. Pusat cluster adalah rata-rata dari semua data/ objek dalam cluster tertentu. Jika dikehendaki bisa juga menggunakan median dari cluster tersebut. Jadi rata-rata (mean) bukan satu-satunya ukuran yang bisa dipakai.
5. Tugaskan lagi setiap objek memakai pusat cluster yang baru. Jika pusat cluster tidak berubah lagi maka proses clustering selesai. Atau, kembali ke Langkah nomor 3 sampai pusat cluster tidak berubah lagi.

## BAB 4. METODE PENELITIAN

Metodologi penelitian dilakukan dengan melakukan proses olah data yang menjadi proses pengolahan data mining model clustering. Datasheet diambil dari kumpulan data IP mahasiswa mulai dari semester 1 sampai 6. Proses penelitian disajikan pada gambar 4.1.



Gambar 4.1. Proses Tahapan Dalam Penelitian

### 4.1. Teknik Pengumpulan Data

#### 1. Observasi Metode

Metode observasi digunakan untuk pengumpulan data. Data diolah dan diambil portal akademik dan dilakukan proses pembersihan data.

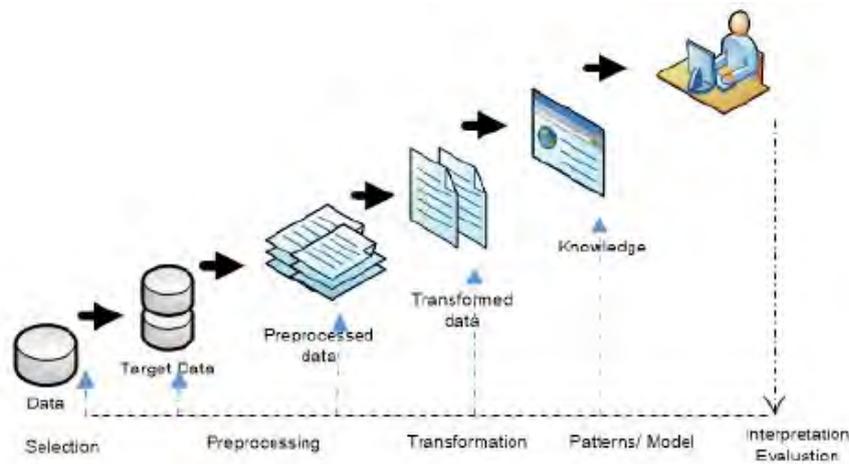
#### 2. Studi Literatur

Studi literatur dilakukan untuk memperoleh data yang sesuai untuk penelitian. Cara ini dilakukan dengan cara mencari dan memahami literatur dari beberapa jurnal yang berkaitan proses pembuatan model clustering

### 4.2. Teknik Analisis Data

Dalam melakukan penelitian ini metode analisis data menggunakan metode KDD. KDD merupakan aktivitas yang mencakup pengumpulan, pemakaian data historis buat memilih keteraturan, pola atau interaksi pada sebuah set data yang ukuran besar. Keluaran

berdasarkan data mining pola dipakai buat dalam proses penentuan keputusan. Gambar 4.2. proses KDD.



Gambar 4.2. Tahapan dalam Metode Knowledge Discovery in Database

Pada penelitian ini digunakan metode clustering K-Means untuk mengelompokkan Indeks Prestasi mahasiswa menjadi beberapa kelompok

#### 1. Selection

Pada proses ini, adalah menyeleksi atribut yang dipakai untuk proses selanjutnya. Hanya atribut yang relevan yang digunakan untuk dianalisis.

#### 2. Preprocessing

Tahap preprocessing ialah tahap pembersihan data dari missing value, noise, outlier, dan regresi. Data yang sudah di seleksi atributnya bisa saja bagiannya tidak lengkap bahkan ada yang hilang. Maka dari itu perlu adanya proses pembersihan atau biasa disebut dengan data cleaning.

#### 3. Transformation

Di tahap transformasi yaitu mengubah tipe data yang masih berbentuk teks menjadi data numerik. Karena dalam penggunaan metode K-Means clustering yang digunakan bisa diolah dengan data dalam bentuk numerik. Sehingga data yang belum berbentuk numerik harus diubah terlebih dahulu. Pada tahap ini penulis menggunakan operator pada RapidMiner untuk mentransormasi data secara otomatis.

#### 4. Data Mining

Ditahap ini dimasukkannya metode data mining yang akan digunakan, dalam penelitian ini

menggunakan K-Means.

#### 5. Evaluasi

Ditahap terakhir yaitu evaluasi. Dimana hasil dari proses yang sudah dirunning mendapatkan berapa cluster. Setelah muncul hasilnya, ditranslasikan ke dalam bentuk kalimat, agar mudah dimengerti oleh orang lain.

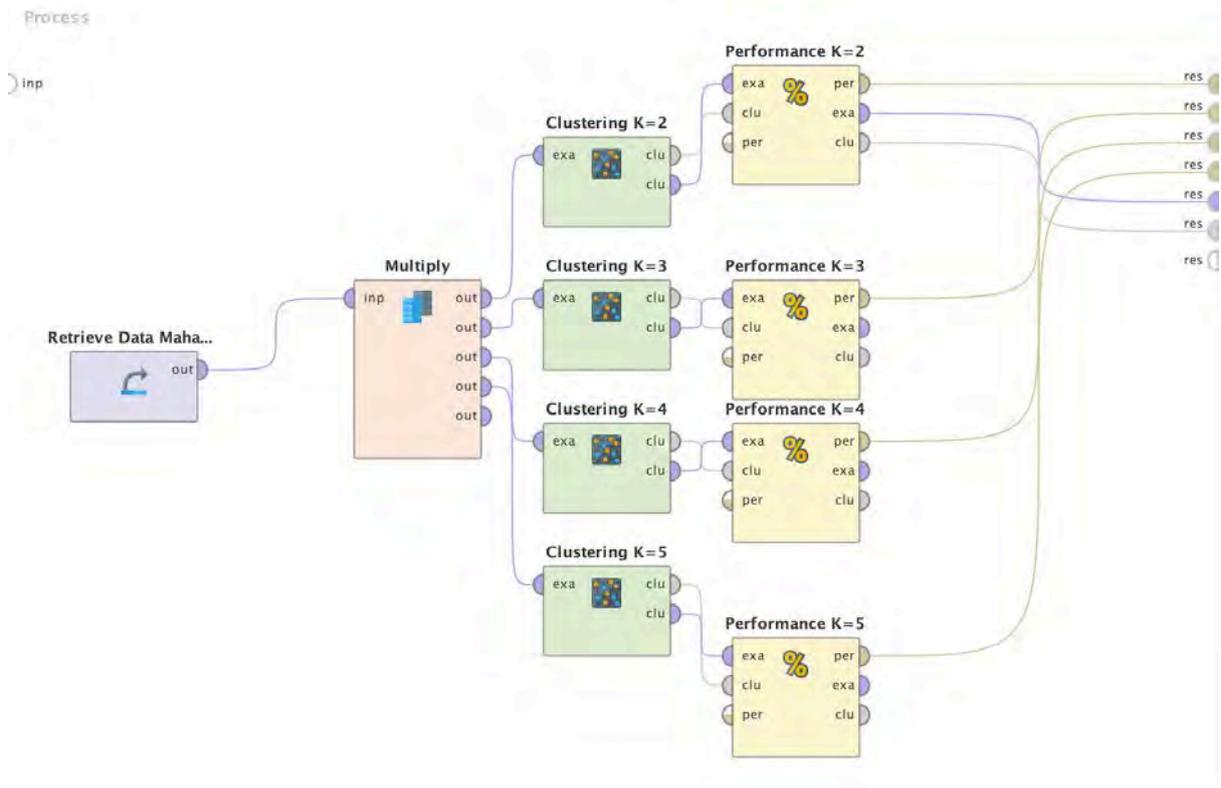
## **BAB 5. HASIL DAN PEMBAHASAN**

### **5.1. HASIL**

Penelitian ini dilakukan dengan mengelompokkan data Indeks Prestasi (IP) mahasiswa di Fakultas Teknologi Informasi pada dua departemen yaitu Departemen Sistem Informasi dan Departemen Teknik Komputer. Data Indeks Prestasi yang digunakan adalah mahasiswa angkatan 2020 mulai dari semester satu sampai dengan semester enam. Pengelompokkan dilakukan untuk menghasilkan kelompok mahasiswa berdasarkan Indeks Prestasi sehingga dapat dirumuskan strategi dalam upaya untuk membantu mahasiswa dapat lulus tepat waktu, sehingga jumlah mahasiswa yang dapat lulus tepat waktu di Fakultas Teknologi Informasi dapat meningkat.

#### **5.1.1 Departemen Sistem Informasi**

Penerapan data mining dengan algoritma k-means clustering dilakukan terhadap data Indeks Prestasi mahasiswa Sistem Informasi angkatan 2020 mulai dari semester satu sampai dengan semester enam. Tahap awal yang dilakukan adalah mempersiapkan data Indeks Prestasi mahasiswa dengan atribut IP1, IP2, IP3, IP4, IP5, dan IP6. Langkah selanjutnya adalah data mining dengan algoritma k-means clustering dengan jumlah kluster yang digunakan adalah  $K = 2$  sampai  $K < 6$ . Model clustering untuk data Indeks Prestasi mahasiswa Departemen Sistem Informasi dapat dilihat pada gambar 5.1.



Gambar 5.1 Model clustering Indeks Prestasi mahasiswa Departemen Sistem Informasi

Evaluasi terhadap model yang telah dihasilkan dilakukan untuk menemukan cluster terbaik dalam proses pengelompokan data yaitu dengan menggunakan Davies Bouldin Index (DBI), dimana cluster yang terbaik adalah nilai evaluasi yang paling kecil. Perbandingan nilai Davies Bouldin untuk setiap cluster dapat dilihat pada tabel 5.1

Tabel 5.1 Perbandingan nilai Davies Bouldin untuk setiap cluster

Jumlah Kluster	Nilai Davies Bouldin
2	- 0,852
3	- 0,577
4	- 0,375
5	- 0,309

Berdasarkan tabel 5.1 dapat dilihat bahwa nilai terkecil dari Davies Bouldin adalah - 0,852 dengan K= 2. Sehingga kluster terbaik adalah dengan K = 2. Hasil dari model K-Means yang dihasilkan dapat dilihat pada gambar 5.2.

## Cluster Model

Cluster 0: 71 items  
Cluster 1: 5 items  
Total number of items: 76

Gambar 5.2 Hasil model k-means clustering

Dari penelitian yang dilakukan, diperoleh 2 cluster untuk mahasiswa angkatan 2020 di Departemen Sistem Informasi, dimana cluster 0 berjumlah 71 mahasiswa dan cluster 1 berjumlah 5 mahasiswa. Data untuk cluster 0 dapat dilihat pada gambar 5.3 dan data untuk cluster 1 dapat dilihat pada gambar 5.4.



Gambar 5.3 Data cluster 0



Gambar 5.4 Data cluster 1

Data cluster dengan IP1, IP2, IP3, IP4, IP5, dan IP6 dapat dilihat pada gambar 5.5. Data centroid untuk masing-masing IP untuk setiap cluster pada gambar 5.6 menunjukkan titik tengah data Indeks Prestasi mahasiswa angkatan 2020 di Departemen Sistem Informasi.

Row No.	NAMA MAH...	cluster	IP1	IP2	IP3	IP4	IP5	IP6
1	TAUFIK HID...	cluster_0	3.700	3.460	3.210	2.260	3.750	3.410
2	BAITUL AZIZ...	cluster_0	1.470	3.610	3.840	3.720	3.630	3.880
3	DINA LATHI...	cluster_0	3.760	3.640	3.750	3.510	3.930	3.810
4	NADILLA SA...	cluster_0	3.920	3.680	3.950	3.590	3.980	3.920
5	ANNISA FARAS	cluster_0	3.750	3.540	3.290	3.360	3.430	3.890
6	REYSHA IRS...	cluster_0	3.960	3.780	4	3.850	3.970	3.980
7	RADITHYA R...	cluster_0	4	3.510	3.470	3.250	3.770	3.800
8	DESRILIA PU...	cluster_0	3.750	3.500	3.760	3.560	3.800	3.810
9	RISKA SHIFA...	cluster_0	4	3.960	3.960	3.630	3.880	3.580
10	DEAN FISABI...	cluster_0	3.960	3.740	4	3.660	3.930	3.790
11	FADILA TRI ...	cluster_0	3.790	3.750	3.730	3.390	3.770	3.850
12	LARA BUSYA...	cluster_0	3.880	3.580	3.880	3.280	3.890	3.780
13	PUTI JUGALO	cluster_0	3.990	3.790	3.880	3.350	3.700	3.880
14	THOMAS AK...	cluster_0	3.710	3.630	3.800	2.990	3.710	3.780

Gambar 5.5 Data cluster dengan IP1, IP2, IP3, IP4, IP5, dan IP6

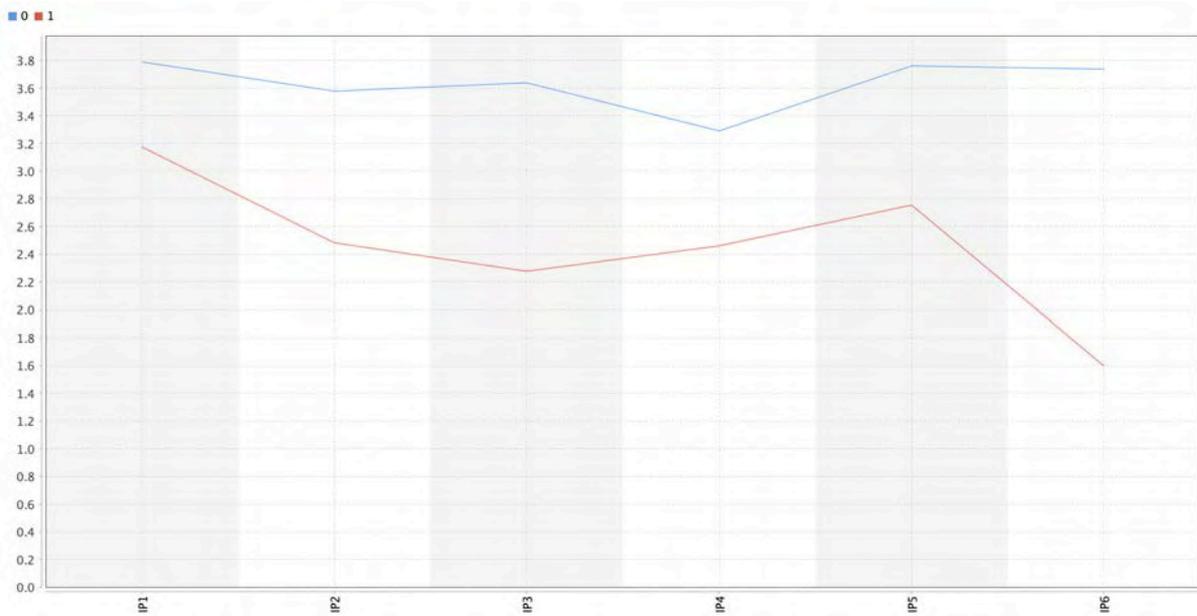
Attribute	cluster_0	cluster_1
IP1	3.788	3.174
IP2	3.577	2.482
IP3	3.637	2.276
IP4	3.290	2.460
IP5	3.759	2.754
IP6	3.736	1.596

Gambar 5.6 data centroid

Nilai minimum, maksimum dan rata-rata Indeks Prestasi mahasiswa pada setiap cluster dapat dilihat pada gambar 5.7. Dari hasil data mining dengan algoritma k-means clustering yang diperoleh dapat dilihat statistik data Indeks Prestasi mahasiswa angkatan 2020 di Departemen Sistem Informasi seperti pada gambar 5.8.

Name	Type	Missing	Statistics		
Id NAMA MAHASISWA	Polynomial	0	Least ZUHA BIMA AL FARUQ ...	Most AINI IZZATHY ISPENDI ...	Values AINI IZZATHY ISPENDI (1), AKMAL RAH...
Cluster cluster	Nominal	0	Least cluster_1 (5)	Most cluster_0 (71)	Values cluster_0 (71), cluster_1 (5)
IP1	Real	0	Min 1.470	Max 4	Average 3.748
IP2	Real	0	Min 1.300	Max 3.960	Average 3.505
IP3	Real	0	Min 0.620	Max 4	Average 3.548
IP4	Real	0	Min 2.140	Max 3.880	Average 3.235
IP5	Real	0	Min 1.680	Max 3.980	Average 3.693
IP6	Real	0	Min 0	Max 4	Average 3.595

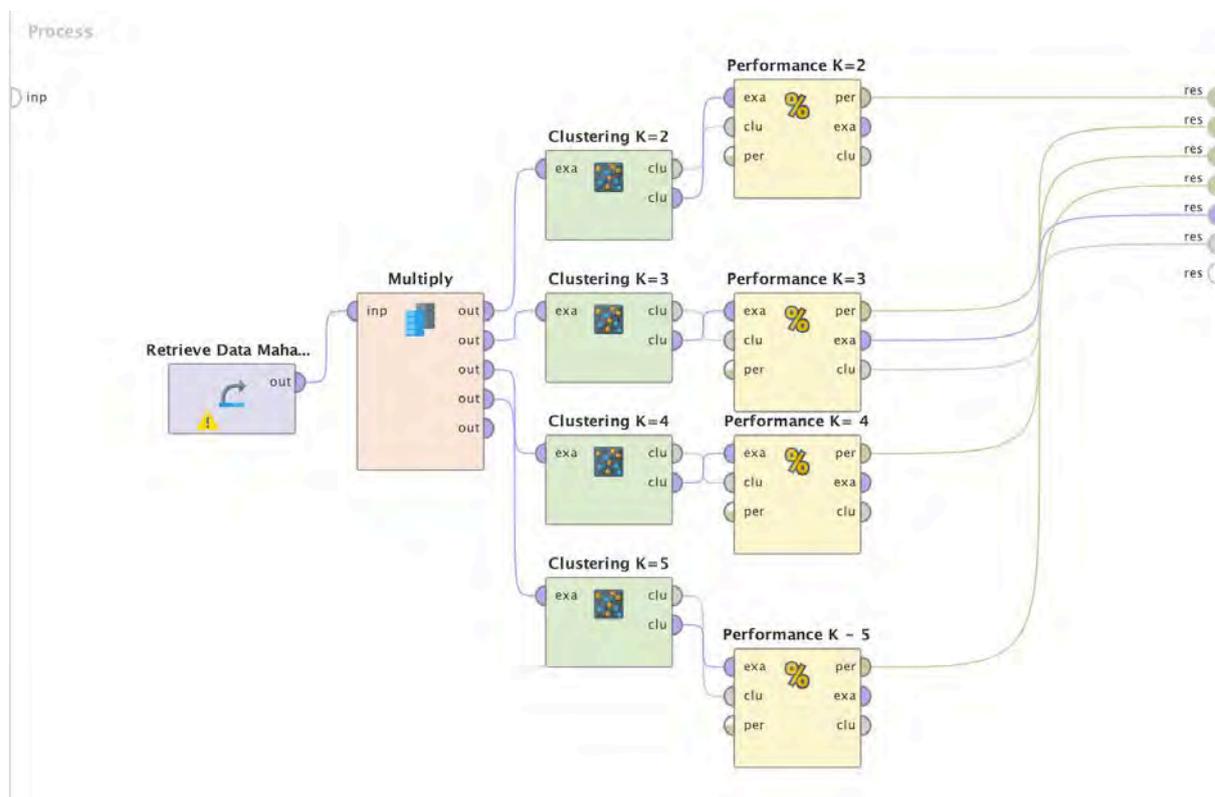
Gambar 5.7 Minimum, maksimum dan rata-rata IP untuk setiap cluster



Gambar 5.8 Statistik Indeks Prestasi mahasiswa Departemen Sistem Informasi

### 5.1.2 Departemen Teknik Komputer

Penerapan data mining dengan algoritma k-means clustering dilakukan terhadap data Indeks Prestasi mahasiswa Teknik Komputer angkatan 2020 mulai dari semester satu sampai dengan semester enam. Tahap awal yang dilakukan adalah mempersiapkan data Indeks Prestasi mahasiswa dengan atribut IP1, IP2, IP3, IP4, IP5, dan IP6. Langkah selanjutnya adalah data mining dengan algoritma k-means clustering dengan jumlah kluster yang digunakan adalah  $K = 2$  sampai  $K < 6$ . Model clustering untuk data Indeks Prestasi mahasiswa Departemen Sistem Informasi dapat dilihat pada gambar 5.9.



Gambar 5.9 Model clustering Indeks Prestasi mahasiswa Departemen Teknik Komputer

Evaluasi terhadap model yang telah dihasilkan dilakukan untuk menemukan cluster terbaik dalam proses pengelompokan data yaitu dengan menggunakan Davies Bouldin Index (DBI), dimana cluster yang terbaik adalah nilai evaluasi yang paling kecil. Perbandingan nilai Davies Bouldin untuk setiap cluster dapat dilihat pada tabel 5.2.

Tabel 5.2 Perbandingan nilai Davies Bouldin untuk setiap cluster

Banyak Kluster	Nilai Davies Bouldin
2	- 0,984
3	- 1,169
4	- 0,971
5	- 1,149

Berdasarkan tabel 5.2 dapat dilihat bahwa nilai terkecil dari Davies Bouldin adalah – 1,169 dengan  $K=3$ . Sehingga kluster terbaik adalah dengan  $K=3$ . Hasil dari model K-Means yang dihasilkan dapat dilihat pada gambar 5.10.

## Cluster Model

```
Cluster 0: 6 items  
Cluster 1: 40 items  
Cluster 2: 25 items  
Total number of items: 71
```

Gambar 5.10 Hasil model k-means clustering

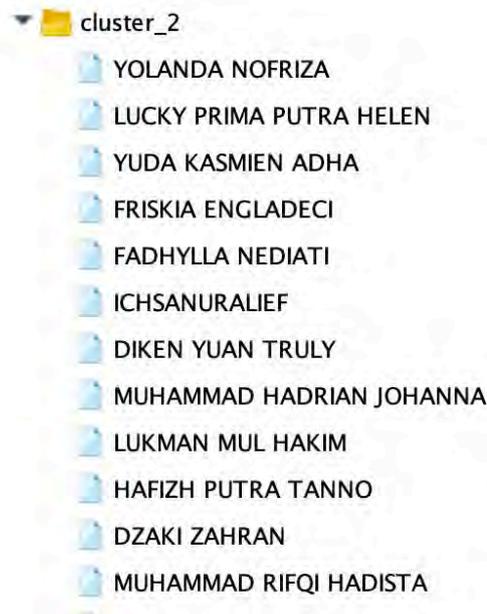
Dari penelitian yang dilakukan, diperoleh 2 cluster untuk mahasiswa angkatan 2020 di Departemen Teknik Komputer, dimana cluster 0 berjumlah 6 mahasiswa, cluster 1 berjumlah 40 mahasiswa dan cluster 2 berjumlah 25 mahasiswa. Data untuk cluster 0 dapat dilihat pada gambar 5.11, data untuk cluster 1 dapat dilihat pada gambar 5.12 dan data untuk cluster 1 dapat dilihat pada gambar 5.13



Gambar 5.11 Data cluster 0



Gambar 5.12 Data cluster 1



Gambar 5.13 Data cluster 2

Data cluster dengan IP1, IP2, IP3, IP4, IP5, dan IP6 dapat dilihat pada gambar 5.14. Data centroid untuk masing-masing IP untuk setiap cluster pada gambar 5.15 menunjukkan titik tengah data Indeks Prestasi mahasiswa angkatan 2020 di Departemen Teknik Komputer.

Row No.	NAMA MAH...	cluster	IP1	IP2	IP3	IP4	IP5	IP6
1	CECEP ARIF ...	cluster_0	3.760	3.200	3.170	2.310	2.120	2.940
2	ZASKIA FAU...	cluster_1	3.640	3.630	3.380	3.570	3.770	3.860
3	ELANDRA M...	cluster_1	3.740	3.810	3.620	3.350	3.240	3.520
4	WELLA KAR...	cluster_1	3.770	3.900	3.520	3.770	3.770	3.950
5	HANIFA AZZ...	cluster_1	3.850	3.410	3.450	3.500	3.640	3.850
6	VITO YUSKA...	cluster_1	3.890	3.560	3.520	3.710	3.950	3.950
7	YOLANDA N...	cluster_2	3.610	3.190	3.350	3.330	3.240	3.640
8	FARANISA R...	cluster_1	3.730	3.630	3.560	3.760	3.880	3.740
9	NATASHA K...	cluster_1	3.550	3.560	3.330	3.570	3.480	3.850
10	HANA PUTRI...	cluster_1	3.660	3.560	3.530	3.520	3.700	3.890
11	FADTINAH	cluster_1	3.660	3.500	3.570	3.630	3.890	3.800
12	FADILLA	cluster_1	3.630	3.730	3.690	3.750	3.770	3.920
13	ABID NAFIN...	cluster_0	3.660	3.740	3.400	2.370	1.600	2.900
14	HAVIS APRI...	cluster_1	3.650	3.610	3.470	3.420	3.890	3.750
15	LUCKY PRIM...	cluster_2	3.590	3.610	3.350	3.080	2.700	3.290

Gambar 5.14 Data cluster dengan IP1, IP2, IP3, IP4, IP5, dan IP6

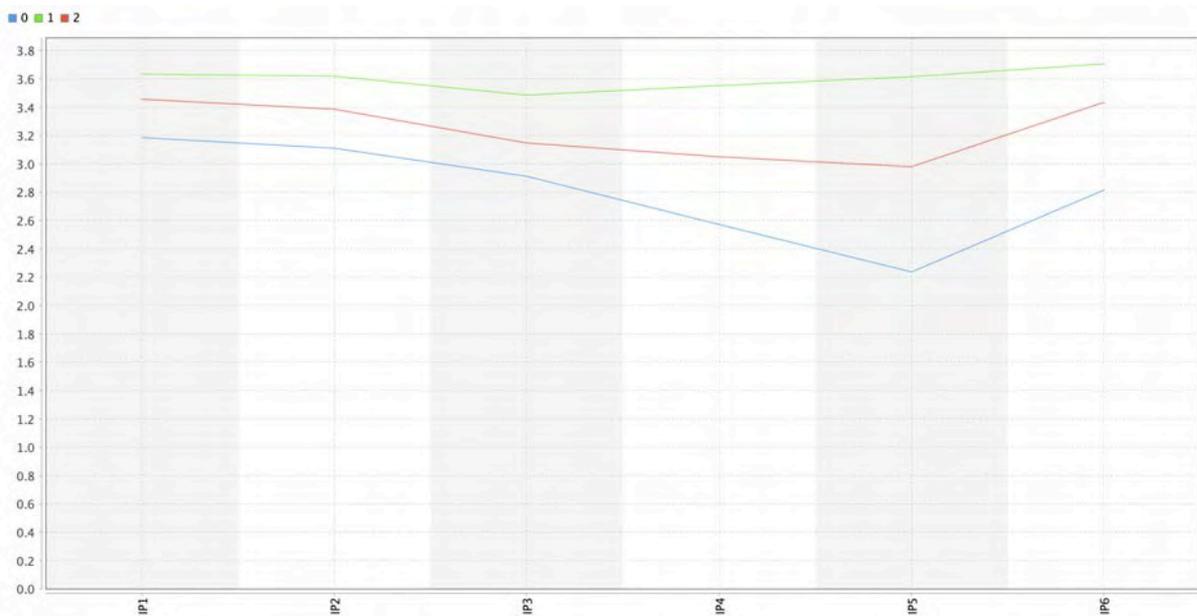
Attribute	cluster_0	cluster_1	cluster_2
IP1	3.185	3.634	3.456
IP2	3.110	3.618	3.386
IP3	2.912	3.485	3.147
IP4	2.572	3.552	3.050
IP5	2.237	3.615	2.980
IP6	2.815	3.706	3.435

Gambar 5.15 data centroid

Nilai minimum, maksimum dan rata-rata Indeks Prestasi mahasiswa pada setiap cluster dapat dilihat pada gambar 5.16. Dari hasil data mining dengan algoritma k-means clustering yang diperoleh dapat dilihat statistik data Indeks Prestasi mahasiswa angkatan 2020 di Departemen Teknik Komputer seperti pada gambar 5.17.

Name	Type	Missing	Statistics		
Id NAMA MAHASISWA	Polynomial	0	Least ZASKIA FAUZANDA (1)	Most ABID NAFINDO AIMAR ...	Values ABID NAFINDO AIMAR (1), ADE AZIZ ARR
Cluster cluster	Nominal	0	Least cluster_0 (6)	Most cluster_1 (40)	Values cluster_1 (40), cluster_2 (25), ...[1 more]
IP1	Real	0	Min 2.760	Max 3.890	Average 3.534
IP2	Real	0	Min 2.170	Max 3.900	Average 3.494
IP3	Real	0	Min 1.930	Max 3.820	Average 3.318
IP4	Real	0	Min 2.280	Max 3.770	Average 3.292
IP5	Real	0	Min 1.600	Max 3.950	Average 3.275
IP6	Real	0	Min 2.340	Max 3.950	Average 3.535

Gambar 5.16 Minimum, maksimum dan rata-rata IP untuk setiap cluster



Gambar 5.17 Statistik Indeks Prestasi mahasiswa Departemen Sistem Informasi

## 5.2. ANALISA

Dari penelitian yang telah dilakukan, *Clustering* dengan Algoritma K-Means untuk Indeks Prestasi mahasiswa telah dihasilkan model clustering untuk Departemen Sistem Informasi dan Departemen Teknik Komputer. Model *Clustering* dengan Algoritma K-Means IP Mahasiswa Departemen Sistem Informasi, diperoleh nilai terkecil  $K=2$  dengan nilai Davies Bouldin = - 0,852. Kluster 0 = 71 mahasiswa dan Kluster 1 = 5 mahasiswa. Dari hasil ini dapat dilihat bahwa ada 71 mahasiswa yang masuk dalam cluster IP yang baik, dan hanya 5 mahasiswa yang masuk ke dalam cluster IP yang tidak baik. Hal ini menunjukkan bahwa potensi mahasiswa angkatan 2020 di Departemen Sistem Informasi untuk dapat menyelesaikan studi tepat waktu sangat tinggi yaitu 93,42 %. Untuk mahasiswa yang berada di cluster 0 yaitu sebanyak 5 orang atau 6,58 % dibutuhkan strategi untuk membantu mahasiswa memperbaiki IP nya sehingga masa studi menjadi tidak terlalu lama.

Model *Clustering* dengan Algoritma K-Means IP Mahasiswa Departemen Teknik Komputer, diperoleh nilai terkecil  $K=3$  dengan nilai Davies Bouldin = - 1,169. Kluster 0 = 6 mahasiswa, Kluster 1 = 40 mahasiswa dan Kluster 2 = 25 mahasiswa. Dari hasil ini dapat dilihat bahwa terdapat 3 cluster Indeks Prestasi mahasiswa angkatan 2020 di Departemen Teknik Komputer. Kelompok dengan IP terbaik berada pada cluster 1 yang terdiri dari 40 mahasiswa, kelompok dengan IP sedang berada pada cluster 2 yang terdiri dari 25 mahasiswa dan kelompok dengan IP rendah berada pada cluster 0 yang terdiri dari 6 mahasiswa. Dari hasil ini dapat dilihat bahwa mahasiswa yang berada pada cluster 1 dan 2 memiliki potensi untuk dapat menyelesaikan studinya tepat waktu, dengan jumlah mahasiswa 65 mahasiswa atau 91,54 %. Untuk mahasiswa yang berada di cluster 0 yaitu sebanyak 6 orang atau 8,46 % dibutuhkan strategi untuk membantu mahasiswa memperbaiki IP nya sehingga masa studi menjadi tidak terlalu lama.

## **BAB 6. KESIMPULAN**

Penelitian yang telah dilakukan untuk penerapan *data mining* terhadap Indeks Prestasi mahasiswa dengan menggunakan algoritma K-Means Clustering telah

1. Data mining dengan algoritma K-Means telah menghasilkan model kluster untuk mengelompokkan Indeks Prestasi mahasiswa angkatan 2020 dari semester 1 sampai 6 di Departemen Sistem Informasi dan Departemen Teknik Komputer
2. Evaluasi model kluster telah dilakukan dengan menggunakan metode Davis Bouldin Index (DBI) yang menghasilkan 2 kluster di Departemen Sistem Informasi dan 3 kluster di Departemen Teknik Komputer.

## DAFTAR PUSTAKA

- [1] P.-N. Tan, M. Steinbach, and V. Kumar, Introduction To Data Mining, Second Edi., vol. 2, no. 1. New York: Pearson Education, Inc, 2006.
- [2] Kusriani and L. E. Taufiq, Algoritma Data Mining. Yogyakarta, Indonesia: Andi - Amikom Yogyakarta, 2009.
- [3] M. Arhami and M. Nasir, Data Mining - Algoritma dan Implementasi. Yogyakarta: Penerbit Andi, 2020.
- [4] M. Arhami and M. Nasir, Data Mining - Algoritma dan Implementasi. Yogyakarta: Penerbit Andi, 2020.
- [5] M. Arhami and M. Nasir, Data Mining - Algoritma dan Implementasi. Yogyakarta: Penerbit Andi, 2020.
- [6] M. Arhami and M. Nasir, Data Mining - Algoritma dan Implementasi. Yogyakarta: Penerbit Andi, 2020.