

Received: February 01, 2021  
Accepted : February 05, 2021  
Published : February 24, 2021

**Conference on Management, Business,  
Innovation, Education and Social Science**  
<https://journal.uib.ac.id/index.php/comblines>

## **Perancangan Model Prediksi Performa Akademik Mahasiswa Menggunakan Algoritma K-Means Clustering (Studi Kasus: Universitas Xyz)**

**Yefta Christian<sup>1</sup> Jimmy<sup>2</sup>**

Email korespondensi : [yefta@uib.ac.id](mailto:yefta@uib.ac.id)[jimmy@uib.edu](mailto:jimmy@uib.edu)

<sup>1</sup>Fakultas Ilmu Komputer, Universitas Internasional Batam, Kota Batam, Indonesia

<sup>2</sup>Fakultas Ilmu Komputer, Universitas Internasional Batam, Kota Batam, Indonesia

### **Abstrak**

Proses pembelajaran dalam sebuah perguruan tinggi menghasilkan data mahasiswa yang berlimpah secara berulang. Data tersebut dapat dimanfaatkan untuk memperoleh informasi yang berguna dan dapat dijadikan sebagai salah satu basis dalam memprediksi performa akademik mahasiswa di suatu universitas. Metode *Data mining* dengan algoritma *K-Means Clustering* yang digunakan untuk mengelola data akademik mahasiswa. Hasil dari penelitian ini merupakan pengelompokan data hasil akademik mahasiswa berdasarkan *cluster* yang diharapkan hasilnya dapat digunakan sebagai model prediksi kinerja akademik mahasiswa.

### **Kata Kunci:**

Performa Akademik, *Data Mining*, *K-Means Clustering*.

### **Pendahuluan**

Perkembangan teknologi informasi pada masa ini sudah berkembang pesat dan banyak sekali data yang dihasilkan hampir dalam berbagai sektor kehidupan (Aziz et al., 2018). Munculnya kebutuhan informasi lebih dari data yang tersimpan dalam gudang data (*database*) dalam dunia bisnis maupun instansi seperti perguruan tinggi. Karena banyak data yang dihasilkan oleh teknologi informasi yang canggih, informasi yang dibutuhkan tidak bisa diperoleh dengan mudah sehingga perlu mengolah ulang dari data tersebut agar dapat menyajikan suatu informasi yang lebih berharga (Sabna & Muhardi, 2016).

Pada institusi pendidikan perguruan tinggi, proses pembelajaran akan menghasilkan data yang berlimpah berupa nilai prestasi dari mahasiswa. Hal seperti ini akan terjadi secara berulang pada sebuah perguruan tinggi. Apabila melakukan pengolahan terhadap data pada sumber data tersebut maka dapat diketahui berbagai informasi yang bermanfaat bagi pihak universitas (Ariasa et al., 2020).

Berdasarkan berlimpahnya data prestasi atau kinerja akademik mahasiswa selama masa studi yang pernah ditempuh, maka dapat dijadikan sebagai salah satu basis untuk memantau perkembangan kinerja akademik mahasiswa dijenjang perguruan tinggi adalah mengelompokkan hasil evaluasi akademik mahasiswa (Fadrial, 2020). Dengan menggunakan teknik *data mining* dapat membantu untuk mengolah data dari siswa yang sudah mengikuti

proses pembelajaran. Banyak sekali teknik *data mining* yang bisa digunakan dalam pengelompokan atau pengklasteran data. Salah satunya yaitu *K-Means Clustering* yang dinilai mampu mengelompokkan data akademik mahasiswa dengan cepat dan efisien dan dapat dijadikan sebagai tolak ukur dalam memonitor performa akademik mahasiswa dalam instansi perguruan tinggi (Syam, 2017).

### Tinjauan Pustaka

Menurut penelitian oleh (Fitri et al., 2018) yang berjudul "Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa (Studi Kasus: UMTAS)" menjelaskan bahwa penambangan informasi dari data jumlah besar yang disebut big data bertujuan untuk dapat menggali data yang memiliki potensi yang tersembunyi dalam gudang data. Teknik yang digunakan untuk menganalisis dan mengestrak informasi dari sejumlah kumpulan data yang berukuran besar dengan menggunakan algoritma dikenal dengan *data mining*.

Menurut penelitian oleh (Fadrial, 2020) yang berjudul "Klasterisasi Hasil Evaluasi Akademik Menggunakan Metode *K-Means* (Studi Kasus: Fakultas Ilmu Komputer UNILAK)" menjelaskan tentang *Educational Data Mining* (EDM) merupakan *sub-domain* dari *data mining* yang berhubungan dengan data dari *database* akademik yang digunakan untuk menemukan berbagai pengetahuan atau pola yang menarik. Pengetahuan tersebut dapat ditujukan kepada para perencana akademik di suatu lembaga pendidikan tinggi untuk mengambil keputusan dalam meningkatkan kinerja akademik mahasiswa serta pohon keputusan untuk memprediksi kinerja siswa dalam hal nilai atau persentase.

Menurut penelitian oleh (Rosadi et al., 2016) yang berjudul "Aplikasi *K-Means Clustering* Untuk Mengelompokkan Data Kinerja Akademik Mahasiswa" mengatakan bahwa ada kalanya data mahasiswa tersebut perlu dikelompokkan terutama data tersebut dengan jumlah yang sangat besar, sehingga pola hubungan data didalam kelompok ataupun antar kelompok dapat diungkap. Dengan bantuan teknik metode *clustering*, mahasiswa dapat dikelompokkan berdasarkan prestasi akademiknya.

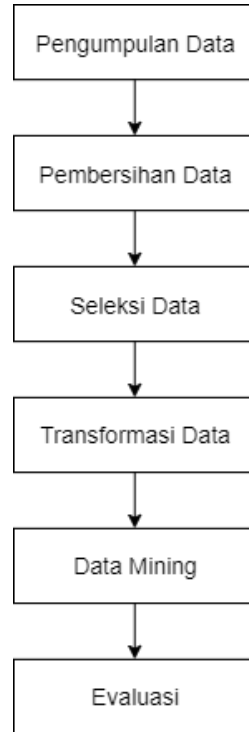
Menurut penelitian oleh yang berjudul "Penerapan Algoritma *K-Means* Untuk Analisis Prestasi Akademik Mahasiswa Fakultas Ilmu Komputer Universitas Klatat" mengatakan bahwa peneliti memanfaatkan data mahasiswa untuk dapat mengetahui mahasiswa yang memiliki prestasi dimata kuliah yang sudah diambil. Dengan menerapkan algoritma *K-Means* untuk menganalisis prestasi akademik mahasiswa Fakultas Ilmu Komputer dan mengetahui dimana mahasiswa tersebut memiliki pengetahuan dan potensi selama masa studi di Universitas Klatat.

Menurut penelitian oleh (Alberto et al., 2017) yang berjudul "Pengelompokkan Performa Akademik Mahasiswa Berdasarkan Indeks Prestasi Menggunakan *K-Means Clustering*" menjelaskan bahwa kualitas mahasiswa ditentukan oleh keterampilan yang dimiliki oleh setiap mahasiswa sebagai indikator untuk menyelesaikan perkuliahan. Universitas menggunakan Indeks Prestasi (IP) untuk mengelompokkan performa akademik mahasiswa akan tetapi, selama ini menggunakan teknik algoritma statistik tidak memberikan banyak hasil yang efektif. Oleh karena itu, kombinasi *K-Means Clustering* digunakan untuk dapat memberikan hasil yang lebih efektif dan juga diimplementasikan pada aplikasi prediksi performa akademik mahasiswa dan dapat digunakan sebagai tolak ukur dalam memonitor performa akademik mahasiswa.

### Metodologi Penelitian

Metode penelitian yang digunakan pada penelitian ini memperhatikan dengan teliti tahap-tahap dilalui yang dibentuk dalam sebuah kerangka penelitian. Tahapan penelitian

berguna agar pelaksanaan penelitian dapat berjalan dengan baik dan sistematis. Berikut alur penelitian dapat dilihat pada gambar 1.



Gambar 1 Alur Penelitian

Pada Gambar 1 ditunjukkan ada enam (6) alur penelitian pada metode ini, antara lain:

#### **A. Pengumpulan Data**

Pada tahap ini adalah proses dimana melakukan pengumpulan data yang akan digunakan dalam proses pengolahan data. Data yang diperoleh dan digunakan oleh peneliti adalah data primer yang berarti didapatkan secara langsung dari instansi terkait. Metode pengumpulan data dilakukan dengan cara permohonan data akademik mahasiswa terlebih dahulu ke pihak IT Center di Universitas.

#### **B. Pembersihan Data**

Pada tahap ini penulis melakukan proses menghilangkan data yang memiliki atribut-atribut tidak sempurna atau jelek. Hal tersebut perlu dilakukan agar tidak mempengaruhi performa proses pengelompokan yang dilakukan.

#### **C. Seleksi Data**

Pada tahap ini penulis melakukan pemilihan data (table, record dan atribut) tertentu yang sesuai untuk dipakai dalam proses *K-Means Clustering*. Karena data yang diperoleh tidak semuanya digunakan dan dipilih sesuai dengan atribut atau variabel yang dibutuhkan dalam penelitian dengan melakukan seleksi data sehingga menjadi *dataset*.

#### D. Transformasi Data

Pada tahap ini penulis mengubah data ke dalam format yang sesuai untuk dapat diproses oleh aplikasi data mining dalam melakukan *K-Means Clustering*. Transformasi data dilakukan dengan memberikan inisialisasi terhadap data yang memiliki nilai nominal menjadi bernilai numerik. Tujuan dari transformasi data untuk menghindari data yang rusak dan tidak valid.

#### E. Data Mining

Pada tahap ini penulis melakukan proses *data mining* yaitu dengan teknik *K-Means Clustering* dari data yang sudah terkumpul. *Output* yang didapatkan adalah hasil pengelompokkan dari *dataset*.

#### F. Evaluasi

Pada tahap ini penulis melakukan evaluasi pola-pola tertentu yang telah dihasilkan oleh proses *K-Means Clustering* untuk memperoleh informasi yang bermanfaat. Evaluasi dilakukan untuk menerjemahkan pola-pola yang dihasilkan dalam format yang lebih mudah dipahami oleh pihak terkait.

### Hasil dan Pembahasan

Sumber data utama yang digunakan dalam penelitian ini adalah *database* alumni Universitas XYZ dengan jumlah keseluruhan data terdapat 1195 *record* dari angkatan 2015 sampai dengan tahun 2016. *Dataset* mahasiswa yang didapatkan berupa dokumen *excel*. Dalam data terdapat nilai yang tidak konsisten pada atribut sehingga penulis melakukan pembersihan data dengan memberikan nilai tertentu pada nilai yang kosong. Setelah itu, tahap seleksi data dilakukan pemilihan atribut yang akan digunakan dalam penelitian ini adalah program studi mahasiswa, tahun angkatan, tahun lulus, usia, lama studi, total SKS dan IPK. Kemudian, melakukan transformasi data pada data yang berjenis nominal seperti program studi dengan cara menginisialisasi ke dalam bentuk angka/numerikal berdasarkan frekuensi terbesar seperti pada tabel 1.

Tabel 1. Inisialisasi Data Program Studi

Program Studi	Frekuensi	Inisial
Manajemen	462	1
Akuntansi	293	2
Ilmu Hukum	172	3
Sistem Informasi	164	4
Teknik Sipil	79	5
Teknik Elektro	25	6

Setelah semua tahapan sebelumnya selesai dilakukan, *dataset* siap digunakan dalam tahap *data mining*. Dalam penelitian ini, penulis menggunakan algoritma *K-Means* serta tiga atribut yang terpilih untuk melakukan proses *data mining*. Berikut ini adalah hasil dari pengolahan data yang telah dilakukan pada WEKA yaitu pengelompokkan (*clustering*) terlihat pada gambar 2 sebagai berikut.

Final cluster centroids:							
Attribute	Full Data (1195.0)	Cluster#					
		0 (168.0)	1 (196.0)	2 (94.0)	3 (80.0)	4 (492.0)	5 (165.0)
Program Studi	2.3138	1.2738	1.4235	4.7021	1.7	2.2988	3.4121
Tahun Angkatan	2015.4117	2015	2015	2015	2015	2016	2015
Tahun Lulus	2019.4736	2018.9762	2018.9439	2019.266	2020	2019.9756	2018.9758
Usia	22.5573	22.3869	22.1633	23.266	23.475	22.5183	22.4667
Lama Studi (Semester)	7.5941	7.756	7.148	8.1383	9.7	7.3313	7.4121
IPK	3.3882	3.0713	3.6561	3.1149	3.0974	3.4659	3.4577
Total SKS	147.4519	146.0774	147.7245	146.0106	146.4	149	145.2424

Gambar 2 Hasil *K-Means Clustering* Data Mahasiswa

Dari hasil *K-Means Clustering* data seluruh mahasiswa (1195 Mahasiswa) menggunakan 6 *cluster* dan 7 atribut antara lain.

1. *Cluster* pertama terdapat 168 mahasiswa, dimana mahasiswanya lebih dominan pada program studi manajemen dengan rata-rata IPK 3.07, berusia 23 tahun, tahun angkatan 2015 dan tahun lulus 2019 serta total SKS 146.
2. *Cluster* kedua terdapat 196 mahasiswa dengan program studi manajemen memiliki rata-rata IPK 3.65, berusia 22 tahun, tahun angkatan 2015, tahun lulus 2019 serta total SKS 147.
3. *Cluster* ketiga terdapat 94 mahasiswa dengan program studi teknik sipil memiliki rata-rata IPK 3.1, berusia 23 tahun, tahun angkatan 2015, tahun lulus 2019 serta total SKS 146.
4. *Cluster* keempat terdapat 80 mahasiswa dengan program studi manajemen memiliki rata-rata IPK 3.09, berusia 23 tahun, tahun angkatan 2015, tahun lulus 2020 serta total SKS 146.
5. *Cluster* kelima terdapat 492 mahasiswa dengan program studi akuntansi memiliki rata-rata IPK 3.46, berusia 22 tahun, tahun angkatan 2016, tahun lulus 2020 serta total SKS 149.
6. *Cluster* keenam terdapat 165 mahasiswa dengan program studi ilmu hukum memiliki rata-rata IPK 3.45, berusia 22 tahun, tahun angkatan 2015, tahun lulus 2019 serta total SKS 145.

Berdasarkan hasil *cluster* yang telah didapatkan di atas, terdapat *cluster* keempat yang mengindikasikan bahwa mahasiswa program studi manajemen dengan rata-rata IPK 3.09 dan lulus tidak tepat waktu. Maka dari itu perlu lebih memonitor akademiknya pada *cluster* tersebut dalam meningkatkan performa akademik mahasiswa pada *cluster* keempat dibandingkan dengan *cluster* lainnya.

Adapun hasil penelitian ini melakukan evaluasi untuk mengetahui nilai keakuratan sebuah model prediksi yang dibangun berdasarkan pengolahan *dataset*. Dalam pengujian menggunakan perangkat lunak WEKA dengan metode klasifikasi dengan algoritma J48 Tree untuk mengukur performa algoritma *K-Means* berupa nilai *precision*, *recall* dan *accuracy*. Dapat dilihat gambar 3 hasil klasifikasi sebagai berikut.

```

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      1180           98.7448 %
Incorrectly Classified Instances     15             1.2552 %
Kappa statistic                     0.9834
Mean absolute error                  0.0048
Root mean squared error              0.0638
Relative absolute error              1.9075 %
Root relative squared error          17.9943 %
Total Number of Instances           1195

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.970	0.003	0.982	0.970	0.976	0.972	0.992	0.977	cluster0
	0.980	0.004	0.980	0.980	0.980	0.976	0.993	0.978	cluster1
	0.968	0.003	0.968	0.968	0.968	0.965	0.984	0.968	cluster2
	1.000	0.001	0.988	1.000	0.994	0.993	0.999	0.975	cluster3
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	cluster4
	0.982	0.004	0.976	0.982	0.979	0.975	0.998	0.981	cluster5
Weighted Avg.	0.987	0.002	0.987	0.987	0.987	0.986	0.996	0.986	

Gambar 3 Hasil *Classification* Algoritma J48 Tree

Dalam hasil klasifikasi algoritma J48 *Tree* memiliki nilai *Precision* : 98.7%, *Recall* : 98.7% dan *accuracy* : 98.75%.

## Kesimpulan

Berdasarkan hasil dan pembahasan yang telah dilakukan, dapat disimpulkan dari penelitian ini bahwa penerapan *data mining* menggunakan algoritma *K-Means Clustering* yang diimplementasi ke dalam program *data mining tools* dapat mengelompokkan data akademik mahasiswa. Dari hasil proses pengelompokkan (*clustering*) terhadap 1195 *dataset* mahasiswa menggunakan aplikasi WEKA mendapatkan pengetahuan baru yang bisa dijadikan sumber informasi serta berfungsi sebagai model acuan bagi perencana akademik untuk memonitor dan memprediksi perkembangan performa akademik setiap mahasiswa.

## Daftar Pustaka

- Alberto, R. Z., Sari, W. K., & Primanita, A. (2017). Pengelompokkan Performa Akademik Mahasiswa Berdasarkan Indeks Prestasi Menggunakan K-Means Clustering. *Jurnal Online Informatika (JOIN)*, 99–104.
- Ariasa, K., Gunadi, I. G. A., & Candiasa, I. M. (2020). *Optimasi Algoritma Klaster Dinamis Pada K-Means Dalam ( Studi Kasus : Universitas Pendidikan Ganesha )*. 9(2), 181–193.
- Aziz, F. N. R. F., Setiawan, B. D., & Arwani, I. (2018). Implementasi Algoritma K-Means untuk Klasterisasi Kinerja Akademik Mahasiswa. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(6), 2243–2251.
- Fadrial, Y. E. (2020). Klasterisasi Hasil Evaluasi Akademik Menggunakan Metode K-Means (Studi Kasus Fakultas Ilmu Komputer UNILAK). *Seminar Nasional Teknologi Informasi & Ilmu Komputer*, 1(1), 53–65.

- Fitri, S., Nurjanah, N., & Astuti, W. (2018). Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa (Studi Kasus: Umtas). *Simetris: Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, 9(1), 633–640.
- Rosadi, R., Akmal, Hidayat, A., & Kharismawan, B. (2016). Aplikasi K-Means Clustering Untuk Mengelompokan Data Kinerja Akademik Mahasiswa. *Seminar Nasional Teknik Elektro, November*, 92–96.
- Sabna, E., & Muhardi, M. (2016). Penerapan Data Mining Untuk Memprediksi Prestasi Akademik Mahasiswa Berdasarkan Dosen, Motivasi, Kedisiplinan, Ekonomi, dan Hasil Belajar. *Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer Dan Teknologi Informasi*, 2(2), 41–44.
- Syam, F. A. (2017). Implementasi Metode Klastering K-Means untuk Mengelompokan Hasil Evaluasi Mahasiswa. *Jurnal Ilmu Komputer Dan Bisnis*, 8(1), 1857–1864.