

PENYEBARAN MAHASISWA ITB AHMAD DAHLAN JAKARTA MENGGUNAKAN ALGORITMA CLUSTERING

Muhamad Toha^{1(*)}, Diana Yusuf², Vany Terisia³

^{1,2,3} Institut Teknologi dan Bisnis Ahmad Dahlan, Jakarta

Abstract

The development of information technology has made it possible to collect and analyze large and complex data. In the context of higher education, information about student distribution patterns can be of added value for institutions to plan and manage their resources effectively. This research aims to design a web mining system that uses a clustering algorithm to grouping the distribution of students at ITB Ahmad Dahlan. In this study, we collected data from open web sources that are relevant to students, such as student profiles, academic preferences, and extracurricular activities. The data is then analyzed using a clustering algorithm to identify patterns and trends in student distribution. By using this approach, we hope to provide useful insights for educational institutions in planning infrastructure, academic programs, and student services.

Kata Kunci: *Web Mining, Student Distribution Grouping, Clustering Algorithm*

Januari – Juni 2024, Vol 5 (1) : hlm 35-46

©2024 Institut Teknologi dan Bisnis Ahmad Dahlan.

All rights reserved.

(*) Korespondensi: muhamadtoha401@gmail.com (Muhamad Toha)

PENDAHULUAN

Teknologi informasi telah mengalami perkembangan yang pesat pada semua aspek kehidupan. Berbagai macam data yang dapat dihasilkan oleh teknologi informasi mulai dari bidang ekonomi, industri, dan teknologi serta bidang kehidupan lainnya. Penggunaan teknologi informasi khususnya dibidang pendidikan juga dapat mengumpulkan jumlah data yang signifikan tentang proses pembelajaran dan mahasiswa. Data di institusi pendidikan tinggi dapat diperoleh dari data historis, yang berarti bahwa data terus bertambah, seperti data mahasiswa. (Yunita, 2018). Perkembangan teknologi informasi di ITB Ahmad Dahlan Jakarta juga sangat berpengaruh pada pengelompokan data mahasiswa yaitu dengan penerapan data *mining* untuk mengetahui sebaran wilayah mahasiswa dengan menggunakan metode algoritma *clustering*.

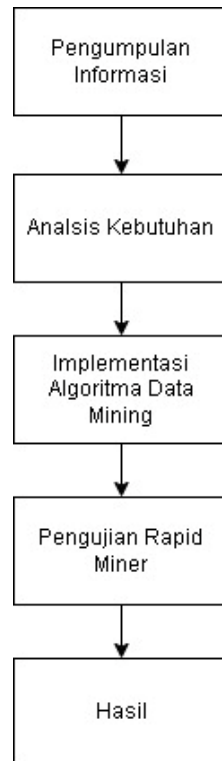
Data *mining* adalah sebuah metode yang dapat mencari korelasi, pola, dan tren baru yang memiliki arti melalui pemisahan data dalam jumlah besar yang disimpan di dalam media penyimpanan, menggunakan metode pengenalan pola serta teknik statistik dan matematika. Salah satu metode pengelompokan data dalam data *mining* adalah algoritma *Clustering*. (Nabila, et al., 2021). *Clustering* adalah proses pemisahan data ke dalam kelas atau *cluster* berdasarkan tingkat kesamaannya. Dalam *clustering*, data yang mempunyai kesamaan dimasukkan ke dalam *cluster* atau kelas yang sama, sedangkan data yang tidak mempunyai kesamaan dimasukkan dalam *cluster* yang berbeda. (Yunita, 2018). Pada metode *Clustering*, terdapat algoritma *K-Means* untuk melakukan proses perhitungan. *K-Means* merupakan algoritma *Clustering* yang berulang-ulang (Putra, dan Wadisman, 2018).

Proses algoritma *k-means* memiliki perbedaan dengan algoritma lain seperti Algoritma *Apriori* yang mencari *frequent* itemsets yang sering muncul dengan model data heterogen juga kemudian dipangkas dan dilakukan perhitungan sesuai jumlah *K* itemsetnya, klasterisasi mengelompokkan sejumlah *n* buah objek ke dalam *k* kelas dengan berdasarkan perhitungan jarak nya dengan pusat *cluster*. (Muliono, dan Sembiring, 2019). Sebagai pembanding data agar lebih akurat, *RapidMiner* dapat digunakan untuk membantu proses tersebut. *RapidMiner* merupakan sebuah aplikasi yang digunakan untuk pengolahan data. *RapidMiner* memudahkan penggunaanya dalam melakukan perhitungan. (CTI, B. R., et, al., 2017). Pada penelitian ini penulis akan memfokuskan pada klasterisasi penyebaran wilayah mahasiswa di ITB Ahmad

Dahlan.

METODE.

Kerangka kerja penelitian ini merupakan langkah-langkah yang diambil oleh penulis dalam menyelesaikan permasalahan yang diteliti. Berikut ini adalah penjelasan lebih detail mengenai kerangka kerja penelitian yang digunakan oleh penulis:



Sumber: Analisis Peneliti

Gambar 1 Kerangka Penelitian

1. Pengumpulan Informasi

Pada tahap ini, peneliti melakukan pengumpulan data melalui 3 metode, yaitu wawancara, observasi, dan studi pustaka. Peneliti melakukan wawancara dan observasi secara langsung di ITB Ahmad Dahlan Jakarta.

2. Analisis Kebutuhan

Setelah melakukan pengumpulan data, peneliti melakukan analisis kebutuhan. Hal ini dimaksudkan agar data yang diperoleh nantinya bisa sesuai dengan kebutuhan dalam mengelompokkan mahasiswa di ITB Ahmad Dahlan Jakarta.

3. Implementasi Algoritma Data Mining

Pada tahap ini, peneliti melakukan penerapan data *mining* pada biodata mahasiswa yang sudah ada. Penerapan algoritma *K-Means Clustering* dilakukan menggunakan aplikasi *RapidMiner* dan juga perhitungan manual dengan menghitung jarak *euclidean* dengan persamaan rumus $d(x_i, \mu_i) = \sqrt{(x_i - \mu_i)^2}$ Setelah itu menghitung nilai pusat *cluster* atau nilai *centroid* baru dengan rumus $C_k = \frac{1}{n_k} \sum d_i$. Kemudian dilanjutkan dengan perhitungan jarak data ke *centroid* menggunakan rumus *Euclidean distance*: $De = \sqrt{(x_i - S_i)^2 + (y_i - t_i)^2}$

4. Pengujian *RapidMiner*

Selanjutnya, melakukan proses pengujian data pada *RapidMiner* sebagai pembandingan agar mendapatkan data yang akurat

5. Hasil

Hasil pengujian dapat dianggap berhasil jika data yang didapatkan telah akurat dari hasil perhitungan manual serta menggunakan aplikasi *RapidMiner*. Dengan mengolah data mahasiswa dari tahun akademik 2019–2023 untuk menghasilkan kelompok mahasiswa yang disarankan berdasarkan wilayahnya.

HASIL DAN PEMBAHASAN

Setelah dilakukan pembersihan data dengan menghapus berbagai kriteria yang tidak diperlukan dan memilih data yang relevan dari biodata mahasiswa, maka didapatkan hasilnya menjadi atribut nim, wilayah, prodi dan IPK yang dapat dilihat pada table berikut:

Tabel. 1 Data Hasil Cleaning

NIM	Wilayah	Prodi	IPK
2262201079	Bekasi	Akuntansi (S1)	3.62
2259201028	Tangerang	Teknologi Informasi (S1)	3.03
2361201059	Tangerang	Manajemen (S1)	1.9
2261201227	Bekasi	Manajemen (S1)	3.51
2357201005	Tangerang	Sistem Informasi (S1)	1.2
2161201769	Bekasi	Manajemen (S1)	3.49
2362201053	Bekasi	Akuntansi (S1)	1.8
2162201242	Bekasi	Akuntansi (S1)	3.92
2262201085	Cilacap	Akuntansi (S1)	3.87
2190241060	Tangerang	Desain Komunikasi Visual (S1)	3.79
2259201003	Tangerang	Teknologi Informasi (S1)	3.66

2259201041	Tangerang	Teknologi Informasi (S1)	3.34
2061201181	Jakarta	Manajemen (S1)	3.43
2361201114	Tangerang	Manajemen (S1)	1.7
2061201142	Garut	Manajemen (S1)	3.75
2261201167	Jakarta	Manajemen (S1)	3.66
2057201019	Bogor	Sistem Informasi (S1)	3.72
2015024118	Bogor	Manajemen (S1)	3.75
2061201145	Bogor	Manajemen (S1)	3.52
2157201005	Depok	Sistem Informasi (S1)	3.58
2361201064	Tangerang	Manajemen (S1)	1.2
2057201015	Tangerang	Sistem Informasi (S1)	3.17
2162201060	Karawang	Akuntansi (S1)	3.65
2059201004	Tangerang	Teknologi Informasi (S1)	3.94
2290241043	Tangerang	Desain Komunikasi Visual (S1)	3.74
2157201012	Jakarta	Sistem Informasi (S1)	3.32
2062201046	Tangerang	Akuntansi (S1)	3.72
.	.	.	.
.	.	.	.
.	.	.	.
2123201005	Sukabumi	Arsitektur (S1)	3.87

Sumber: analisis literatur, 2023

1. Menentukan Jumlah *Cluster*

Pada tahap awal algoritma *K-means*, langkah pertama yang harus dijalankan adalah menentukan jumlah *cluster* yang akan digunakan. Dalam perhitungan ini dibuat menjadi 2 *Cluster* sebagai berikut:

1. *Cluster* Pertama (C1): Direkomendasikan
2. *Cluster* Kedua (C2): Tidak Direkomendasikan

2. Menentukan *Centroid*

Pada langkah ini, akan ditentukan nilai pusat *cluster* (*centroid*) dari data yang digunakan. Penentuan *centroid* dilakukan berdasarkan kriteria berikut pada skripsi ini:

1. Wilayah Mahasiswa (WM)
2. Program Studi (PS)
3. Index Prestasi Kumulatif (IPK)

3. Menentukan Jarak Data Dengan *Centroid*

Selanjutnya dilakukan proses perhitungan jarak dari setiap data ke pusat *cluster* dengan menggunakan rumus *Euclidean distance*. Perhitungan jarak data ke *centroid* menggunakan rumus *Euclidean distance*:

$$De = \sqrt{(xi - Si)^2 + (yi - ti)^2}$$

Dimana:

$D_e = \text{Euclidean Distance}$

$i = \text{banyak objek}$

$(x,y) = \text{koordinat object}$

$(s,t) = \text{koordinat centroid.}$

Selanjutnya proses *iterasi* dilakukan mulai dari *iterasi* 1 sampai mendapatkan data hasil proses perhitungan jarak data mahasiswa yaitu pada *iterasi* 6. Proses *iterasi* dapat dilihat pada tabel di bawah ini:

Tabel 2. Hasil Iterasi 1

	C1	C2	Class
M1	3.605551	2	2
M2	1.732051	4.242641	1
M3	2.44949	5	1
M4	3.162278	3.605551	1
M5	2.236068	4.472136	1
M6	3.162278	3.605551	1
M7	4.123106	0	2
M8	3.605551	2	2
M9	4.472136	2.236068	2
M10	3.162278	4.582576	1
M11	1.414214	4.582576	1
M12	1.414214	4.582576	1
M13	1	4.690416	1
M14	2.44949	5	1
M15	4.123106	3.741657	2
M16	1	4.690416	1
M17	1	3.464102	1
M18	1.414214	4.123106	1
M19	1.414214	4.123106	1
M20	2	3	1
•	•	•	•
•	•	•	•
•	•	•	•
M52	7.211103	4.582576	2

Sumber: analisis literatur, 2023

Tabel 3. Hasil Iterasi 2

	C1	C2	Class
M1	3.605551	2	2
M2	1.732051	4.242641	1
M3	2.44949	5	1
M4	3.162278	3.605551	2
M5	2.236068	4.472136	1
M6	3.162278	3.605551	2
M7	4.123106	0	2

M8	3.605551	2	2
M9	4.472136	2.236068	2
M10	3.162278	4.582576	1
M11	1.414214	4.582576	1
M12	1.414214	4.582576	1
M13	1	4.690416	1
M14	2.44949	5	1
M15	4.123106	3.741657	2
M16	1	4.690416	1
M17	1	3.464102	1
M18	1.414214	4.123106	1
M19	1.414214	4.123106	1
M20	2	3	2
.	.	.	.
.	.	.	.
.	.	.	.
M52	7.211103	4.582576	2

Sumber: analisis literatur, 2023

Tabel 4. Hasil Iterasi 3

	C1	C2	Class
M1	3.5919	0.746181	2
M2	0.917136	4.510683	1
M3	2.288029	5.348777	1
M4	3.704042	2.466395	2
M5	1.721148	4.976677	1
M6	3.704042	2.466395	2
M7	3.820826	1.927351	2
M8	3.5919	0.746181	2
M9	4.521858	0.845388	2
M10	2.539134	4.699266	1
M11	0.99632	4.446045	1
M12	0.99632	4.446045	1
M13	1.756008	4.196358	1
M14	2.288029	5.348777	1
M15	4.61144	2.498199	2
M16	1.756008	4.196358	1
M17	1.535021	2.815175	1
M18	2.151515	3.430352	1
M19	2.151515	3.430352	1
M20	2.429351	2.020669	2
M21	2.288029	5.348777	1
M22	1.124892	4.648593	1
M23	4.521858	0.845388	2
M24	0.99632	4.446045	1
M25	2.539134	4.699266	1
.	.	.	.
.	.	.	.
.	.	.	.
M52	6.912641	4.826702	2

Sumber: analisis literatur, 2023

Tabel 5. Hasil Iterasi 4

	C1	C2	Class
M1	3.5919	0.746181	2
M2	0.917136	4.510683	1
M3	2.288029	5.348777	1
M4	3.704042	2.466395	2
M5	1.721148	4.976677	1
M6	3.704042	2.466395	2
M7	3.820826	1.927351	2
M8	3.5919	0.746181	2
M9	4.521858	0.845388	2
M10	2.539134	4.699266	1
M11	0.99632	4.446045	1
M12	0.99632	4.446045	1
M13	1.756008	4.196358	1
M14	2.288029	5.348777	1
M15	4.61144	2.498199	2
M16	1.756008	4.196358	1
M17	1.535021	2.815175	1
M18	2.151515	3.430352	1
M19	2.151515	3.430352	1
M20	2.429351	2.020669	2
•	•	•	•
•	•	•	•
•	•	•	•
M52	6.881591	4.620201	2

Sumber: analisis literatur, 2023

Tabel 6. Hasil Iterasi 5

	C1	C2	Class
M1	3.64086	0.807775	2
M2	0.832682	4.433114	1
M3	2.292566	5.239513	1
M4	3.79221	2.335059	2
M5	1.697309	4.883902	1
M6	3.79221	2.335059	2
M7	3.857572	1.962779	2
M8	3.64086	0.807775	2
M9	4.576392	0.975961	2
M10	2.476057	4.674666	1
M11	0.938541	4.364917	1
M12	0.938541	4.364917	1
M13	1.821636	4.068476	1
M14	2.292566	5.239513	1
M15	4.697697	2.398437	2
M16	1.821636	4.068476	1
M17	1.621992	2.693047	1
M18	2.237378	3.294313	1
M19	2.237378	3.294313	1
M20	2.513635	1.884808	2
•	•	•	•

•	•	•	•
•	•	•	•
M52	6.887914	4.748947	2

Sumber: analisis literatur, 2023

Tabel 7. Hasil Iterasi 6

	C1	C2	Class
M1	3.690977	0.879342	2
M2	0.809671	4.343786	1
M3	2.337096	5.114117	1
M4	3.857638	2.280749	2
M5	1.744323	4.762143	1
M6	3.857638	2.280749	2
M7	3.944464	1.854699	2
M8	3.690977	0.879342	2
M9	4.632712	1.096273	2
M10	2.418494	4.645726	1
M11	0.829352	4.305246	1
M12	0.829352	4.305246	1
M13	1.83446	3.989499	1
M14	2.337096	5.114117	1
M15	4.766559	2.372843	2
M16	1.83446	3.989499	1
M17	1.668712	2.620779	1
M18	2.281218	3.216313	1
M19	2.281218	3.216313	1
M20	2.573579	1.815779	2
•	•	•	•
•	•	•	•
•	•	•	•
M52	6.912641	4.826702	2

Sumber: analisis literatur, 2023

Dari tabel iterasi 6 di atas dapat diketahui bahwa hasil perhitungan berhenti pada iterasi 6 yang menunjukkan bahwa 30 data mahasiswa dengan wilayah Tangerang, Jakarta, Depok, dan Bogor masuk ke dalam cluster 1; direkomendasikan. Dan 22 mahasiswa dengan wilayah Bekasi, dan lain-lain masuk ke dalam cluster 2; tidak direkomendasikan. Berdasarkan 52 data sampel mahasiswa.

Selanjutnya proses data *mining* dilakukan dengan menggunakan *RapidMiner* sebagai pembanding data agar mendapatkan data yang akurat. Proses menggunakan *RapidMiner* dapat dilihat pada gambar berikut:

Nama Category	Wilayah Number	Prodi Number	IPK Number	id Number	cluster Category
ALDI FATHURRAHMAN	3	1	4	1	cluster_0
ARDINI ANNISYAH AZZAHRA	6	8	3	2	cluster_1
ARISTA SARASYFA RAHMA NUG...	1	2	4	3	cluster_0
ALFITO FATIHUL IHSAN	1	2	4	4	cluster_0
AKMAL ANNAFIS	3	2	4	5	cluster_0
ALIFA NURUL INAYAH	1	3	4	6	cluster_0
ATIULLAH	6	4	4	7	cluster_1
AI ITA ROSITA	6	1	4	8	cluster_1
ARFAN MAULANA	6	1	4	9	cluster_1
ALDIANSYAH	3	1	4	10	cluster_0
ARI TRISNA MUKTI	2	5	4	21	cluster_0
AGUNG MUHAMMAD NATSIR	1	5	4	22	cluster_0
ARYANTO	6	1	4	23	cluster_1
AGNES VINCENTIA GULTOM	5	4	4	24	cluster_1
ASRYANA	4	4	3	25	cluster_1
AULIA PUTRI AZZAHRA	1	5	4	26	cluster_0
AYU NADYA	3	4	4	27	cluster_0
ARYA HARD MANDALA	6	1	4	28	cluster_1
ALI ZARKASI	6	4	4	29	cluster_1
ANISAH	6	1	4	30	cluster_1

Gambar 2 Hasil RapidMiner

Setelah menghasilkan hasil perhitungan *K-Means clustering* melalui *RapidMiner*, dilakukan pengujian untuk memastikan bahwa hasil perhitungan dari aplikasi tersebut cocok dengan perhitungan manual. Hasil menunjukkan bahwa rata-rata jarak terdekat berdasarkan "*Avg. within centroid distance*" adalah pada *Cluster 0*, yang menunjukkan daerah yang direkomendasikan. Sebaliknya, jarak terjauh ditunjukkan oleh *Cluster 1* dengan nilai "*Avg. within centroid distance*", mengindikasikan daerah yang tidak direkomendasikan. Semakin rendah nilai *Indeks Davies Bouldin*, semakin baik hasil *cluster* yang diperoleh dari metode *K-Means*. Dalam hal ini, nilai *Davies Bouldin* sebesar 1.142 menunjukkan bahwa objek-objek dalam cluster tersebut memiliki kesamaan yang cukup baik karena mendekati angka 0.

Setelah dilakukan proses *Iterasi* dan perhitungan melalui *RapidMiner*, diketahui bahwa hasil yang didapat memiliki kesamaan yaitu *cluster 0* memiliki data 31 mahasiswa dengan wilayah Tangerang, Jakarta, dan Bogor adalah wilayah yang

direkomendasikan, dan *Cluster 1* memiliki data 22 mahasiswa dengan wilayah Bekasi, dan Lain-lain yang tidak direkomendasikan.

KESIMPULAN

Hasil dari penelitian yang dilakukan telah mengarah pada beberapa kesimpulan yang dapat disimpulkan. Berikut adalah rangkuman dari kesimpulan yang diperoleh:

1. Untuk memenuhi kebutuhan data dalam mengelompokkan mahasiswa di ITB Ahmad Dahlan telah berhasil dengan metode algoritma *K-means Clustering*. Diharapkan analisa ini akan membantu dalam proses menentukan sebaran pengelompokan mahasiswa, sehingga promosi ITB Ahmad Dahlan akan menjadi lebih efisien, akurat, dan efektif.
2. Sistem data *mining* ini menggunakan algoritma *K-means clustering* di mana hasil dari mengelompokkan terbagi menjadi 2 (Dua) kriteria yaitu direkomendasikan dan tidak direkomendasikan.
3. Algoritma *clustering K-means* telah digunakan untuk mengetahui sebaran mahasiswa di ITB Ahmad Dahlan. Hasil menunjukkan bahwa 30 mahasiswa yang wilayahnya adalah Tangerang, Jakarta, Depok, Bogor masuk ke dalam *Cluster 1* yang direkomendasikan dan 22 mahasiswa berasal dari wilayah Bekasi dan lain-lain termasuk dalam *Cluster 2* yang tidak direkomendasikan, berdasarkan 52 data sampel.

DAFTAR PUSTAKA

- CTI, B. R., Gafar, A. A., Fajriani, N., Ramdani, U., Uyun, F. R., Purnamasari, Y., & Ransi, N. (2017, May). Implementasi k-means clustering pada rapidminer untuk analisis daerah rawan kecelakaan. In Prosiding Seminar Nasional Riset Kuantitatif Terapan (pp. 58-62).
- Muliono, R., & Sembiring, Z. (2019). Data Mining Clustering Menggunakan Algoritma K-Means Untuk Klasterisasi Tingkat Tridarma Pengajaran Dosen. CESS (Journal of Computer Engineering, System and Science), 4(2), 272-279.

- Nabila, Z., Isnain, A. R., Permata, P., & Abidin, Z. (2021). Analisis data mining untuk clustering kasus covid-19 di Provinsi Lampung dengan algoritma k-means. *Jurnal Teknologi Dan Sistem Informasi*, 2(2), 100-108.
- Putra, R. R., & Wadisman, C. (2018). Implementasi Data Mining Pemilihan Pelanggan Potensial Menggunakan Algoritma K Means. *INTECOMS: Journal of Information Technology and Computer Science*, 1(1), 72-77.
- Yunita, F. (2018). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru. *Sistemasi: Jurnal Sistem Informasi*, 7(3), 238-249.