

Klasterisasi Menggunakan Algoritma *K-Means Clustering* Untuk Memprediksi Kelulusan Mata Kuliah Mahasiswa

Ahziril Pria Adisty^{1*}, Novara Lutfiyani¹, Perahim Tara¹, Rifaldi¹, Richky Adriyan¹, Perani Rosyani¹

¹Fakultas Ilmu Komputer, Teknik Informatika, Universitas Pamulang, Jl. Raya Puspiptek No. 46, Kel. Buaran, Kec. Serpong, Kota Tangerang Selatan. Banten 15310, Indonesia

Email: ^{1*}ahziriladisty25@gmail.com, ²storagerandom1@gmail.com, ³perahimsaber@gmail.com, ⁴rivalz1973@gmail.com, ⁵mixmorgum@gmail.com, ⁶dosen00837@unpam.ac.id

(* : coresponding author)

Abstrak– Kelulusan mahasiswa yang tepat waktu akan menguntungkan pihak mahasiswa dan perguruan tinggi. Salah satu faktor yang mempengaruhi kelulusan mahasiswa yang sesuai dengan waktu studi adalah lulus tidaknya mahasiswa dalam sebuah mata kuliah, karena kelulusan mahasiswa dalam sebuah mata kuliah sangat mempengaruhi rata-rata IPK akhir mereka. Clustering merupakan suatu metode untuk pengelompokan dokumen dimana dokumen dikelompokkan dengan konten untuk mengurangi ruang pencarian yang diperlukan dalam merespon suatu query (Natalius, 2010). Dalam penelitian ini menggunakan tiga kategori kelulusan yaitu nilai bagus, cukup, dan tidak lulus. Hasil perhitungan dataset dengan metode K-Means didapatkan hasil Accuracy 81%, hal ini menunjukkan bahwa sistem dapat mengklasifikasikan data secara benar, namun dalam pengelompokannya belum optimal, karena terdapat data yang terklasifikasi benar masuk ke klasifikasi salah dan data terklasifikasi salah masuk ke klasifikasi benar.

Kata Kunci: Kelulusan, Clustering, K-Means

Abstract– Student graduation on time will benefit both the student and the university. One of the factors that influence student graduation according to the time of study is whether or not students graduate in a course, because student graduation in a course greatly affects their average final GPA. Clustering is a method for grouping documents where documents are grouped by content to reduce the search space required in responding to a query (Natalius, 2010). In this study, three categories of graduation were used, namely good grades, enough, and did not pass. The results of calculating the dataset using the K-Means method yield an accuracy of 81%, this indicates that the system can classify data correctly, but the grouping is not optimal, because there are data that are classified correctly into the wrong classification and data that are classified incorrectly into the correct classification.

Keywords: Graduation, Clustering, K-Means

1. PENDAHULUAN

Kelulusan mahasiswa yang tepat waktu akan menguntungkan pihak mahasiswa dan perguruan tinggi. Untuk mahasiswa, semakin cepat lulus maka kesempatan untuk mengikuti berbagai seleksi dalam mencari pekerjaan semakin banyak. Sedangkan bagi perguruan tinggi, kelulusan mahasiswa dengan tepat waktu dapat memajukan kualitas, meningkatkan reputasi, dan juga berpengaruh pada akreditasi perguruan tinggi tersebut. (Premashanti, Gita. 2015).

Untuk mengetahui tingkat kelulusan mahasiswa dalam satu tahun ajaran dapat dilakukan suatu prediksi berdasarkan data-data mahasiswa pada tingkat atau tahun ajaran pertama. Ada beberapa faktor yang mempengaruhi prediksi kelulusan mahasiswa yang sesuai dengan waktu studi, diantaranya : rata-rata IPK terakhir, jumlah SKS. (Kondo, Ferry. 2015) Dalam menentukan prediksi kinerja mahasiswa dengan akurasi yang tinggi bermanfaat untuk mengidentifikasi siswa dengan prestasi akademik (Suhartinah dan Emastuti, 2010).

Untuk menyelesaikan permasalahan pada penjelasan diatas adalah dengan pemanfaatan algoritma K-Means Clustering. Algoritma K-Means adalah algoritma klastering yang paling sederhana dibanding algoritma klastering yang lain. Algoritma ini mempunyai kelebihan mudah diterapkan dan dijalankan, relatif cepat, mudah untuk diadaptasi, dan paling banyak dipraktikkan dalam tugas data mining.

Clustering merupakan suatu metode untuk pengelompokan dokumen dimana dokumen dikelompokkan dengan konten untuk mengurangi ruang pencarian yang diperlukan dalam merespon suatu query (Natalius, 2010).

2. METODE PENELITIAN

Dalam penelitian ini dilakukan beberapa tahapan yang dilakukan secara sistematis, bertujuan untuk mempermudah peneliti dalam merngkat prosedur dan rangkaian dalam sebuah penelitian dan lebih terarah, penelitian ini merupakan penelitian yang menggunakan model penelitian kuantitatif. Model kuantitatif ini membantu dan mengarahkan dalam sebuah penelitian dalam bentuk penyelesaian sebuah masalah dalam bentuk penelitian agar pencapaian tujuan penelitian tersebut lebih terarah dengan baik. Dalam penelitian ini dilakukan pengambilan data dengan menggunakan beberapa teknik penambilan data sebagai berikut:

2.1 Studi Literatur

Studi Literatur adalah cara untuk menyelesaikan persoalan dengan menelusuri sumber-sumber tulisan yang pernah dibuat sebelumnya. Dengan kata lain, istilah Studi Literatur ini juga sangat familiar dengan sebutan studi pustaka. Dalam sebuah penelitian yang akan dijalankan, tentunya seorang peneliti harus memiliki wawasan yang luas terkait objek yang akan diteliti. Jika tidak, maka dapat dipastikan dalam presentasi yang besar bahwa penelitian tersebut akan gagal. Sumber-sumber yang diteliti pun tidak boleh sembarangan.

2.2 Clustering

Clustering yang berarti cluster atau pengelompokan adalah salah satu metode dalam sebuah pengelompokan data mining yang digunakan untuk mengelompokkan sebuah data yang banyak untuk melihat setiap data memiliki kemiripan dari masing-masing nilai yang bersifat homogen dan didapat melalui cara pengamatan terhadap objek-objek, data-data yang ingin diketahui cluster dari tiap data tersebut, dalam proses clustering biasanya sebuah data dibantu menggunakan pendekatan sebuah algoritma bertujuan untuk lebih memperinci dan target dalam pengelempokan data jauh lebih terstruktur dalam sistem matematis.

2.3 Algoritma K-Means

Pengelompokan data mahasiswa dengan menggunakan Algoritma K-means, yaitu terdiri dari kehadiran, nilai tugas, nilai UTS dan nilai UAS. Untuk penelitian ini kami menggunakan data dari jurnal-jurnal yang pernah meneliti objek yang sama. Algoritma K-means merupakan algoritma yang di mana akan mengelompokkan data kedalam cluster – cluster sehingga jika ada data yang memiliki karkter yang sama maka akan berada pada cluster yang sama dan data yang memiliki ketidak samaan akan berada di dalam cluster yang lain. Istilah-istilah dalam K-Means:

1. N data : data set yang akan diolah sebanyak N data dimana N data tersebut terdiri dari atribut- atributnya.
2. K centroid : Inisialisasi dari pusat cluster data adalah sebanyak K dimana pusat-pusat awal tersebut digunakan sebagai banyaknya kelas yang akan tercipta. Centroid didapatkan secara random dari N data set yang ada.
3. Euclidian Distance: merupakan jarak yang didapat dari perhitungan antara semua N data dengan K centroid dimana akan memperoleh tingkat kedekatan dengan kelas yang terdekat dengan populasi data tersebut.

Langkah-langkah dalam Algoritma K-Means Clustering:

1. Menentukan jumlah cluster .
2. Menentukan nilai centroid. Dalam menentukan nilai centroid untuk awal iterasi, nilai awal centroid dilakukan secara acak. Sedangkan jika menentukan nilai centroid yang merupakan tahap dari iterasi, maka digunakan rumus sebagai berikut :
3. Menghitung jarak antara titik centroid dengan titik tiap objek.
4. Pengelompokan objek untuk menentukan anggota cluster adalah dengan memperhitungkan jarak minimum objek.
5. Kembali ke tahap 2, lakukan perulangan hingga nilai centroid yang dihasilkan tetap dan anggota cluster tidak berpindah ke cluster lain.

3. HASIL DAN PEMBAHASAN

Tabel 1. Data Mahasiswa

NIM	Objek	Kehadiran	Tugas	UTS	UAS
1482002	A1	96	100	100	100
1481007	A2	54	64	75	67
1382002	A3	89	95	75	69
1381019	A4	96	96	100	94
1482017	A5	45	54	100	77
1382009	A6	64	95	50	78
1381030	A7	71	94	100	81
1482014	A8	100	94	100	92
1482018	A9	32	24	25	0
1481002	A10	100	96	100	90

Variabel yang digunakan dalam pengelompokan atau *clustering* yaitu Kehadiran, Tugas, UTS dan UAS. Pada Tabel 1 adalah tabel data mahasiswa yang akan dikelompokkan. Langkah Metode K-Means *Clustering* Untuk dapat melakukan pengelompokan data-data tersebut menjadi beberapa langkah yaitu:

1. Tentukan jumlah cluster yang diinginkan. Dalam penelitian ini data-data yang ada akan dikelompokkan menjadi 3 Cluster.
2. Tentukan titik pusat awal cluster (centroid). Dalam penelitian ini titik pusat awal ditentukan secara random atau acak.

Tabel 2. Titik Pusat Awal Setiap Cluster

M1 = Sebagai Mahasiswa yang lulus dengan nilai Bagus (Centroid awal adalah data ke 1)
M2 = Sebagai Mahasiswa yang lulus dengan nilai Cukup (Centroid awal adalah data ke 5)
M3 = Sebagai Mahasiswa yang tidak lulus (Centroid awal adalah data ke 9)

3. Setelah menentukan centroid awal, maka setiap data akan menemukan centroid terdekatnya yaitu dengan menghitung jarak setiap data ke masing-masing centroid menggunakan rumus korelasi euclidean distance. Adapun perhitungan centroid awal, yakni sebagai berikut:

Tabel 3. Hasil Penghitungan Jarak Untuk Masing-Masing Cluster

Jarak C1	Jarak C2	Jarak C3
0	72	160
69	30	95
41	66	125
7	68	154
72	0	112
64	67	113
32	48	136
11	70	154
160	112	0
12	70	153

- Setelah menghitung jarak data ke centroidnya, maka langkah berikutnya adalah mengelompokkan data berdasarkan jarak minimumnya. Berdasarkan nilai minimum yang telah dihasilkan pada penentuan nilai centroid maka diperoleh hasil pengelompokkan seperti terlihat pada tabel di bawah ini.

Tabel 4. Pengelompokkan Data

Jarak C1	Jarak C2	Jarak C3
0	72	160
69	30	95
41	66	125
7	68	154
72	0	112
64	67	113
32	48	136
11	70	154
160	112	0
12	70	153

- Tahap selanjutnya hitung centroid yang baru untuk setiap cluster berdasarkan data yang bergabung pada setiap clusternya.

Tabel 5. Titik Pusat Setiap Cluster Yang Baru

Keterangan	Kategori	Kehadiran	Tugas	UTS	UAS
Cluster 1	Bagus	88	96	89	86
Cluster 2	Cukup	49.5	59	87.5	72
Cluster 3	Tidak Lulus	32	24	25	0

- Setelah didapatkan centroid baru langkah berikutnya kembali lagi ke langkah 3, yakni menghitung jarak setiap data ke masing-masing centroid menggunakan rumus kolerasi antar dua obyek yaitu Euclidean Distance berdasarkan centroid baru.

Tabel 6. Pengelompokkan Data Berdasarkan Jarak Minimumnya Pada Iterasi Terakhir

Jarak C1	Jarak C2	Jarak C3
20	69	160
52	15	95
22	55	125
15	65	154
61	15	112
47	54	113
21	44	136
17	66	154
140	103	0
16	66	153

Pada penelitian ini iterasi terhenti pada iterasi ke 2 karena tidak ada lagi anggota cluster yang berpindah terbukti pada tabel di atas.

4. IMPLEMENTASI

Pengujian yang dilakukan dalam penelitian ini, iterasi clustering data nilai IPK terjadi sebanyak 2 kali iterasi. Hasil akhir yang diperoleh adalah 3 cluster, dengan:

1. Cluster pertama memiliki pusat (88, 96, 89, 86) yang dapat diartikan sebagai kelompok mahasiswa yang akan lulus dengan nilai bagus. Ada 7 mahasiswa yang termasuk dalam kelompok ini.
2. Cluster kedua memiliki pusat (49.5, 59, 87.5, 72) yang dapat diartikan sebagai kelompok mahasiswa yang akan lulus dengan nilai cukup. Ada 2 mahasiswa yang termasuk dalam kelompok ini.
3. Cluster ketiga memiliki pusat (32, 24, 25, 0) yang dapat diartikan sebagai kelompok mahasiswa yang akan tidak lulus. Ada 1 mahasiswa yang termasuk dalam kelompok ini.

Hasil Cluster juga dipengaruhi dari nilai centroid awal yang dipakai dan jumlah data yang dipakai, perbedaan pengambilan data pusat centroid awal yang dipakai juga akan mempengaruhi hasil centroid akhirnya.

Dengan menggunakan confusion matrix menunjukkan bahwa tingkat akurasi prediksi terhadap data aktual adalah 81% yang artinya algoritma K-mean cukup baik untuk memprediksi kelulusan mata kuliah.

5. KESIMPULAN

Pada penelitian penentuan predikat kelulusan Mahasiswa menggunakan pemodelan K-Means dapat diambil kesimpulan sebagai berikut:

1. Input data yang digunakan dalam penelitian ini yaitu Kehadiran Mahasiswa, Nilai Tugas, UTS dan UAS dan output ditampilkan dari data yang digunakan yaitu berupa pengelompokan data berdasarkan algoritma K-Means.
2. Minimnya data yang digunakan sebagai sampel yaitu 10 data yang sederhana agar memudahkan dalam proses analisis dan penjelasan terhadap perhitungan dalam menggunakan algoritma K-Means terhadap data yang digunakan serta untuk memudahkan pemahaman penggunaan pemodelan K-Means.
3. Pemodelan K-Means yang digunakan dalam penelitian ini memiliki keakuratan yang cukup baik terhadap permasalahan yang terjadi terkait predikat kelulusan Mahasiswa dengan hasil 81% dapat mengenali data pada 10 data yang digunakan sebagai sampel.
4. Metode lain dapat digabungkan atau dikembangkan dalam membangun sistem agar dapat diketahui perbandingan hasilnya seperti menggabungkan dengan metode subtractive clustering atau juga fuzzy dalam fuzzy C-Means.

REFERENCES

- V. Novita Sari, Y. Yupianti, and D. Maharani, (2018). "Penerapan Metode K-Means Clustering Dalam Menentukan Predikat Kelulusan Mahasiswa Untuk Menganalisa Kualitas Lulusan," *Jurteks*, vol. 4, no. 2, pp. 133–140, doi: 10.33330/jurteks.v4i2.53.
- M. R. A. Fernanda, P. Sokibi, and R. Fahrudin, (2021). "Sistem Prediksi Ketepatan Kelulusan Mahasiswa Berdasarkan Data Akademik Dan Non Akademik Menggunakan Metode K-Means (Studi Kasus : Universitas Catur Insan Cendekia)," *J. Digit*, vol. 11, no. 1, p. 89, doi: 10.51920/jd.v11i1.182.
- G. A. Pradnyana and A. A. J. Permana, (2018). "Sistem Pembagian Kelas Kuliah Mahasiswa Dengan Metode K-Means Dan K-Nearest Neighbors Untuk Meningkatkan Kualitas Pembelajaran," *JUTI J. Ilm. Teknol. Inf.*, vol. 16, no. 1, p. 59. doi: 10.12962/j24068535.v16i1.a696.
- R. Rosmini, A. Fadlil, and S. Sunardi, (2018). "Implementasi Metode K-Means Dalam Pemetaan Kelompok Mahasiswa Melalui Data Aktivitas Kuliah," *It J. Res. Dev.*, vol. 3, no. 1, pp. 22–31. doi: 10.25299/itjrd.2018.vol3(1).1773.



- R. M. Sagala, (2021). "Prediksi Kelulusan Mahasiswa Menggunakan Data mining Prediction of college subject using K-means Algorithm in Data mining," *J. TeIKa*, vol. 11, no. 2, pp. 131–142.
- B. G. Sudarsono and S. P. Lestari, (2021). "Clustering Penerima Beasiswa Yayasan Untuk Mahasiswa Menggunakan Metode K-Means," *J. Media Inform. Budidarma*, vol. 5, no. 1, p. 258. doi: 10.30865/mib.v5i1.2670.
- Halilintar R and Farina Nur I, (2018). "Penerapan Algoritma K-Means Clustering Untuk Prediksi Prestasi Nilai Akademik Mahasiswa," *J. Sains dan Inform.*, vol. 4, no. 2.
- H. Priyatman, F. Sajid, and D. Haldivany, (2019). "Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan Mahasiswa," *J. Edukasi dan Penelit. Inform.*, vol. 5, no. 1, p. 62. doi: 10.26418/jp.v5i1.29611.
- sL. Rusdiana, T. Informatika, J. G. Obos, and N. Palangka, (2009). "Kelulusan Mahasiswa Stmik Palangkaraya," no. 114.
- A. R. JANNAH, D. ARIFANTO, AND M. KOM, (2015). "PENERAPAN METODE CLUSTERING DENGAN ALGORITMA K-MEANS UNTUK PREDIKSI KELULUSAN MAHASISWA JURUSAN TEKNIK INFORMATIKA DI UNIVERSITAS MUHAMMADIYAH JEMBER," *J. MANAJ. SIST. INF. DAN TEKNO.*, VOL. 1, NO. 1210651237, PP. 1–10, 2015.