

# Penerapan Data Mining Untuk Nilai Akhir Mata Kuliah Menggunakan Algoritma K-Means Clustering

Yudi Santoso, Nurwati

Program Studi Sistem Informasi,  
Fakultas Teknologi Informasi, Universitas Budi Luhur  
Jl. Ciledug Raya, RT.10/RW.2, Jakarta Selatan, 12260  
[yudi.santoso@budiluhur.ac.id](mailto:yudi.santoso@budiluhur.ac.id)

**Abstrak**— Nilai kehadiran/absensi seringkali diabaikan oleh mahasiswa karena dianggap tidak penting. Padahal dengan absen merupakan bukti kehadiran online pada kelas yang mereka ikuti. Penelitian ini mengamati, mempelajari dan menganalisa tingkat kerajinan kehadiran mahasiswa pada perkuliahan yang mereka ambil. Kemudian data tersebut diinisiasi pengelompokkan waktu kehadiran dengan jeda 10 menit setiap kelompoknya dari jadwal kuliah 08:00 sampai dengan 10:40 wib. Dari data yang direkam pada elearning kemudian diolah dengan k-means diperoleh 2 clustering. Clustering C1 rajin berjumlah 21 mahasiswa dan clustering C2 cukup rajin 5 mahasiswa. Dengan adanya clustering ini maka Dosen pengampu mata kuliah tersebut mempertimbangkan untuk memberikan nilai tambah pada clustering rajin dan cukup rajin sehingga dapat memberikan nilai positif dalam penilaian. Penambahan nilai pada akhir mata kuliah merupakan kebijakan dosen masing-masing pengampu mata kuliah.

**Kata kunci** — Absensi; clustering; k-means

## I. PENDAHULUAN

Absensi memiliki peranan penting dalam menentukan nilai akhir mata kuliah selain nilai tugas, nilai Ujian Tengah Semester dan Nilai Akhir Semester. Namun pada penelitian ini kami meneliti nilai absensi kehadiran mahasiswa untuk mengelompokkan tingkat kerajinan mahasiswa pada tingkat rajin dan cukup rajin. Kegiatan absensi perkuliahan digunakan sebagai bukti kehadiran mahasiswa pada kelas perkuliahan yang diambil. Data yang diolah melihat waktu absensi kehadiran pada saat mahasiswa absen. Untuk data yang diambil pada pertemuan genap perkuliahan karena pada pertemuan genap ini biasanya diberikan tugas yang dikirim lewat online dengan terbatas waktu pengiriman. Tujuan penelitian ini digunakan oleh dosen sebagai salah satu pertimbangan untuk menentukan atau menambah nilai bagi mahasiswanya.

Menurut (Putri & Erlinda, 2014) memaparkan seberapa besar pengaruh kehadiran terhadap nilai akhir mahasiswa. Metode yang digunakan Fuzzy Quantification Theory I. Adapun kuantifikasi merupakan metode yang menggunakan data-data kasar seperti evaluasi dan pendapat orang yang mana kuantitas dan pemahaman tentang data-data tersebut tidak secara normal diekspresikan secara numeris. Untuk membandingkan pendapat atau evaluasi yang dipresentasikan dalam bentuk kualitatif secara linguistic maka akan lebih mudah jika ekspresi kualitatif tersebut diganti dengan bentuk numeris. Suksesnya proses belajar bagi mahasiswa dapat dilihat dengan alat ukur berupa nilai akhir yang diperoleh mahasiswa. Kelebihan pada penelitian ini adalah menghasilkan nilai kehadiran mahasiswa rata-rata hanya mampu mempengaruhi sebesar 52,70% terhadap nilai

akhir masyarakat karena faktor utama yang paling berpengaruh adalah pada nilai UAS, karena presentasi penilaiannya yang paling tinggi yaitu 35%. Faktor – faktor kualitatif tersebut sangat berpengaruh apabila kehadiran mahasiswa lebih dari 4 kali pertemuan dan dibawah 70%, maka mahasiswa tidak bisa mengikuti ujian akhir semester (UAS) [1].

B. Poerwanto dan R. Y. Fa'rifah (2016) menyatakan mengenai pengelompokkan kemampuan mahasiswa berdasarkan penilaian sebelum UTS. Mata kuliah yang diteliti dalam mata kuliah statistika yang merupakan mata kuliah wajib program studi teknik informatika Universitas Cokroaminoto Palopo (UNCP) yang diprogramkan pada semester kedua. Pembelajaran statistika mahasiswa angkatan 2015 masih tergolong rendah, hal ini dibuktikan dari perolehan nilai rata-rata dari 114 mahasiswa yaitu 69,87. Ada pun metode yang dapat digunakan dalam mengelompokkan kemampuan mahasiswa, salah satu yang sering digunakan adalah analisis cluster k-means. Hasil analisis cluster k-means dalam mengelompokkan mahasiswa ke dalam tiga kelompok berdasarkan hasil belajarnya. Setelah dikelompokkan, terdapat 3 orang yang masuk pada kategori rendah, 27 orang pada kategori sedang dan lebih dari 70% pada kategori tinggi [2].

A. Wardhani (2016) menyatakan mengenai perangkan kesehatan pada suatu daerah masih menggunakan cara manual yaitu perhitungan yang masih menggunakan rata-rata seluruh hasil indikator atau didasarkan pada distribusi data pada setiap daerah, pengolahan data indikator data juga masih menggunakan teknik statistik dasar, ini menghasilkan output yang kurang maksimal dan memiliki permasalahan pada konsistensi data pada setiap Dinas Kesehatan [3]. Tumpukan data yang berada pada dinas kesehatan, poliklinik maupun rumah sakit dan puskesmas. Sehingga dapat dirumuskan kesimpulan inialisasi jumlah cluster sebanyak 2 buah sesuai dengan pendefinisian nilai k dengan jumlah cluster akut ada 376 item, cluster tidak akut ada 624 item dengan total jumlah data adalah 1000.

Penelitian ini menggunakan teknik data mining yaitu Clustering. Diharapkan dengan data mining ini dapat mengubah data menjadi pengetahuan, data harus diolah menjadi pengetahuan sehingga bermanfaat bagi manusia. Informasi dari sudut pandang sistem informasi adalah data yang telah diolah menjadi sebuah bentuk yang berarti bagi penerimanya dan bermanfaat dalam mengambil keputusan saat ini atau mendatang [4].

## II. METODE

Dalam kegiatan pengumpulan data untuk penelitian ini pengumpulan studi pustaka yang mana pada metode ini kegiatan

dilakukan adalah mempelajari, mencari dan mengumpulkan data yang berhubungan dengan penelitian ini [5]. Data yang diperoleh kemudian akan diolah menggunakan metode k-means dengan mengambil nilai-nilai dari setiap atribut pada data untuk mengelompokkan data kehadiran mahasiswa.

### A. 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 [5]. Pada dasarnya algoritma k-means hanya mengambil sebagian dari banyaknya komponen yang didapatkan untuk kemudian dijadikan pusat cluster awal, pada penentuan pusat cluster ini dipilih secara acak dari populasi data. Kemudian algoritma k-means akan menguji masing-masing dari setiap komponen dalam populasi data tersebut dan menandai komponen tersebut ke dalam salah satu pusat cluster yang telah didefinisikan sebelumnya tergantung dari jarak minimum antar komponen dengan tiap – tiap pusat cluster. Selanjutnya posisi pusat cluster akan dihitung kembali sampai semua komponen data digolongkan ke dalam tiap-tiap cluster dan terakhir akan terbentuk cluster baru [6].

Langkah-langkah melakukan clustering dengan metode K-Means [7] adalah sebagai berikut:

- Pilih jumlah cluster k.
- 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.
- Alokasikan semua data/ objek ke cluster terdekat.

$$D(i, j) = \sqrt{(C_{1i} - C_{1j})^2 + (C_{2i} - C_{2j})^2 + \dots + (C_{ki} - C_{kj})^2} \quad (1)$$

Keterangan:

Dimana:

$D(i, j)$  = jarak data ke  $i$  ke pusat cluster  $j$

$C_{ki}$  = Data ke  $i$  pada atribut data ke  $k$

$C_{kj}$  = Data ke  $j$  pada atribut data ke  $k$

- 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.
- 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.

### B. Pengertian Rajin atau Tepat Waktu Absen Kehadiran

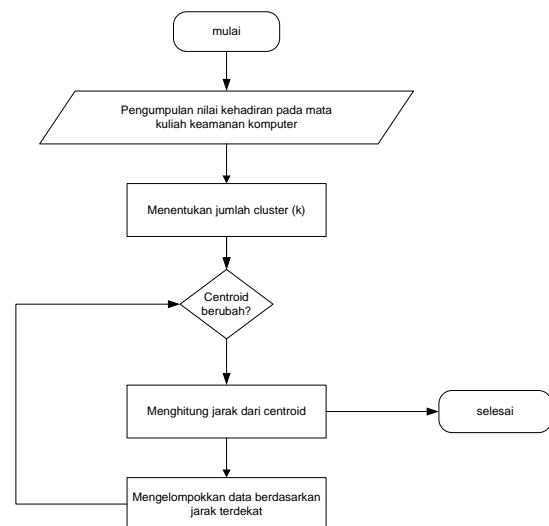
Prestasi belajar siswa dalam waktu tertentu dapat dijadikan tolak ukur tingkat keberhasilan belajar siswa di sekolah dimana tingkat kehadiran siswa di sekolah maupun didalam kelas juga merupakan salah satu faktor yang mempengaruhi prestasi belajar siswa. Kehadiran di sekolah merupakan faktor penting dalam keberhasilan sekolah. Seorang guru tidak hanya menilai prestasi siswa hanya berdasarkan nilai yang diperolehnya melalui tes ataupun ujian tetapi juga melakukan penilaian yang salah satunya berasal dari tingkat kehadiran siswa. Siswa yang rajin masuk memberikan nilai positif tersendiri dalam penilaian. Selanjutnya, disiplin dalam kehadiran di sekolah sangat dibutuhkan untuk meningkatkan hasil belajar yang optimal [8].

### C. Variabel Penelitian

Penelitian ini menggunakan rekaman data di elearning Universitas Budi Luhur semester genap tahun akademik 2020/2021 mata kuliah Keamanan Komputer program studi Teknik Informatika. Jumlah mahasiswa yang menjadi obyek penelitian adalah sebanyak 31 mahasiswa. Pengumpulan data dimulai dari pertemuan 2 tanggal 8 Maret 2021 hingga pertemuan 14 tanggal 31 Mei 2021. Nama variable di Tabel 1 berasal dari kode untuk durasi kehadiran absen pada *e-learning*.

TABEL 1. VARIABEL PENELITIAN

Variabel	Nama variable
x1	id-2
x2	id-4
x3	id-6
x4	id-8
x5	id-10
x6	id-12
x7	id-14



Gambar 1. Alur algoritma k-means

Penelitian ini akan mengelompokkan mahasiswa berdasarkan nilai kehadiran mahasiswa pada mata kuliah keamanan komputer dengan menggunakan algoritma K-Means. Adapun alur algoritma k-means penelitian ini ada pada Gambar 1.

### III. HASIL DAN DISKUSI

Data penelitian yang sedang dilakukan merupakan data *e-learning* mahasiswa pada nilai kehadiran sebanyak 31 mahasiswa. Pengelompokan tersebut berdasarkan pada level waktu kehadiran yang dikelompokkan dalam durasi 10 menit. Kehadiran di atur sesuai dengan jadwal kelas perkuliahan. Untuk penelitian ini jadwal kelas nya mulai pukul 08:00-10.40 wib. Atribut pada Tabel 2 Data Kehadiran Mahasiswa adalah No, Nama, P-2 (Pertemuan ke 2), P-4 (Pertemuan ke 4), P-6 (Pertemuan ke 6), P-8 (Pertemuan ke 8), P-10 (Pertemuan ke 10), P-12 (Pertemuan ke 12), P-14 (Pertemuan ke 14). Data kemudian dikelompokkan kriteria “RAJIN (C1)” dan “KURANG RAJIN (C2)” pengelompokkan tersebut berdasarkan atribut id. Tabel 3 merupakan pengelompokan waktu kehadiran.

TABEL 2. DATA KEHADIRAN MAHASISWA

No	Nama	P-2	P-4	P-6	P8	P-10	P-12	P-14
1	Adhitya Alamsyah	08:09	08:34	08:03	08:11	09:29	08:31	08:56
2	Adnan Yazid	08:43	08:17	09:04	09:09	09:27	08:53	08:57
3	Andri Maskhuli	-	-	09:01	08:15	10:27	-	-
4	Ardin Noor	08:00	08:15	09:05	08:12	09:58	-	09:02
...	...	...	...	...	...	...	...	...
28	Titan Ardhihiro	09:01	08:02	08:07	08:23	13:40	09:25	08:51
29	Widya Sari	08:02	08:16	08:09	16:29	09:29	08:18	09:11
30	Yovani Ekabahari	10:11	08:13	08:12	09:29	10:39	-	08:45
31	Zulfa Khalid	08:44	08:49	08:05	08:18	09:32	08:40	08:49

TABEL 3. INISIASI PENGELOMPOKAN WAKTU KEHADIRAN

Level	Jam-1	Jam-2
1	08:00	08:10
2	08:11	08:20
3	08:21	08:30
4	08:31	08:40
5	08:41	08:50
6	08:51	09:00
7	09:01	09:10
8	09:11	09:20
9	09:21	09:30
10	09:31	09:40
11	09:41	09:50
12	09:51	10:00
13	10:01	10:10
14	10:11	10:20
15	10:21	10:30
16	10:31	10:40
17	0	0

Berikut Tabel 4 Data Kehadiran Mahasiswa Yang Sudah di Inisiasi Pengelompokan Waktu Kehadiran. Tabel 4 ini diolah

dengan Ms Excel dahulu menggunakan formula vlookup. Kode id-2 adalah kode unik untuk pertemuan ke 2. Kode id-4 adalah kode unik untuk pertemuan ke 4 dan seterusnya hingga id-14 adalah kode unik untuk pertemuan ke 14. Contoh mahasiswa no 1 pada pertemuan 2 absen pada pukul 08:09. Pada tabel inisiasi pengelompokkan pukul 08:09 masuk pada level 1 karena level 1 berjeda 08:00 sampai dengan 08:10, sehingga data mahasiswa no 1 pada pertemuan 2 masuk pada level 1. Pada pertemuan 4 absen mahasiswa no 1 pada pukul 08:34. Pada tabel inisiasi pengelompokkan waktu pukul 08:34 masuk pada level 4 yaitu pukul 08:31 sampai dengan 08:40, sehingga data mahasiswa pada pertemuan 4 masuk pada level 4 dan seterusnya hingga pada pertemuan 14 absen mahasiswa no 1 pada pukul 08:56. Pada tabel inisiasi pengelompokkan waktu pukul 08:56 masuk pada level 6 yaitu pukul 08:51 sampai dengan 09:00, sehingga data mahasiswa pada pertemuan 14 masuk pada level 6. Setelah di saring dengan atribut id-n nilai id 17 muncul sebanyak 3 kali maka tidak masuk dalam perhitungan, sehingga jumlah data sekarang ada 26 mahasiswa

TABEL 4. DATA KEHADIRAN MAHASISWA YANG SUDAH DI INISIASI PENGELOMPOKAN WAKTU KEHADIRAN

No	Nama	id-2	id-4	id-6	id-8	id-10	id-12	id-14
1	Adhitya Alamsyah	1	4	1	2	9	4	6
2	Adnan Yazid	5	2	1	7	9	6	6
3	Andri Maskhuli	17	17	7	2	15	17	17
4	Ardin Noor	1	2	7	2	12	17	7
...	...	...	...	...	...	...	...	...
28	Titan Ardhihiro	7	1	1	3	17	9	6
29	Widya Sari	1	2	1	17	9	2	8
30	Yovani Ekabahari	14	2	2	9	16	17	5
31	Zulfa Khalid	5	5	1	2	10	4	5

Penentuan pusat awal cluster diberikan oleh Tabel 5.

TABEL 5. TITIK PUSAT AWAL CLUSTER

No	id-2	id-4	id-6	id-8	id-10	id-12	id-14
1	1	4	1	2	9	4	6
8	2	2	17	14	12	17	16

Selanjutnya pusat cluster dihitung menggunakan Persamaan (1) menggunakan rumus Euclidean distance, sehingga diperoleh hasil berikut.

$$c1 = \sqrt{(1-1)^2 + (4-4)^2 + (1-1)^2 + (2-2)^2 + (9-9)^2 + (4-4)^2 + (6-6)^2}$$

$$c1 = 0$$

$$c2 = \sqrt{(2-1)^2 + (2-4)^2 + (17-1)^2 + (14-2)^2 + (12-9)^2 + (17-4)^2 + (16-6)^2}$$

$$c2 = 20$$

Dan seterusnya dilanjutkan menghitung untuk data ke-2 hingga ke-n terhadap pusat awal cluster hingga didapatkan matrik jarak.

Jarak hasil perhitungan pada langkah sebelumnya akan dilakukan perbandingan, seperti pada Tabel 6 dan dipilih jarak yang paling dekat antara data dengan pusat cluster. Jarak paling

dekat antara satu data dengan satu *cluster* tertentu akan menentukan suatu data masuk dalam *cluster* mana.

TABEL 6. PENGELOMPOKKAN DATA BERDASARKAN CLUSTER TERDEKAT

No	C1	C2
1	1	
2	1	
3	1	
4	1	
5	1	
6	1	
7		1
8		1
9	1	
10	1	
11	1	
12	1	
13	1	
14		1
15	1	
16	1	
17	1	
18	1	
19	1	
20		1
21		1
22	1	
23	1	
24	1	
25	1	
26	1	

Berdasarkan Tabel 6 didapatkan pengeompokan sebagai berikut:

C1: 1,2,3,4,5,6,9,10,11,12,13,15,16,17,18,19,22,23,24,25, 26

C2: 7,8,14,20,dan 21

Selanjutnya, cluster badi ditentukan dengan rincian sebagai berikut:

C1: (4,3 ; 1,9 ; 3 ; 4,6 ;13,1 ;7,8 ;8,2)

C2: (3 ; 3,8 ; 11,4 ; 9,4 ; 14,4 ; 7,4 ; 14,8)

Maka didapat pusat cluster baru, seperti pada Tabel 7 berikut:

TABEL 7. USAT CLUSTER BARU

C1	4,3	1,9	3	4,6	13,1	7,8	8,2
C2	3	3,8	11,4	9,4	14,4	7,4	14,8

Iterasi selanjutnya dilakukan dengan cara yang sama hingga tidak ada perubahan data dalam cluster.

#### IV. KESIMPULAN

Dari hasil penelitian yang telah dilakukan dapat disimpulkan bahwa:

1. Pada penelitian ini dilakukan iterasi sebanyak 3 kali iterasi. Karena tidak ada perubahan data dalam cluster. Hasil pengelompokan data berdasarkan cluster terdekat C1: data no 1,2,3,4,5,6,9,10,11,12,13,15,16,17,18,19,22,23,24,25 dan 26 dan C2: data no 7,8,14,20,dan 21. Sama halnya nilai pusat

cluster tidak berubah yaitu C1= (4,3 ; 1,9 ; 3 ; 4,6 ;13,1 ;7,8 ;8,2) dan C2= (3 ; 3,8 ; 11,4 ; 9,4 ; 14,4 ; 7,4 ; 14,8).

2. Pada penelitian ini menghasilkan pengelompokan clustering. Terdapat 2 clustering yaitu clustering rajin dan clustering cukup rajin.
3. Nilai absensi kehadiran mahasiswa untuk pengelompokan tingkat kerajinan mahasiswa pada tingkat rajin (C1) sebanyak 21 mahasiswa dan tingkat kerajinan cukup rajin(C2) sebanyak 5 mahasiswa.
4. Sesuai dengan tujuan penelitian ini digunakan oleh dosen sebagai salah satu pertimbangan untuk menentukan atau menambah nilai akhir matakuliah bagi mahasiswa sebanyak 21 mahasiswa yang berada pada clustering 1. Penambahan nilai pada akhir mata kuliah merupakan kebijakan dosen masing-masing pengampu mata kuliah.

#### DAFTAR PUSTAKA

- [1] A. E. Putri, A. Eka, dan S. Erlinda, "Menggunakan Metode Fuzzy Quantification Theory 1," *J. SATIN - Sains dan Teknol. Inf.*, vol. Vol. 3, No, hal. 32–39, 2014.
- [2] B. Poerwanto dan R. Y. Fa'rifah, "Analisis Cluster K-Means dalam Pengelompokan Kemampuan Mahasiswa," *J. Sci. Pini*, vol. 2, no. 2, hal. 92–96, 2016.
- [3] A. K. Wardhani, "K-Means Algorithm Implementation for Clustering of Patients Disease in Kajen Clinic of Pekalongan," *J. Transform.*, vol. 14, no. 1, hal. 30, 2016.
- [4] S. Ati, Nurdien, Kistanto, dan A. Taufik, "Pengantar Konsep Informasi, Data, dan Pengetahuan," *Univ. Terbuka*, hal. 230, 2014.
- [5] J. O. Ong, "Implementasi Algoritma K-means clustering untuk menentukan strategi marketing president university," *J. Ilm. Tek. Ind.*, vol. 12, no. no. juni, hal. 10–20, 2013.
- [6] S. Agustina, D. Yhudo, H. Santoso, N. Marnasusanto, A. Tirtana, dan F. Khusnu, "CLUSTERING KUALITAS BERAS BERDASARKAN CIRI FISIK MENGGUNAKAN METODE K-MEANS Algoritma," *Clust. K-Means*, 2012.
- [7] Agus Nur Khormarudin, "Teknik Data Mining: Algoritma K-Means Clustering," *J. Ilmu Komput.*, hal. 1–12, 2016.
- [8] A. Yudiawan, "Analisis Korelasi Tingkat Absensi dengan Hasil Belajar Siswa MTs . Sains al-Gebra Kota Sorong Papua Barat," *J. Kependidikan*, vol. 11, no. 2, hal. 353–373, 2019.