

Analisis Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Keaktifan Peserta Pelatihan Pembuatan Dokumen Ilmiah Secara Daring

Abdul Rohman¹, Sri Mujiyono²

^{1,2}SI Teknik Informatika, Universitas Ngudi Waluyo

Email: abdulrohman15@gmail.com¹, mujiyn80@gmail.com²

Abstrak— Pelatihan online atau daring pada masa pandemik covid-19 merupakan salah satu solusi kegiatan pengembangan diri dan keterampilan yang mematuhi protokol kesehatan, akan tetapi sejauhmana keaktifan peserta dalam kegiatan tersebut. Untuk mengetahui hal tersebut diperlukan suatu penelitian yang menganalisis cluster yang tergolong aktif dan tidak aktif menggunakan data mining algoritma K-Means, dimana algoritma ini membentuk cluster dengan adanya kemiripan dataset. Dalam penelitian ini menggunakan atribut kehadiran, nilai tugas dan keterangan kelulusan untuk mengetahui keaktifan peserta dalam pelatihan. Hasil yang diperoleh dari 194 peserta diperoleh cluster aktif sebanyak 182 orang dan cluster tidak aktif yaitu 11 orang.

Kata kunci: Algoritma, K-Means, Keaktifan, Peserta, Pelatihan

I. PENDAHULUAN

Pada saat pandemi covid-19 ini, banyak kegiatan pembelajaran, seminar, pelatihan dan lain-lain dilaksanakan secara daring atau online. Hal ini dilakukan untuk menghindari kerumunan masa yang memungkinkan akan penyebaran virus covid-19.

E-guru.id merupakan salah satu organisasi disemarang yang berdiri tahun 2019 memberikan beragam fasilitas untuk menunjang peningkatan kompetensi akademisi baik guru, dosen, instruktur maupun pelaku Pendidikan lainnya, yaitu dengan pelatihan, seminar, workshop secara daring atau online.

Salah satu pelatihan yang diminati adalah "Pelatihan Pembuatan Dokumen Ilmiah Serba Otomatis dengan Ms Word" dengan jumlah peserta sekitar 190 sampai dengan 300 peserta yang tersebar diseluruh Wilayah Indonesia, dengan durasi waktu 36 jam dalam 6 hari. Pelaksanaan dilakukan secara daring/online dengan menggunakan telegram.

Dalam pelatihan ini, setiap peserta akan memiliki tingkat aktifannya masing-masing, yang dapat dilihat dari persentase kehadiran dan pengumpulan tugas. Sehingga dengan dua aspek tersebut menentukan kelulusannya.

Yang menjadi permasalahannya adalah bagaimana menentukan dan menganalisa tingkat aktifan peserta dalam pelatihan secara daring, sehingga dapat memberikan informasi yang berguna untuk masukan dalam penyelenggaraan pelatihan online.

Algoritma K-Means merupakan metode data mining yang banyak digunakan untuk mengidentifikasi suatu kelompok yang alami dari sebuah kasus yang didasarkan pada pengelompokkan data peserta pelatihan yang memiliki kemiripan sehingga hasilnya dari pengelompokkan dapat dianalisis [1].

Maka dalam penelitian ini, penulis akan mengolah data peserta pelatihan menggunakan metode data mining algoritma k-means dengan menggunakan dua aspek yaitu kehadiran dan tugas.

II. TINJUAN PUSTAKA

1. Keaktifan Peserta Pelatihan

Menurut Nana Sudjana, keaktifan siswa atau peserta dalam pembelajaran atau pelatihan ditentukan dalam keikutsertaan dalam pemecahan masalah dan penugasan sehingga tujuan pembelajaran dapat tercapai [2].

Maka dalam penelitian ini aspek kehadiran merupakan salah satu kegiatan dalam keikutsertaan dalam pemecahan masalah yaitu bagaimana cara membuat dokumen ilmiah secara otomatis menggunakan Ms Word. Dan aspek penugasan memperlihatkan keaktifan peserta dalam pelatihan yang sekaligus untuk mengetahui kemampuan dan keterampilan peserta dan menyerap materinya.

2. Algoritma K-Means

Algoritma k-means merupakan metode data mining clustering dengan mengolah dataset menjadi kelompok-kelompok yang berdasarkan dari tingkat kemiripan atribut-atribut data [3].

Untuk menerapkan algoritma K-means dalam pengolahan dataset terdapat 6 langkah [3]:

- 1) Menyiapkan Dataset dalam penelitian ini menggunakan dataset rekap kelulusan pelatihan.
- 2) Menentukan jumlah klaster atau kelompok, dalam penelitian ini ditentukan 2 klaster yaitu klaster Aktif dan Klaster Tidak aktif.
- 3) Memilih titik centroid secara acak

- 4) Mengelompokkan data sehingga terbentuk kluster dengan dengan titik centroid dari setiap kluster yaitu dengan menggunakan rumus:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

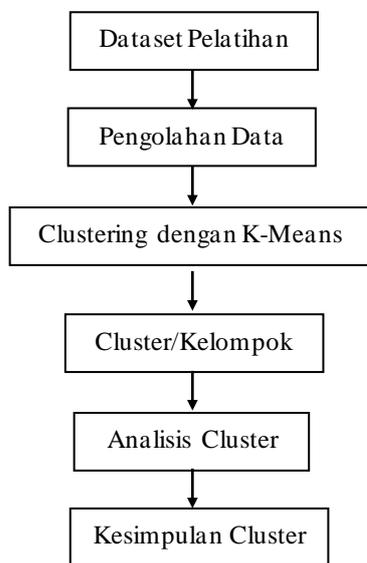
- 5) Memperbaharui nilai titik centroid, dengan rumus:

$$\mu = \frac{1}{N_k} \sum_{i=1}^{N_k} x_i$$

- 6) Dan terakhir mengulangi Langkah yang ke-3 sampai ke-5, sampai menghasilkan nilai dari titik centroid tidak berubah [3].

III. METODE PENELITIAN

Untuk mencapai tujuan penelitian ini yaitu menganalisis pola dataset dari pelatihan maka kerangka pemikiran adalah sebagai berikut:



Gambar 1 Kerangka Pemikiran

Dalam penelitian ini menggunakan data primer, yaitu data yang diambil langsung dari tempat penelitian [4].

Bentuk dataset dalam penelitian ini yaitu rekap kelulusan yang didapat dari salah satu Angkatan di pelatihan membuat dokumen ilmiah serba otomatis

dengan Ms Word di e-guru.id dengan jumlah peserta 194 orang yang terdiri dari guru, dosen dan pegawai/karyawan. Atribut yang digunakan adalah (1) status kehadiran dari pertemuan pertama sampai dengan pertemuan ke enam, (2) Nilai Tugas, (3) Jumlah Kehadiran, dan (4) keterangan lulus/tidak lulus.

Untuk tahap pengolahan data dilakukan pengskoran ditiap masing-masing atribut yaitu:

- Atribut status kehadiran diberikan skor 1 untuk status hadir dan skor 0 untuk tidak hadir
- Atribut Tugas diberikan skor 1 untuk peserta yang mengumpulkan tugas dan skor 0 untuk peserta yang tidak mengumpulkan tugas.
- Jumlah kehadiran dengan skor 1,2,3,4,5 dan 6
- Keterangan lulus/tidak lulus dengan skor 1 untuk peserta yang lulus dan 0 untuk peserta yang tidak lulus.

Tahap selanjutnya clustering menggunakan algoritma k-means dengan beberapa kluster yang disesuaikan dengan tujuan penelitian, dimana kluster yang terbentuk memiliki kesamaan dan kemiripan data [5].

Dan pada tahap clustering dan analisis dataset untuk menentukan kluster keaktifan peserta dalam pelatihan menggunakan software framework yaitu Rapid Miner yang cocok untuk penelitian data mining [6][7] [8][9].

IV. HASIL DAN PEMBAHASAN

Dataset yang pada awalnya terdapat atribut yaitu nama, asal institusi, jenis kelamin, status pekerjaan,, presensi kehadiran, jumlah kehadiran, nilai tugas dan keterangan kelulusan, diolah menjadi 4 atribut yang terkait dengan keaktifan peserta dalam pelatihan yaitu; status kehadiran perpertemuan, jumlah/nilai presensi kehadiran, nilai tugas dan status keterangan kelulusan

Tabel 1 Dataset Rekap Kelulusan Peserta Pelatihan

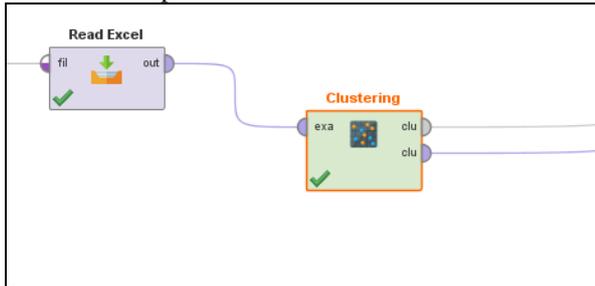
No	P1	P2	P3	P4	P5	P6	Tgs	Ktrg	Khdr
1	1	1	1	1	1	1	1	1	6
2	1	1	1	1	1	1	0	0	6
3	1	1	1	1	1	1	1	1	6
4	1	1	1	1	1	1	1	1	6
5	1	1	0	0	0	1	0	0	3
6	1	1	1	1	1	1	1	1	6
7	1	1	1	1	1	1	1	1	6
8	1	1	1	1	1	1	1	1	6
9	0	1	0	0	0	1	0	0	2

Dan seterusnya

Tabel 1. Diatas memiliki 4 atribut status kehadiran dari pertemuan 1 sampai dengan pertemuan 6 (P1-P6) yang memiliki skor 1 untuk status hadir dan skor 0 untuk status tidak hadir. Atribut tugas (tgs) terdapat 2 skor yaitu skor 1

untuk selesainya penugasan dan skor 0 untuk tidak pengumpulan tugas, atribut keterangan kelulusan (ktrg) dengan skor 1 untuk yang lulus dan skor 0 untuk yang tidak lulus. Dan atribut kehadiran (Khdr) untuk nilai kehadiran yaitu 1-6.

Selanjutnya dilakukan tahap eksperimen clustering dengan algoritma K-Means yang dikelompokkan menjadi 2 cluster dengan asumsi yaitu; (1) kelompok yang aktif dan (2) kelompok yang tidak aktif. Dalam penelitian ini eksperimen dilakukan dengan menggunakan software framework Rapid Miner Versi 9.3.



Gambar 2 Proses Clustering K-Means

Proses Clustering k-means tersebut diatas menggunakan iterasi maksimal 10 kali untuk menentukan clusternya [10], [11].

Dari eksperimen tersebut diatas menghasilkan 2 cluster dengan data yaitu sebagai berikut:

Tabel 2 Hasil Clustering

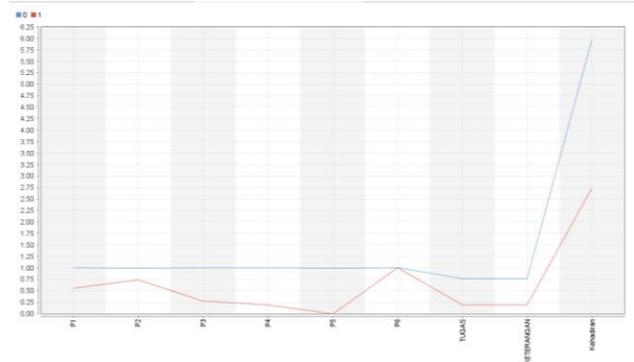
Cluster	Jumlah
0	183
1	11

Dari table 2 tersebut diatas menghasilkan nilai centroid sebagai berikut:

Tabel 3. Nilai Centroid Clustering K-Means

Attribute	cluster_0	cluster_1
P1	0.995	0.545
P2	0.989	0.727
P3	1	0.273
P4	0.995	0.182
P5	0.984	0
P6	1	1
TUGAS	0.760	0.182
KETERANGAN	0.760	0.182
Kehadiran	5.962	2.727

Secara visual dapat dilihat penyebaran cluster_0 dan cluster_1



Gambar 3. Grafik Penyebaran Cluster_0 dan Cluster_1

Dengan melihat tabel 3 dan grafik tersebut diatas, bahwa nilai tertinggi terdapat pada cluster_0 yaitu 5,962, sedangkan nilai terendah di cluster_1 yaitu 0. Maka dapat diartikan cluster_0 adalah peserta pelatihan yang aktif dengan jumlah 183 orang, sedangkan cluster_1 adalah peserta pelatihan yang tidak aktif dengan jumlah 11 orang.

Tabel 4. Jumlah Peserta pelatihan yang aktif dan tidak aktif

Cluster	Aktif	Tidak Aktif
0	183	
1		11

V. KESIMPULAN

Dari hasil penelitian mengelompokkan peserta pelatihan “membuat dokumen ilmiah serba otomatis dengan word” dalam 2 cluster yaitu kelompok 1 yang aktif berjumlah 183 orang dan yang tidak aktif berjumlah 11 orang. Sehingga dengan hasil tersebut dapat dijadikan evaluasi penyelenggara terhadap aspek-aspek yang perlu diperhatikan dan dipertimbangkan dalam pelaksanaan berikutnya.

VI. REFERENSI

- [1] R. T. Vuldari, *Data Mining Teori dan Aplikasi Rapidminer*. Yogyakarta: Gava Media, 2017.
- [2] N. Sudjana, *Dasar-dasar Proses Belajar Mengajar*. Bandung: Sunar Baru Algensindo, 2005.
- [3] Joko Suntoro, *Data Mining Algoritma dan Implementasi dengan Pemrograman PHP*. Jakarta: Elex Media Computindo, 2019.
- [4] W. Purhantara, *Metode penelitian kualitatif untuk bisnis*. Yogyakarta: Graha Ilmu, 2010.
- [5] S. Gusriani, K. D. K. Wardhani, and M. I. Zul, “Analisis Sentimen Terhadap Toko Online di Sosial Media Menggunakan Metode Klasifikasi Naïve Bayes (Studi Kasus: Facebook Page BerryBenka),” *4th Appl. Bus. Eng. Conf.*, vol. 1, no. 1, pp. 1–7, 2016.
- [6] Soni and Ganatra, “Categorization of several clustering algorithms from different perspective,” *J.*

Adv. Res. Comput. Sci., pp. 63–68, 2012.

- [7] A. Nurzahputra, M. A. Muslim, and M. Khusniati, “Penerapan Algoritma K-Means Untuk Clustering Penilaian Dosen Berdasarkan Indeks Kepuasan Mahasiswa,” *Techno.Com*, vol. 16, no. 1, pp. 17–24, 2017, doi: 10.33633/tc.v16i1.1284.
- [8] F. Febrian, “Algoritma Klasifikasi Data Mining Pada Akseptasi Data Fakultatif,” *Program*, 2011.
- [9] B. Rifai, P. Pascasarjana, M. Ilmu, S. Tinggi, M. Informatika, and D. a N. Komputer, “Model Neural Network Berbasis Metode Model Neural Network Berbasis Metode,” 2012.
- [10] Y. Prastyo, “Pembagian Tingkat Kecanduan Game Online Menggunakan K-Means Clustering Serta Korelasinya Terhadap Prestasi Akademik,” *Elinvo (Electronics, Informatics, Vocat. Educ.)*, vol. 2, no. 2, p. 138, 2017, doi: 10.21831/elinvo.v2i2.17307.
- [11] B. R. Setiyono, Budi; Isnanto, “Pembagian kelas kuliah mahasiswa menggunakan algoritma pengklasteran fuzzy c-means,” *Pros. Semin. Nas. Teknoin 2008*, no. 1, pp. 15–20, 2008.