

PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA K-MEANS UNTUK MENENTUKAN JUDUL SKRIPSI DAN JURNAL PENELITIAN (STUDI KASUS FTIK UNPRI)

Cornelia Selvi Dinta Sembiring, Latifah Hanum, Saut Parsaoran Tamba
Program Studi Sistem Informasi, Fakultas Teknologi dan Ilmu Komputer, Universitas Prima Indonesia
Email: sautparsaorantamba@unprimdn.ac.id

ABSTRAK- Kemajuan teknologi saat ini berpengaruh pesat termasuk dalam bidang pendidikan khususnya dalam perkuliahan untuk menentukan judul skripsi dan jurnal penelitian. Dalam hal ini para pengembang menemukan suatu pola untuk mempermudah dalam pencarian ide judul untuk menyelesaikan syarat kelulusan perkuliahan dalam lingkup jurusan Sistem Informasi di Universitas Prima Indonesia. Untuk mengatasi permasalahan tersebut dengan menerapkan metode Algoritma K-Means. Metode tersebut bertujuan untuk mengelompokkan data mahasiswa seesuai dengan skill dan basic yang didominasi pada mata kuliah yang paling banyak diminati sebagai acuan dalam pengembangannya. Dengan adanya pengolahan data yang dilakukan dapat memberikan solusi kepada mahasiswa dan lingkungannya untuk mengetahui ide judul skripsi dan jurnal penelitian. Maka hasil uji coba mendapatkan perbandingan score dalam pembagian clustering yaitu pada 29% C1, 21% C2, 22% C3, 13% C4, 15% C5

Kata kunci : Data Mining, Clustering, K-Means

1. PENDAHULUAN

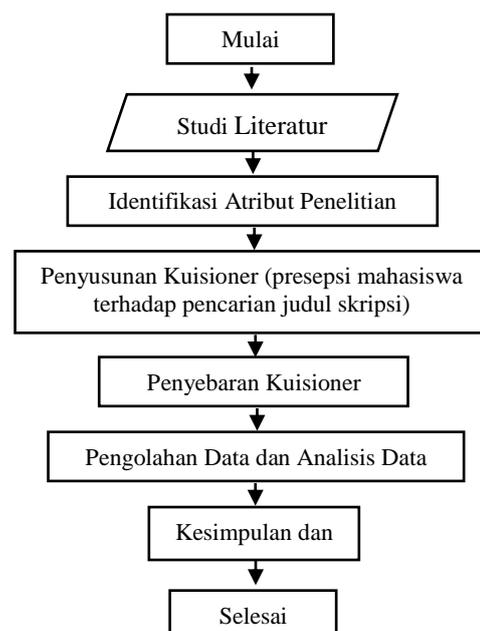
Skripsi adalah suatu karya tulis ilmiah yang digunakan oleh mahasiswa/i untuk menyelesaikan pendidikan di jenjang sarjana dengan memaparkan tulisan hasil penelitian yang membahas suatu permasalahan dalam bidang ilmunya. Dalam waktu yang sudah ditentukan untuk menyelesaikan jenjang sarjana, para mahasiswa selalu bingung untuk menentukan penelitian yang ingin dibuat. Dalam tahap pertama mahasiswa harus menentukan judul skripsi yang akan dikembangkan dan konsultasi dengan dosen akan hal penelitian tersebut. Sangat banyak problem yang terjadi ketika karya tulis ilmiah yang merupakan skripsi tersebut mahasiswa tidak mengerti untuk mencari judul. Sehingga terhambatnya dalam menyelesaikan skripsi.

Sistem yang dirancang bertujuan untuk membantu mahasiswa dalam mencari ide judul skripsi menggunakan Data Mining dengan Clustering K-Means. Sehingga terdapatnya kriteria pendukung untuk menentukan judul skripsi maupun jurnal penelitian. Adapun penelitian sebelumnya dengan metode yang berbeda, yaitu Sistem Pendukung Keputusan untuk Menentukan Judul Skripsi Jurusan Teknik Informatika Menggunakan Metode Topsis dan Sistem Rekomendasi Penentuan Judul Skripsi Menggunakan Algoritma Decision Tree. Pada metode Topsis memiliki hasil akhir berupa perankingan bidang studi yang dapat dijadikan acuan dalam menentukan judul skripsi, sehingga mahasiswa dibantu dalam melihat fokus bidang ilmu yang seharusnya dijadikan judul skripsi sesuai dengan kemampuan dan nilai dari kriteria yang ada. Sedangkan pada metode Algoritma Decision

Tree memiliki hasil rekomendasi, mahasiswa memilih minat tema skripsi yang mau diambil.

2. METODE PENELITIAN

Pada penelitian ini, peneliti menggunakan jenis penelitian secara kuantitatif dengan metode algoritma K-Means). Proses penelitian yang dilakukan dalam metode tersebut yang dimulai dari input, proses, dan output. Dalam penelitian tersebut menggunakan proses prosedural menggunakan flowchart. Flowchart merupakan langkah awal dalam proses perancangan program dan dengan adanya flowchart akan lebih jelas proses kegiatannya



Gambar 1. Alur Metodologi Penelitian

Algoritma K-Means merupakan salah satu algoritma clustering yang sangat umum dalam mengelompokkan data sesuai dengan kesamaan karakteristik. Kelompok data yang dihasilkan disebut sebagai cluster/klaster. Algoritma ini memiliki kelebihan, yaitu algoritma yang mudah diimplementasikan dan memiliki kompleksitas waktu dan ruang yang relatif kecil. Sehingga cukup efisien dalam komputasinya dan memberikan hasil yang cukup baik dan memuaskan jika klasternya compact, hyperspherical dalam shape dan mampu memisahkan fitur – fitur ruangnya dengan baik. Menurut Hasn & Kamber, algoritma K-Means bekerja dengan membagi data ke dalam k buah klaster yang telah ditentukan. Langkah – langkah dasar untuk algoritma k-means adalah :

- 1 Tentukan nilai titik k klaster sesuai dengan yang diinginkan
- 2 Pilih titik-titik atau sampel yang menjadi anggota klaster secara acak.
- 3 Tentukan nilai centroid atau titik tengah dari klaster tersebut dengan rumus :

$$M_k = \left(\frac{1}{n_k}\right) \sum_{i=1}^{n_k} x_{ik} \dots \dots \dots (1)$$

Dimana X_{ik} = sampel yang ada dalam klaster C_k .

- 4 Hitung square error untuk setiap klaster C_k yang merupakan jumlah kuadrat jarak Euclidean antara tiap sampel dalam C_k dan titik tengahnya (centroid). Error ini juga dikenal dengan nama within cluster variation (WCV), yaitu :

$$e_k^2 = \sum_{i=1}^{n_k} (x_{ik} - M_k)^2 \dots \dots \dots (2)$$

Setelah melakukan beberapa penelitian dan berhasil mengidentifikasi bahwa ada 4 dimensi karakteristik dalam mengevaluasi kualitas pelayanan. Keempat dimensi karakteristik kualitas pelayanan tersebut yaitu :

1. Reliability (Keandalan).
2. Responsiveness (Daya Tanggap)
3. Assurance (Jaminan)
4. Emphaty (Empati)



Gambar 2. Use Case Diagram

3. HASIL DAN PEMBAHASAN

3.1 Analisis Data

Data penelitian tersebut didapat dari mahasiswa/i FTIK UNPRI, dimana data tersebut dikumpulkan dengan cara penyebaran questioner dan memperoleh 1400 data. Output yang diharapkan memiliki 5 cluster yaitu (C1) Data Mining Berbasis Web, (C2) Aplikasi Android, (C3) Sistem Pakar Berbasis Web, (C4) Sistem Pakar Berbasis Android, dan (C5) Sistem Pendukung Keputusan Berbasis Web. Variabel yang digunakan dalam pengelompokan data terdiri dari **Bahasa Pemograman, Perancangan Aplikasi, Nilai Akademik, Jumlah Sertifikat Pelatihan, dan Peminatan Mata Kuliah**. Dibawah berikut tabel Sampel Data.

Tabel 1. Sampel Data

Nim	Nama	B.P	P.A pk	N.A	JS	P. Mk
2005103015	A1	2	1	1	1	4
203303040301	A2	2	1	1	1	4
203303041005	A3	1	1	1	1	4
203303043001	A4	4	2	1	4	1
203303043002	A5	4	2	1	4	1
203303043004	A6	2	1	1	1	4
203303043005	A7	3	2	2	2	2
203303043006	A8	3	1	2	2	2
213303040466	A9	4	2	2	2	3
213303043001	A10	1	1	2	1	1
213303043005	A11	4	2	1	4	1
213303043006	A12	1	1	1	3	1
213303043007	A13	1	1	1	5	1

213303	A14	1	1	1	5	1
043009						

3.2 Penerapan K-Means Clustering

Dalam penelitian ini ada 5 cluster, yaitu C1 = Data Mining Berbasis Web, C2 = Aplikasi Android, C3 = Sistem Pakar Berbasis Web, C4 = Sistem Pakar Berbasis Android, C5 = Sistem Pendukung Keputusan Berbasis Web. Langkah pertama yaitu menentukan titik pusat awal dari setiap cluster secara random sebagai berikut.

Table 2. Titik Pusat Cluster Awal

Cluster	Bahasa Pemograman	Perancangan Aplikasi	Nilai Akademik	Jumlah Sertifikat	Peminatan Mata Kuliah
Cluster 1	2	1	1	1	4
Cluster 2	4	2	2	2	3
Cluster 3	4	2	1	4	1
Cluster 4	1	1	1	3	1
Cluster 5	1	1	1	5	1

Kemudian hitung jarak data terdekat dengan titik pusat awal pada data ke -1 terhadap pusat cluster.

Tabel 3. Data ke-1 dan 2

Nama	Bahasa Pemograman	Perancangan Aplikasi	Nilai Akademik	Jumlah Sertifikat	Peminatan Mata Kuliah
Velinda C.	2	1	1	1	4
Krista yanti G.	4	2	1	2	4

Cluster1

$$\sqrt{(2-2)^2 + (1-1)^2 + (1-1)^2 + (1-1)^2 + (4-4)^2} = 0$$

Cluster2

$$\sqrt{(2-4)^2 + (1-2)^2 + (1-2)^2 + (1-2)^2 + (4-3)^2} = 2.828$$

Cluster3

$$\sqrt{(2-4)^2 + (1-2)^2 + (1-1)^2 + (1-4)^2 + (4-1)^2} = 4.795$$

Cluster4

$$\sqrt{(2-1)^2 + (1-1)^2 + (1-1)^2 + (1-3)^2 + (4-1)^2} = 3.741$$

Cluster5

$$\sqrt{(2-1)^2 + (1-1)^2 + (1-1)^2 + (1-5)^2 + (4-1)^2} = 5.099$$

Jarak dari data ke-2 terhadap pusat cluster

Cluster1

$$\sqrt{(4-2)^2 + (2-1)^2 + (1-1)^2 + (2-1)^2 + (4-4)^2} = 2.449$$

Cluster2

$$\sqrt{(4-4)^2 + (2-2)^2 + (1-2)^2 + (2-2)^2 + (4-3)^2} = 1.414$$

Cluster3

$$\sqrt{(4-4)^2 + (2-2)^2 + (1-1)^2 + (2-4)^2 + (4-1)^2} = 3.605$$

Cluster4

$$\sqrt{(4-1)^2 + (2-1)^2 + (1-1)^2 + (2-3)^2 + (4-1)^2} = 4.472$$

Cluster5

$$\sqrt{(4-1)^2 + (2-1)^2 + (1-1)^2 + (2-5)^2 + (4-1)^2} = 5.291$$

Hasil untuk perhitungan selengkapnya untuk iterasi 1 diatas dapat dilihat pada tabel 4.

Tabel 4. Hasil Perhitungan pada Iterasi 1

Nama	B.P	P.Apk	N.Ak	J.S	PMK	C1	C2	C3	C4	C5	Hsl
A1	2	1	1	1	4	0	2.828	4.796	3.742	5.099	Cluster-1
A2	2	1	1	1	4	0	2.828	4.796	3.742	5.099	Cluster-1
A3	1	1	1	1	4	1	3.606	5.292	3.606	5	Cluster-1
A4	4	2	1	4	1	4.796	3	0	3.317	3.317	Cluster-3
A5	4	2	1	4	1	4.796	3	0	3.317	3.317	Cluster-3
A6	2	1	1	1	4	0	2.828	4.796	3.742	5.099	Cluster-1
A7	3	2	2	2	2	2.828	1.414	2.646	2.828	4	Cluster-2
A8	1	1	2	1	1	3.317	3.873	4.472	2.236	4.123	Cluster-4
A9	4	2	1	4	1	4.796	3	0	3.317	3.317	Cluster-3
A10	1	1	1	3	1	3.742	4	3.317	0	2	Cluster-4
A11	1	1	1	5	1	5.099	4.899	3.317	2	0	Cluster-5
A12	3	1	2	2	2	2.646	1.732	2.828	2.646	3.873	Cluster-2

Kemudian ulangi perhitungan dengan titik pusat yang baru sampai hasil cluster tidak berubah lagi dan juga tidak ada data yang berpindah dari cluster ke cluster.

Setelah Semua data ditempatkan ke dalam cluster terdekat, kemudian hitung kembali pusat cluster yang baru berdasarkan rata-rata anggota yang ada pada cluster tersebut.

Pembangkitan ulang centroid baru dihitung dengan rumus sebagai berikut :

$$C = \frac{\sum m}{n} \dots \dots \dots (3)$$

Keterangan :

$C = \text{Centroid data}$

$m =$ anggota data yang termasuk kedalam centroid tertentu

$n +$ jumlah data yang menjadi anggota centroid tertentu

Berdasarkan Tabel3, titik pusat cluster baru untuk paramater stok awal memiliki 3 anggota :

$$C1 = (3.316 + 3.741 + 5.099) / 3 = 8.756,666$$

Cluster Kedua (C2) memiliki 8 anggota

$$C2 = (2.828 + 2.828 + 3.605 + 2.828 + 3.872 + 3.741 + 5.099 + 1.732) / 8 = 3.308,375$$

Cluster Ketiga (C3) memiliki 9 anggota

$$C3 = (4.795 + 4.795 + 5.291 + 0 + 0 + 4.795 + 3.316 + 0 + 2) / 9 = 2.554,888$$

Cluster Keempat (C4) memiliki 6 anggota

$$C4 = (3.741 + 3.741 + 3.605 + 5.099 + 4.123 + 3.316) / 6 = 3.937,5$$

Cluster Kelima (C5) memiliki 6 anggota

$$C5 = (3.316 + 3.316 + 4.123 + 3.316 + 2 + 0) / 6 = 2.345,5$$

Tabel 5. Titik Pusat Baru Cluster

Clus ter1	8.75 6,66	1,2777 77778	1,5	1,3888 88889	3,3888 88889
Clus ter2	3.30 8,37 5	1,6	1,8666 66667	1,8666 66667	2,8
Clus ter3	2.55 4,88 8	2	1	4	1
Clus ter4	3.93 7,5	1	1,7142 85714	1,8571 42857	1,1428 57143
Clus ter5	2.34 5,5	1	1	5	1

Setelah didapatkan titik pusat yang baru dari setiap cluster , lakukan kembali dari langkah ketiga hingga titik pusat dari setiap cluster tidak berubah lagi dan tidak ada lagi data yang berpindah dari satu cluster ke cluster yang lain.

3.3 Hasil

Berdasarkan hasil pengelompokan data menggunakan metode K-Means clustering, didapatkan hasil dengan iterasi ke-2. Hasil Cluster yang terbentuk dapat dilihat ditabel bawah tersebut.

Tabel 6. Hasil Akhir Pengelompokan

Nama	B.P	P.A pk	N.A	J.S	P.Mk	Cluster
A1	2	1	1	1	4	Cluster-1
A2	2	1	1	1	4	Cluster-1
A3	1	1	1	1	4	Cluster-1
A4	4	2	1	4	1	Cluster-3
A5	4	2	1	4	1	Cluster-

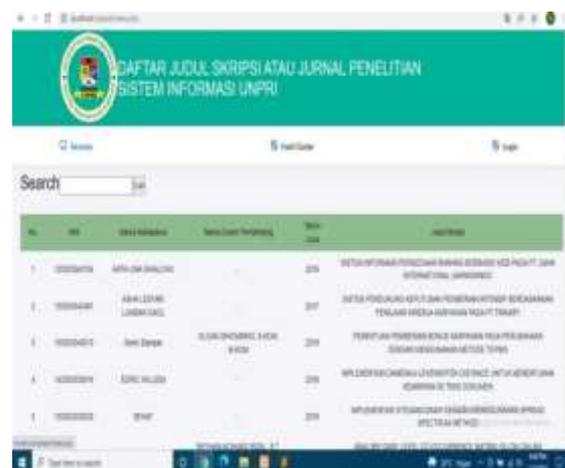
						3
A5	2	1	1	1	4	Cluster-1
A6	3	2	2	2	2	Cluster-2
A7	3	1	2	2	2	Cluster-2
A8	4	2	2	2	3	Cluster-2
A9	1	1	2	1	1	Cluster-4
A10	4	2	1	4	1	Cluster-3
A11	1	1	1	3	1	Cluster-4
A12	1	1	1	5	1	Cluster-5
A13	1	1	1	5	1	Cluster-5

Berdasarkan dari hasil uji coba tersebut dapat dilihat bahwa pengelompokan cluster diatas bahwa (C1) Data Mining Berbasis Web dengan peminat 29% , (C2) Aplikasi Android dengan peminat 21%, (C3) Sistem Pakar Berbasis Web dengan peminat 22%, (C4) Sistem Pakar Berbasis Android dengan peminat 13%, dan (C5) Sistem Pendukung Keputusan Berbasis Web dengan peminat 15%.

3.4 Analisa sistem

3.4.1 Analisa Input

Pada Sistem Penerapan data mining menggunakan algoritma K-Means untuk menentukan judul skripsi dan jurnal penelitian terdapat beberapa bagian menu yaitu tampilan judul skripsi hanya dapat dilihat oleh pengunjung.



Gambar 3. Halaman Utama

dan user yang telah melakukan pendaftaran dapat melihat tampilan judul skripsi dan login user. Pada halaman login "user" dapat dilakukan jika pengguna tersebut melakukan konsultasi terlebih dahulu ke admin untuk cluster dan data yang sudah didaftarkan oleh admin, maka pengguna bisa

4. PENUTUP

4.1 Kesimpulan

Dari Hasil Penelitian ini penulis menuliskan ada beberapa kesimpulan pada penelitian ini antara lain :

1. Dengan adanya sistem ini, mahasiswa dapat lebih mudah dalam menentukan judul skripsi dan jurnal penelitian sesuai dengan basic dan skill yang dimiliki agar dapat melakukan pengajuan penelitian ataupun pengembangannya, serta admin dapat lebih mudah dalam penyimpanan lampiran-lampiran judul skripsi dan jurnal dari setiap mahasiswa yang telah menyelesaikan skripsi.
2. Hasil pengujian sistem penerapan data mining menggunakan algoritma K-Means untuk menentukan judul skripsi dan jurnal penelitian (Studi Kasus FTIK UNPRI) menunjukkan sistem dapat berjalan dengan baik. hasil uji coba mendapatkan perbandingan score mencapai 60% tampilan aplikasi, 70% kinerja program, dan 70% manfaat program.

4.2 Saran

Untuk penelitian lebih lanjut diharapkan agar dilanjutkan dengan metode lain untuk pengelompokan data serta sistem yang lebih akurat dan gampang diakses ataupun digunakan.

DAFTAR PUSTAKA

- [1] Adrianto, R. (2016). Penerapan Metode Clustering Dengan Algoritma K-means untuk Rekomendasi Pemilihan Jalur Peminatan Sesuai Kemampuan Pada Program Studi Teknik Informatika-S1 Universitas Dian Nuswantoro. *Journal JOINS Udinus*, 101-116.
- [2] Dr Suyanto, S. (2017). Data mining untuk klasifikasi dan klasterisasi data. Bandung: Informatika Bandung. Diambil kembali dari https://suyanto.staff.telkomuniversity.ac.id/t_extbook-data-mining/
- [3] E. Turban, R. E. (2010). *Decision Support Systems and Intelligent Systems* (9th Edition). Prentice Hall.
- [4] Eko, P. (2014). *Data Mining Konsep dan Aplikasi menggunakan MATLAB*. Yogyakarta: Andi.
- [5] Farokhah, L., &. (2017). Implementasi K-Means Clustering untuk Rekomendasi Tugas Akhir Pada STMIK Asia Malang. *Jurnal Teknologi Dan Manajemen Informatika*, 3(2), 142-148. Diambil kembali dari <https://doi.org/10.26905/jtmi.v3i2.1329>
- [6] M. Sholehudin, M. (2018). Implementasi Metode Text Mining dan K-Means Clustering untuk Pengelompokan Dokumen Skripsi . (Studi Kasus : Universitas Brawijaya) "vol. 2, no.11,pp. 5518-5524.
- [7] M. Sholehudin, M. F. (2018). Implementasi Metode Text Mining and K-Means Clustering untuk pengelompokan Dokumen Skripsi. (Studi Kasus Universitas Brawijaya) ,"vol.2, no. 11, pp. 5518-5524.
- [8] O. Somantri, S.Wiyono, a. (2016). "Metode K-Means untuk Optimasi Klasifikasi Tema Tugas Akhir Mahasiswa Menggunakan Support Vector Machine (SVM)". *Sci. J. Informatics*, vol. 3, no.1, pp.34-45.
- [9] S. Sharma and ShikhaRai. (2017). "Genetic K-Means Algorithm -Implementation and Analysis". *Int. J. Recent Technol. Eng.*, vol. 1, no.2, pp. 117-120.
- [10] V.K. Bakti, J. (2017). Klasterisasi Dokumen Tugas Akhir Menggunakan K-Means clustering sebagai analisis Penerapan Sistem Temu Kembali . *J.Ilm. Manaj. Inform. and Komput.*, vol. 1, no. 1, pp. 31-34.
- [11] Tamba, S. P., & Kesuma, F. T. (2019). PENERAPAN DATA MINING UNTUK MENENTUKAN PENJUALAN SPAREPART TOYOTA DENGAN METODE K-MEANS CLUSTERING: data mining; k-means-clustering. *Jurnal Sistem Informasi dan Ilmu Komputer Prima (JUSIKOM PRIMA)*, 2(2), 67-72.
- [12] Tamba, S. P., Batubara, M. D., Purba, W., Sihombing, M., Siregar, V. M. M., & Banjarnahor, J. (2019, July). Book data grouping in libraries using the k-means clustering method. In *Journal of Physics: Conference Series* (Vol. 1230, No. 1, p. 012074), IOP Publishing.