

ANALISIS DATA PENERIMAAN MAHASISWA BARU BERBASIS K-MEANS

Anastasia L Maukar¹, Fitri Marisa^{2*}, Anang Aris Widodo³,
Nurinkamilaningtyas⁴, dan Novian Didik Nugraha⁵

¹Teknik Industri, Universitas Presiden,

^{2*,4,5}Teknik Informatika, Universitas Widyagama Malang,

³Teknologi Informasi, Universitas Merdeka Pasuruan

Email: almaukar@president.ac.id¹, fitrimarisa@gmail.com², anangariswido@gmail.com³,
nurinkamilaningtyas@gmail.com⁴, noviandidiknugraha@gmail.com⁵

Abstrak

Pendaftaran mahasiswa pada perguruan tinggi yang di selenggarakan setiap tahunnya membuahkan hasil penumpukan informasi calon mahasiswa baru. Informasi yang berlimpah tersebut alangkah baiknya diolah untuk menghasilkan analisa data yang berguna. Penelitian ini menggunakan sample data penerimaan mahasiswa baru di Universitas Widyagama Malang. Metode dilakukan dengan cara pengelompokan data (cluster) calon mahasiswa baru melalui tahap pemrosesan data / informasi menggunakan algoritma K-Means Clustering. Atribut yang digunakan dalam tahap pemrosesan data ini adalah Nama Calon Mahasiswa, jurusan yang dipilih dan Sekolah Asal. Data yang digunakan adalah data mahasiswa Angkatan 2020/2021 dengan mengambil sampel 100 baris data. Sampel yang didapatkan ini diolah dengan cara membagi 3 cluster sampai didapatkan pemodelan data yang sesuai dengan aturan algoritma K-Means Clustering. Cluster1 menunjukkan peminat jurusan ekonomi banyak dari SMA dan SMK, Cluster 2 Menunjukkan peminat jurusan Teknik dari SMK dan MAN, Cluster 3 menunjukkan peminat jurusan hukum dari SMA. Output penelitian ini dapat digunakan sebagai bahan pertimbangan untuk mengatur strategi pengenalan institusi ke sekolah menengah atas yang ada di Kota Malang khususnya. Kekhasan cluster yang dihasilkan dapat dijadikan rujukan untuk menentukan strategi promo yang lebih tepat sasaran.

Kata Kunci: Clustering, K-Means, Student Admission

Abstract

Student Admission at universities which is held annually results in the accumulation of information on prospective new students. This abundant information would be better processed to produce useful data analysis. This study uses a data sample of new student admissions at the University of Widyagama Malang. The method is carried out by grouping data (clusters) of prospective new students through the data / information processing stage using the K-Means Clustering algorithm. The attributes used in this data processing stage are the Name of the Prospective Student, the major chosen and the School of Origin. The data used is student data for the 2020/2021 class by taking 100 data lines as samples. The sample obtained is processed by dividing 3 clusters until the data modeling is obtained in accordance with the rules of the K-Means Clustering algorithm. Cluster1 shows that there are a lot of economics majors from SMA and SMK, Cluster 2 shows those who are interested in Engineering majors from SMK and MAN, Cluster 3 shows those who are interested in law majors from high school. The output of this research is used as material for consideration to regulate the strategy of introducing institutions to senior high schools in Malang City in particular. The uniqueness of the resulting cluster can be used as a reference to determine a more targeted promo strategy.

KeyWords : Clustering, K-Means, Student Admission

I. PENDAHULUAN

Tingkat perkembangan suatu daerah bisa dilihat dari kemajuan bidang pendidikan salah satunya. Dimana daerah tersebut sudah ada tingkat pendidikan dari tingkat bawah sampai tinggi, dengan banyaknya sekolah tinggi menjadi salah satu tolak ukur bahwa daerah tersebut sudah menyadari akan pentingnya pendidikan tinggi. Di Kota Malang ini terdapat banyak perguruan tinggi swasta dan negeri, yang mana ikut memberikan kontribusi menyerap lulusan dari sekolah menengah yang ingin melanjutkan ke jenjang pendidikan tinggi. Lulusan sekolah menengah baik SMA, SMK, dan MAN yang melanjutkan pendidikan tinggi kebanyakan di Kota Malang saja. Salah satu tujuan pendidikan Tinggi adalah Universitas Widyagama Malang yang terletak di Kecamatan Blimbing Kota Malang. Universitas Widyagama Malang adalah perguruan tinggi swasta yang termasuk salah satu kampus terbesar di kota Malang. Mahasiswa yang masuk ke Universitas Widyagama Malang tidak hanya siswa lulusan dari Kota Malang saja, akan tetapi dari berbagai daerah yang ada diluar Kota Malang, bahkan sampai luar pulau Jawa.

Penerimaan mahasiswa baru yang selalu diselenggarakan sebelum awal semester Gasal di tiap tahun mengumpulkan informasi calon Mahasiswa baru dari berbagai daerah. Berdasarkan informasi yang terkumpul tersebut, terdapat hal penting yang dapat diambil kesimpulan dengan cara mengolah data / informasi agar menghasilkan informasi – informasi baru. Informasi yang diperoleh dapat dipakai sebagai penunjang keputusan bagi Universitas. Pengolahan data dalam ukuran besar yang didalamnya terdapat banyak record tidak bisa dilakukan dengan mudah dan dalam waktu yang singkat untuk mendapatkan informasi yang diinginkan. Tentunya pengolahan data ini dapat dilakukan dengan menggunakan tool untuk mengolah data dalam ukuran yang besar, yaitu dengan menerapkan metode Data Mining. Penelitian dengan tema analisis data mahasiswa dengan data mining

menggunakan algoritma K-Means clustering sebelumnya telah banyak dilakukan, tetapi tempat penelitian dan hasil informasi yang berbeda. Penelitian [1] Menggunakan metode K-Means Clustering dalam menganalisa prediksi kelulusan mahasiswa dalam lingkup jurusan untuk membantu dalam pengelompokan mahasiswa yang lulus tepat waktu dan tidak. Penelitian [2] Menggunakan metode K-Means Clustering untuk mengelompokkan mahasiswa, agar mahasiswa tahu keahlian apa yang perlu mereka tingkatkan atau kembangkan diambil dari matakuliah dan nilai yang di peroleh. Sementara penelitian tentang pengelompokan data menggunakan K-Means juga telah dilakukan pada penelitian-penelitian sebelumnya di berbagai bidang. Penelitian [11] menggunakan K-Means untuk mencari pola cluster pelanggan dengan empat kriteria sehingga dapat menjadi acuan pemilik usaha untuk menentukan strategi promo yang lebih efektif. Penelitian [12] Menggunakan K-Means dikombinasi dengan metode LRFM untuk mencari pola kluster pelanggan berdasarkan tingkat kepuasannya, sehingga dapat dijadikan acuan untuk menentukan strategi pelayanan kepada pelanggan agar lebih optimal dan meningkatkan loyalitas pelanggan. Dengan demikian K-Means dapat dijadikan metode yang efektif untuk mencari pola pengelompokan data berdasarkan kriteria tertentu yang hasilnya dapat dianalisis lebih jauh sesuai kebutuhan data.

Untuk itu penelitian ini bertujuan menerapkan algoritma K-Means Clustering terhadap data yang dihasilkan dari penerimaan calon mahasiswa baru Universitas Widyagama Malang agar hasil cluster dapat dianalisis lebih lanjut untuk keperluan pengambilan keputusan strategis. Atribut data yang akan digunakan antara lain nama mahasiswa, jurusan, lulusan sekolah dan IPK. Data tersebut diperoleh dari data mahasiswa Angkatan 2020/2021 yang didapatkan dari bagian akademik dan administrasi Universitas Widyagama Malang. Pengolahan data mentah ini mengadopsi Teknik data mining algoritma k-means clustering. Hasil penelitian ini diharapkan dapat bermanfaat bagi Universitas Widyagama untuk membantu meningkatkan promosi kampus dan meningkatkan jumlah calon mahasiswa baru.

II. METODE PENELITIAN

Metode penelitian dilakukan dalam beberapa tahapan dengan mengikuti langkah metode Data Mining *K-Means Clustering*.

A. Tahapan KDD dalam Data Mining

Proses *Knowledge Discovery in Database* (KDD) adalah tahapan global dari penggalian data yang terdiri dari beberapa langkah / tahapan, diantaranya pemilihan data mentah, pra pengolahan data (preprocessing), transformasi data, data mining, dan evaluasi hasil dari pengolahan informasi [1] [3]. Maka dapat dikatakan bahwa data mining merupakan bagian dari *Knowledge Discovery in Database* (KDD). Dalam bagian proses data mining perlu menentukan jenis metode yang akan digunakan dengan menyesuaikan karakteristik data dan kebutuhan analisis.

B. Metode Clustering

Salah satu metode yang dikenal di dalam data mining adalah clustering. Pengertian clustering adalah pengelompokan sejumlah data atau objek ke dalam kluster (*group*) agar setiap kluster tersebut berisi data yang semirip mungkin dan berbeda dengan data / objek dalam cluster lainnya. Pengelompokan (*clustering*) merupakan bagian dari metode data mining yang sifatnya tanpa arahan (*unsupervised*) [2]. Pemilihan metode ini berdasarkan kebutuhan analisis dan karakteristik data. Dalam metode ini terdapat beberapa algoritma yang dapat dipilih sesuai kebutuhan analisis.

C. Algoritma K-Means

K-Means Clustering dapat diartikan sebagai berikut, K adalah sebagai Konstanta dari jumlah cluster yang diinginkan, *means* diartikan nilai rata - rata dari suatu group data, sehingga K-Means diartikan sebagai metode analisa data dengan melakukan proses pemodelan tanpa adanya supervisi untuk dijadikan kelompok data atau sistem partisi [3]. Metode K-Means ini mengarahkan untuk mengelompokkan data yang ada kedalam beberapa kelompok, yang mana data dalam satu kelompok memiliki karakteristik sama, dan berbeda karakteristik dengan data di kelompok yang lain. Berikut ini adalah rumus perhitungan Euclidean Distance :

$$D(c, p)_n = \sqrt{\sum_{i=0}^n (p_i - c_i)^2} \quad (1)$$

Berdasarkan rumus, berikut ilustrasi data untuk mendapatkan $D(S_n, C_1)$:

$$\begin{aligned} D(S_1, C_1) &= \sqrt{(S_{1a} - C_{1a})^2 + (S_{1b} - C_{1b})^2 + (S_{1c} - C_{1c})^2 + (S_{1d} - C_{1d})^2} \\ &= \sqrt{(2 - 2)^2 + (1 - 1)^2 + (1 - 1)^2 + (1 - 1)^2} = 0 \\ D(S_2, C_1) &= \sqrt{(S_{2a} - C_{1a})^2 + (S_{2b} - C_{1b})^2 + (S_{2c} - C_{1c})^2 + (S_{2d} - C_{1d})^2} \\ &= \sqrt{(2 - 2)^2 + (1 - 1)^2 + (3 - 1)^2 + (1 - 3)^2} = 0.828 \end{aligned} \quad (2)$$

Dalam rumus Euclidean distance dinotasikan D adalah titik dokumen, p adalah data record, c adalah centroid, n adalah jumlah data dan i adalah iterasi.

D. Dataset

Dataset dalam penelitian ini berisikan profil mahasiswa baru pada universitas Widyagama Malang tahun ajaran 2020/2021 yang diambil dari pihak akademik Universitas Widyagama Malang. Dipilih beberapa atribut penting yang dibutuhkan dalam penelitian ini antara lain nama mahasiswa, Jurusan, Sekolah Asal dan IPK. Table I menyajikan sample 10 dari 100 dataset yang diproses.

Tabel I: Dataset Mahasiswa

No.	Nama	Jurusan	Sekolah asal	IPK
1	Adel	Teknik	SMK	2.25
2	Adji	Ekonomi	SMA	1.8
3	Alya	Hukum	SMA	3.5
4	Anel	Teknik	MAN	2.89
5	Darul	Teknik	SMK	2.45
6	Firadi	Ekonomi	SMK	1.9
7	Irwansyah	Hukum	SMA	3.3
8	Ismail	Teknik	SMA	2
9	Khadja	Ekonomi	SMA	2.44
10	M Alief	Hukum	SMK	3.22

E. Transformasi Data

Dari informasi yang di peroleh, maka dilakukan perubahan (transformasi) data. Transformasi data bertujuan agar data calon mahasiswa baru dapat diolah dengan memakai metode K-Means Clustering. Berikut (Tabel II - IV) adalah pengelompokan tiap data yang ditransformasi.

Tabel II: Inisialisasi Jurusan

Jurusan	Frekuensi	Inisial
Ekonomi	38	1
Hukum	29	2
Teknik	33	3

Tabel III: Inisialisasi Asal Sekolah

Asal sekolah	Frekuensi	Inisial
SMA	54	1
SMK	29	2
MAN	17	3

Tabel IV: Inisialisasi IPK

IPK	Frekuensi	Inisial
SMA	54	1
SMK	29	2
MAN	17	3

Hasil Transformasi data yang dihasilkan dapat dilihat pada Tabel V berikut (sample 10 data dari 100 hasil):

Tabel V: Hasil Transformasi

Id	Jurusan	Asal Sekolah	IPK
1	3	2	2
2	1	1	1
3	2	1	3
4	3	3	2

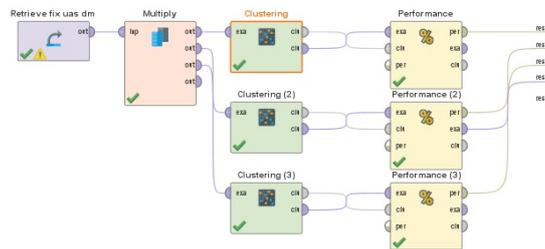
... berlanjut di halaman selanjutnya

Tabel V – ...lanjutan

Id	Jurusan	Asal Sekolah	IPK
5	3	2	2
6	1	2	1
7	2	1	3
8	3	1	2
9	1	1	2
10	2	2	3

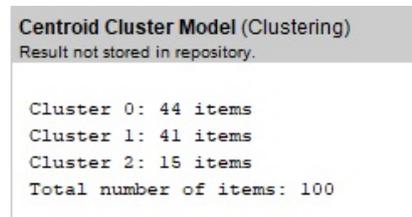
III. HASIL DAN PEMBAHASAN

Hasil pengolahan data calon mahasiswa baru menggunakan metode algoritma *K-Means Clustering* dan bantuan perangkat lunak *RapidMiner* dapat dilihat pada Gambar 1 berikut.



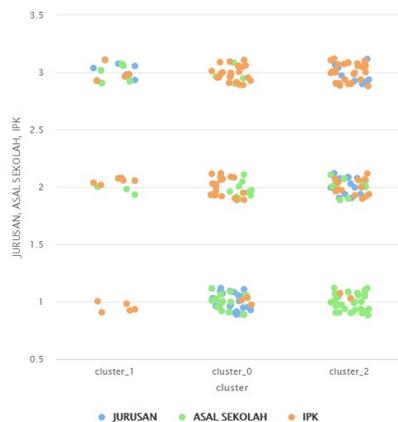
Gambar 1: Pemodelan K-Means di Rapid Miner

Pengolahan data mahasiswa baru menggunakan pemodelan K-Means clustering dengan jumlah sampel data 100 dan jumlah cluster sebanyak 3, dengan mendefinisikan nilai k jumlah cluster_0 : 44 items , jumlah cluster_1: 41 items dan jumlah cluster_2: 15 items, terlihat pada Gambar 2.



Gambar 2: Cluster Model

Berikut (Gambar 3) adalah hasil pemodelan k-means clustering persebaran cluster_0, cluster_1 dan cluster_2 dari sampel data sebanyak 100 items dengan menggunakan RapidMiner.



Gambar 3: Pemodelan Metode K-Means Clustering Menggunakan RapidMiner

Hasil pemodelan pada RapidMiner menunjukkan bahwa data mahasiswa baru menghasilkan tiga kelompok data. Kelompok pertama yaitu jurusan yang ditandai menggunakan warna biru muda, kelompok kedua merupakan kelompok asal sekolah yang ditunjukkan dengan warna hijau muda, dan yang terakhir adalah kelompok ketiga, yaitu kelompok IPK yang ditandai warna jingga. Tabel VI adalah table pengelompokkan berdasarkan kedekatan jarak antara centroid dengan data mahasiswa baru pada setiap atributnya.

Tabel VI: Hasil Analisis Cluster Satu (Cluster_0)

Jurusan	Jumlah	Asal Sekolah dan Jumlah
Cluster_0		
Ekonomi	41	SMA=16, SMK=14, MAN=8
Hukum	3	SMA=0, SMK=3, MAN=0
Teknik	0	SMA=0, SMK=0, MAN=0
Cluster_1		
Ekonomi	0	SMA=0, SMK=0, MAN=0
Hukum	2	SMA=0, SMK=1, MAN=1
Teknik	13	SMA=0, SMK=5, MAN=8
Cluster_1		
Ekonomi	0	SMA=0, SMK=0, MAN=0
Hukum	24	SMA=18, SMK=6, MAN=0
Teknik	20	SMA=17, SMK=3, MAN=0

Dari hasil clustering yang sudah dilakukan, dapat di lihat untuk jurusan ekonomi banyak peminat dari SMA dan SMK, jurusan hukum dari SMA dan SMK, dan jurusan Teknik dari SMK dan MAN. Dari hasil tersebut, pihak Universitas Widyagama Malang dalam melakukan promosi kampus dengan cara pengenalan program studi di masing-masing sekolah, sesuai dengan hasil pengelompokan yang paling banyak diminati berdasarkan cluster yang terbentuk. Dengan data analisa ini, dari pihak penerimaan mahasiswa baru bisa dengan mudahnya memberikan saran terhadap calon mahasiswa baru dalam menentukan jurusan yang sesuai dengan kemampuan dan hasil studi rata-rata ketika sudah menjadi Mahasiswa. Dengan demikian calon mahasiswa yang dibidik oleh institusi juga tidak melenceng jauh dari hasil analisa agar tepat sasaran. Dengan menambahkan item pemberian beasiswa mahasiswa berprestasi untuk promosi peminat jurusan yang sudah ada, maka akan semakin menarik pula jumlah dari calon mahasiswa baru. Dengan demikian, analisa dari data yang sudah diolah ini akan semakin banyak manfaatnya. Selain mahasiswa semakin termotivasi, image institusi juga akan terbentuk dengan sendirinya.

IV. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilaksanakan, dapat disimpulkan beberapa pengetahuan. Berdasarkan hasil cluster pertama (cluster_0) pendaftar didominasi calon mahasiswa yang mengambil jurusan ekonomi (43) yang tersebar cukup merata di ketiga jenis sekolah. Kemudian cluster kedua (cluster_1) didominasi jurusan teknik (13) yaitu tidak sebesar jumlah calon mahasiswa yang mengambil ekonomi pada cluster_0 dan asal sekolah menengah hanya dari SMA dan MAN. Sementara cluster ketiga (cluster_2) didominasi dua jurusan (hukum = 24, teknik = 20) dan asal sekolah menengah dari SMA dan SMK. Dari ketiga hasil cluster nampak dua pengetahuan yang dihasilkan yaitu peminat terhadap belum merata dan asal sekolah pada peminatan masing-masing jurusan juga belum merata. Untuk itu perlu ditekankan promosi kampus terhadap sekolah-sekolah dengan lebih menekankan penjelasan tentang jurusan yang ada secara lebih detil. Secara umum peminat dari jurusan teknik cukup banyak namun hasil data berbanding terbalik untuk itu perlu evaluasi penyebab sedikitnya peminat jurusan teknik. Terakhir adalah dari hasil cluster dihasilkan peminat dari madrasah paling sedikit, untuk itu perlu meningkatkan promosi yang lebih optimal. Sementara dari sisi algoritma, bahwa penentuan titik centroid (titik pusat) pada awal proses sangat berpengaruh pada hasil cluster. Maka pada penelitian yang akan datang dapat dilakukan beberapa ujicoba penentuan cluster yang berbeda-beda kemudian diamati dan dianalisis hasil masing-masing agar dapat menghasilkan analisis cluster yang tepat.

PUSTAKA

- [1] Priyatman, H., Sajid, F., & Haldivany, D. (2019). Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan Mahasiswa. *JEPIN (Jurnal Edukasi Dan Penelitian Informatika)*, 5(1), 62–66.
- [2] Nas, C. (2020). Data Mining Pengelompokan Bidang Keahlian Mahasiswa Menggunakan Algoritma K-Means (Studi Kasus: Universitas Cic Cirebon). In *Syntax: Jurnal Informatika (Vol. 09, Issue 1)*.
- [3] A. E. Wicaksono, "Implementasi Data Mining Dalam Pengelompokan Data Peserta Didik Di Sekolah Untuk Memprediksi Calon Penerima Beasiswa Dengan Menggunakan Algoritma K-Means (Studi Kasus SMAN 16 Bekasi)," *Jurnal Teknologi Rekayasa*, vol. 21 No. 03, pp. 206–216, 2016.
- [4] Lestari, W., Bina, S., & Kendari, B. (2019). Clustering Data Mahasiswa Menggunakan Algoritma K-Means Untuk Menunjang Strategi Promosi (Studi Kasus: STMIK Bina Bangsa Kendari). *SIMKOM*, 4(2), 2715–2906.
- [5] Hartati, T., Nurdianawan, O., Wiyandi, E., & Cirebon, S. I. (2021). ANALISIS DAN PENERAPAN ALGORITMA K-MEANS DALAM STRATEGI PROMOSI KAMPUS AKADEMI MARITIM SUAKA BAHARI. In *Jurnal Sains Teknologi Transportasi Maritim (Vol. 3, Issue 1)*.
- [6] Y. Darmi and A. Setiawan, "Penerapan Metode Clustering K-Means Dalam Pengelompokan Penjualan Produk," *Jurnal Media Infotama*, vol. 12 No. 02, pp. 148–157, 2016, ISSN: 1858-2680.
- [7] Mahmudi, I., Dwi Indriyanti, A., & Lazulfa, I. (n.d.). PENERAPAN ALGORITMA K-MEANS CLUSTERING SEBAGAI STRATEGI PROMOSI PENERIMAAN MAHASISWA BARU PADA UNIVERSITAS HASYIM ASY'ARI JOMBANG.

- [8] E. W. Astuti, “Clustering Program Keahlian Pada Pendaftaran Siswa Baru (PSB) Dengan Menggunakan Algoritma K-Means: Studi Kasus Di SMKN 14 Chairun Nas 1 Nguling,” *Jurnal SPIRIT.*, vol. 7 No. 02, pp. 58–65, 2015, ISSN: 2085- 3092.
- [9] Wulan Sari, R., Wanto, A., & Perdana Windarto, A. (n.d.). KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer) IMPLEMENTASI RAPIDMINER DENGAN METODE K-MEANS (STUDY KASUS: IMUNISASI CAMPAK PADA BALITA BERDASARKAN PROVINSI).
- [10] G. F. Mandias et al., “Penerapan Algoritma K-Means Untuk Analisis Prestasi Akademik Mahasiswa Fakultas Ilmu Komputer Universitas Klabat Application of K-Means Algorithm for Academic Achievement Analysis of Faculty of Computer Science Universitas Klabat,” pp. 230–239, 2017.
- [11] F. Marisa, S. S. S. Ahmad, Z. I. M. Yusof, T. M. Akhriza, R. K. Pandey, and W. Purnomowati, “The Analyze of Relationship between Revenue and Customer Payment Methods in Small Medium Enterprise Based on Clustering K-Means,” in *Journal of Physics: Conference Series*, 2021, no. Wcse, pp. 28–30, doi: 10.1088/1742-6596/1908/1/012021.
- [12] F. Marisa, S. S. S. Ahmad, Z. I. M. Yusof, Fachrudin, and T. M. A. Aziz, “Segmentation model of customer lifetime value in Small and Medium Enterprise (SMEs) using K-Means Clustering and LRFM model,” *Int. J. Integr. Eng.*, vol. 11, no. 3, pp. 169–180, 2019, doi: 10.30880/ijie.2019.11.03.018.