

Analisis Data Mining dengan Metode Algoritma K-Means sebagai penentu pemilihan program Studi Teknik Informatika pada Fakultas Teknik Universitas PGRI Madiun

Fatim Nugrahanti¹, Eka Resty Novieta Sari²

^{1,2}Jurusan Teknik Informatika, Universitas PGRI Madiun

e-mail: fatim@unipma.ac.id, Ekaresty@unipma.ac.id

Intisari

Pemilihan program Studi dalam suatu lembaga adalah hal yang sangat penting untuk memastikan prodi yang akan dipilih, para pendaftar mahasiswa baru. Universitas PGRI Madiun mempunyai beberapa program studi yang salah satunya prodi teknik Informatika pada Fakultas Teknik. Dalam pemilihan program studi juga harus memiliki kepeminatan yang harus dipikirkan. Tujuan pemilihan program studi teknik Informatika dalam memilih program studi sangatlah penting untuk menunjang pembelajaran selanjutnya. Dalam pemilihan program studi sangat penting dalam menunjang pembelajaran pada perguruan tinggi. Selain itu mahasiswa lebih semangat mengembangkan ilmu dibidang yang telah mahasiswa pilih. Oleh karena itu setiap mahasiswa yang mendaftar sebagai mahasiswa baru dikelompokkan berdasarkan program studi yang telah dipilih. Dalam analisa data Mining pemilihan program studi menggunakan metode K-Means dimana metode K-Means ini merupakan metode yang sangat populer dalam clustering, Metode Algoritma K-Means yang merupakan salah satu metode terbaik dan paling populer dalam algoritma *clustering* dimana K-Means mencari partisi yang optimal dari data dengan meminimalkan kriteria jumlah kesalahan kuadrat dengan prosedur iterasi yang optimal. Hasil dari analisis penelitian dengan metode Algoritma K-Means ini akan mengelompokkan sesuai dengan prodi yang dipilih mahasiswa tersebut.

Kata kunci— Analisis ,Data Mining, Metode Algorithma K-Means

Abstract

The selection of study programs in an institution is very important to ensure the study program to be chosen, new student registrants. PGRI Madiun University has several study programs, one of which is the Informatics Engineering Study Program at the Faculty of Engineering. In choosing a study program, you must also have interests that must be considered. The purpose of selecting the Informatics engineering study program in choosing a study program is very important to support further learning. In choosing a study program, it is very important to support learning in higher education. In addition, students are more enthusiastic about developing knowledge in the field they have chosen. Therefore, every student who registers as a new student is grouped based on the chosen study program. In data mining analysis, the selection of study programs uses the K-Means method where the K-Means method is a very popular method in clustering, the K-Means algorithm method is one of the best and most popular methods in the clustering algorithm where K-Means looks for the partition that optimally from the data by minimizing the criterion of the number of squared errors with the optimal iteration procedure. The results of the research analysis using the K-Means Algorithm method will group according to the study program chosen by the student.

Keywords— *Analysis, Data Mining, K-Means . Algorithm Method*

PENDAHULUAN

Pendidikan merupakan hal sangat diperlukan saat ini mengingat jenjang pendidikan sangat diperlukan bagi setiap manusia. Dalam usahanya untuk meningkatkan kualitas pendidikan Universitas PGRI Madiun yang mempunyai banyak program studi tentunya penerapan teknologi informasi dalam dunia pendidikan juga dapat menghasilkan data yang berlimpah mengenai mahasiswa dan proses pembelajaran yang dihasilkan. Pada institusi pendidikan perguruan tinggi, data dapat diperoleh berdasarkan data mahasiswa yang mendaftar, sehingga data akan bertambah secara terus menerus, misalnya data mahasiswa baru tersebut.

Berdasarkan pemilihan program studi mahasiswa, akan mendapatkan informasi yang sesuai dengan pilihannya yang dapat diketahui pengolahan terhadap data mahasiswa baru sehingga berguna bagi pihak universitas dalam mengelompokkan sesuai program studi. Pengolahan data mahasiswa perlu dilakukan untuk mengetahui informasi penting berupa.([1] Fitri Yunita, 2018)

Dalam pemilihan program studi di Universitas PGRI Madiun sangat diperlukan pengelompokan pada masing-masing program studi. Data Mining adalah data yang memberikan informasi yang sangat dibutuhkan. K-means merupakan salah satu algoritma clustering. Tujuan algoritma ini yaitu untuk membagi data menjadi beberapa kelompok. Algoritma ini menerima masukan berupa data tanpa label kelas. Metode *K-Means* digunakan untuk mengelompokkan instruktur berdasarkan materi diklat. Untuk menetapkan satu nama instruktur.

Dalam penelitian sebelumnya pengelompokan data berdasarkan nilai, dapat menghasilkan keputusan yang berminat untuk memilih jurusan yang diinginkan. Penerapan Metode clustering merupakan salah satu langkah untuk menyelesaikan masalah tentang pengelompokan data siswa. Sistem yang pengelompokan ini dapat mempermudah yang berminat sesuai jurusan masing-masing yang dipilih. Sehingga lebih efisien daripada sebelumnya ([2] Eferoni Ndruru, Riswan Limbong, 2018).

Dari penelitian yang dilakukan oleh [3](Agus Perdana W, 2017) yang berjudul Penerapan Data Mining Pada Ekspor Buah-Buahan Menurut Negara Tujuan Menggunakan *K-Means* Clustering. Dalam melakukan clustering data yang diperoleh akan dihitung terlebih dahulu berdasarkan ekspor buah buahan berdasarkan negara tujuan, dan hasil penjumlahan menghasilkan 2 kriteria penilaian . Dari hasil penelitian dapat digunakan jumlah ekspor buah-buahan menurut negara tujuan.

Penelitian (Totok Suprawoto, 2016)[4] menyatakan dalam jurnalnya yang berjudul ‘ Klasifikasi data mahasiswa menggunakan metode k-means untuk menunjang pemilihan strategi pemasaran’ bahwa hasil klusterisasi dibagi 3 kelompok data yaitu Mahasiswa rendah dan jarang ke perpustakaan, dengan pusat cluster (1.48, 1.4), mahasiswa dengan ipk tinggi dan cukup sering ke perpustakaan , dan juga mahasiswa, mahasiswa yang ipk tinggi dan sering berkunjung ke

perpustakaan.

TINJAUAN PUSTAKA

DATA MINING

Dalam data Mining merupakan yang suatu hal yang tidak sulit dan baru. Dari definisi *Data Mining* adalah kenyataan bahwa *Data Mining* mewarisi banyak aspek dan teknik dari bidang-bidang ilmu yang sudah mapan terlebih dahulu. Berawal dari beberapa disiplin ilmu, *Data Mining* bertujuan untuk memperbaiki teknik tradisional sehingga bisa menangani:

- a. Jumlah data yang sangat besar
- b. Dimensi data yang tinggi
- c. Data yang heterogen dan berbeda sifat

Pengelompokan *Data Mining* dibagi menjadi beberapa kelompok yaitu :

- a. Deskripsi

Deskripsi merupakan cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data yang dimiliki.

- b. Estimasi

Estimasi dapat juga dikatakan perkiraan atau klasifikasi, kecuali *variable* target estimasi lebih ke arah numerik daripada ke arah kategori. Demikian juga model yang dibangun merupakan record yang lengkap sehingga dalam data mining menyediakan nilai variabel target yang merupakan sebagai nilai prediksi.

- c. Prediksi

Merupakan memperkirakan nilai yang belum diketahui untuk nilai yang akan datang.

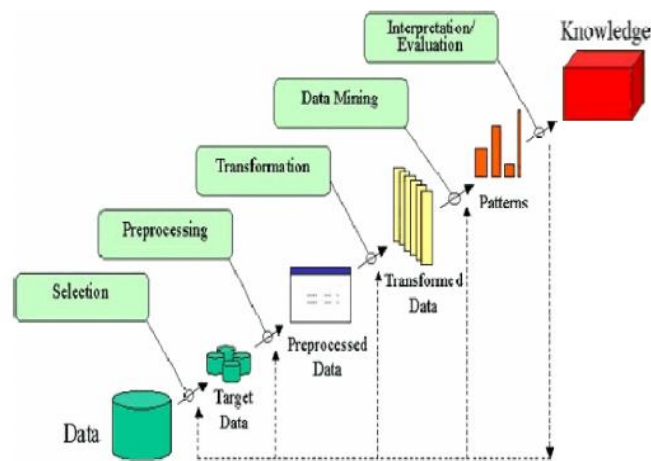
- d. Klasifikasi

Penggolongan yang terdiri dari tiga kategori yaitu tinggi, sedang dan rendah, merupakan klasifikasi terdapat target *variable*.

- e. Pengkalsteran

Merupakan pengelompokan *record*, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan.

Data Mining bukanlah sesuatu yang baru, bahwa Data Mining mewarisi bidang bidang teknik dan disiplin ilmu sehingga sulit untuk mendefinisikan karena terlebih dahulu lebih mapan. Tujuan dari Data Mining adalah ilmu yang memperbaiki teknik tradisional yang biasa menangani jumlah data yang besar, data yang mempunyai dimensi tinggi dan data yang berbeda sifat dan heterogen. (Alfanisa Annurulah Fajrin, Algifanri Maulana, 2018)[5]



Gambar.1, Aliran Informasi dalam Data Mining (Fina Nasari1), Surya Darma, 2015)[6]

Algoritma K-Means

Penggunaan Algoritma K-Means kita dapat mengelompokkan data secara maksimal dalam kemiripan data dalam satu cluster dan meminimalkan data dalam antar cluster. Dengan algoritma ini mempunyai keefektifan dalam mengolah data mining dalam pengelompokan (clustering) data yang akan dianalisa. Dalam jurnal Metode K-Means berusaha mengelompokkan data yang ada kedalam satu kelompok, dimana data dalam satu kelompok mempunyai karakteristik yang berbeda dengan data yang ada didalam kelompok yang lain. Dasar algoritma k-measn adalah sebagai berikut :

1. Tentukan nilai k sebagi jumlah klaster yang ingin dibentuk.
2. Inisialisasi k sebagai centroid yang dapat dibangkitkan secara random.

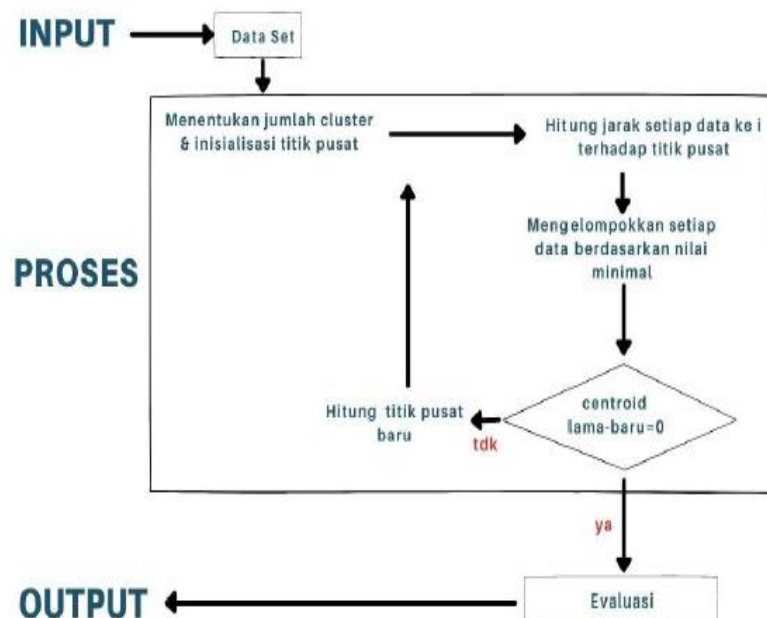
3. Hitung jarak setiap data ke masing-masing centroid menggunakan persamaan Eulidiean Distance yaitu sebagai berikut :

$$d(x_j, c_j) = \sqrt{\sum_{j=1}^n (x_j - c_j)^2}$$

.....(1)

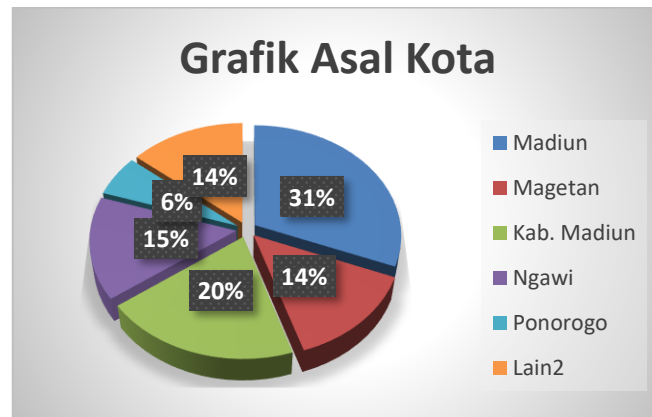
METODE PENELITIAN

Dalam penelitian ini menggunakan metode penelitian menggunakan diagram alir yang digambarkan dibawah ini :



Gambar 2. Diagram Alir Penelitian

Adapun data yang penulis gunakan , data dari bagian Penerimaan mahasiswa baru Universitas PGRI Madiun . Sumber data diambil langsung dari data mahasiswa yang mendaftar pada tahun 2020 yang sekarang ini pada semester 4. Dan dibawah akan disampaikan grafik asal mahasiswa pendaftar :



Gambar 3. Grafik Prosentase Sekolah Asal

HASIL DAN PEMBAHASAN

Sumber Data

Penelitian ini bertujuan untuk mengelompokkan mahasiswa baru yang mendaftar pada Universitas PGRI Madiun tahun 2019/2020 khususnya Fakultas Teknik. Dalam tahap pengumpulan data melalui mahasiswa yang mendaftar yang bersumber pada bagian Penerimaan Mahasiswa Baru (PMB). Data mahasiswa yang mendaftar di prodi prodi Fakultas Teknik yang terdiri dari Prodi Teknik Informatika, Sistem Informasi, Teknik Elektro, Teknik Kimia dan Teknik Industri, dan dari sekolah asal yaitu SMU, SMAK, dan sekolah menengah atas lainnya. Metode K –Mean dipergunakan untuk mengelompokkan data dalam beberapa kelompok.

Tabel 1. Data Pendaftar Mahasiswa Baru

No	Asal Kota	Program Studi	Sekolah Asal
1	Madiun	Teknik Informatika	SMK Negeri 3 KIMIA Madiun
2	Ngawi	Teknik Informatika	SMK PGRI 1 NGAWI
3	Magetan	Teknik Informatika	SMAN 1 SUKOMORO
4	Madiun	Teknik Informatika	SMKN 1 MADIUN

Analisis Data Mining dengan Metode Algoritma K-Means sebagai penentu pemilihan program Studi Teknik Informatika pada Fakultas Teknik Universitas PGRI Madiun

(Fatim Nugrahanti.Eka Resty Novieta Sari)

5	Kab. Madiun	Teknik Informatika	SMAN 1 KAWEDANAN
6	Kab. Madiun	Teknik Informatika	SMAN 1 DAGANGAN
7	Ngawi	Teknik Informatika	SMK NEGERI 1 NGAWI
8	Mojokerto	Teknik Informatika	SMK MUHAMMADIYAH 1 NGORO
9	Madiun	Teknik Informatika	SMAN 6 MADIUN
10	Kab. Madiun	Teknik Informatika	SMKN 1 MEJAYAN
11	Kab. Madiun	Teknik Informatika	SMKN 1 KEBONSARI
32			
100	Madiun	T.Informatika	SMK CENDEKIA MADIUN

Transformasi Data

Pada tahap ini penulis mengelompokkan sehingga dapat dianalisa dengan menggunakan K-Means. Dan sebagai variabel untuk kota asal dikelompokkan menjadi 6 kelompok, yaitu Madiun, Kab. Madiun, Magetan, Ngawi, Ponorogo dan kota-kota diluar dari ke 6 kota tersebut. Dan untuk kota Madiun ditransformasikan menjadi 1, Kabupaten Madiun ditransformasikan menjadi 2, Magetan ditransformasikan menjadi 3, Ngawi ditransformasikan dengan 4 dan seterusnya wilayah Ponorogo 5, sampai dengan wilayah lainnya diluar Kota Madiun yaitu 6. Variabel Prodi dikelompokkan menjadi 5 kelompok. Kelompok urutan 1 prodi teknik Informatika, 2 untuk prodi Sistem Informasi, 3 prodi Teknik Industri, 4, prodi Teknik Elektro, dan 5 prodi Teknik Kimia menjadi 5, untuk pengelompokkan variabel asal sekolah dikelompokkan menjadi SMA ditransformasikan menjadi 1, transformasi untuk SMK 2, dan untuk sekolah selain SMA, SMK ditransformasikan menjadi 3. Dan setelah semua ditransformasikan seperti tabel dibawah ini :

Tabel 2. Transformasi Data

ID	Asal sekolah	Program Studi	Sekolah asal
1	1	1	2

2	4	1	2
3	3	1	1
4	1	1	1
5	2	1	1
6	2	1	1
7	4	1	2
8	6	1	2
9	1	1	1
10	2	1	2
.....			
100	1	1	2

Peng-clusteran dengan menggunakan *Algoritma K-Means*

Untuk proses selanjutnya setelah dilakukan transformasi kemudian pengolahan data selanjutnya menggunakan *Algoritma K-Means*. Dalam pengolahan data dibentuk cluster baru yang dapat dibuat dari data yang telah ditransformasikan tersebut diatas. Pada penelitian ini didapat 3 *cluster*. Kemudian langkah berikutnya jika sudah mendapatkan 3 *cluster* akan ditentukan 3 titik pusat dari *cluster* tersebut. Tabel dibawah menunjukkan titik pusat dari 3 *cluster* yang disebut *centroid*.

Tabel 3. .Titik Pusat awal *cluster*

Centroid 1	4	1	2
Cebtroid 2	3	2	1
Centroid 3	1	1	2

Langkah berikutnya menentukan titik pusat menghitung jarak objek ke *centroid*, pada penelitian ini aplikasi yang digunakan untuk membantu perhitungan persamaan dengan menggunakan *excel* perhitungan untuk Iterasi pertama. Dari hasil penelitian dinyatakan dalam jurnalnya yang berjudul ‘Analisis dan implementasi data mining dalam menganalisa Kendala akademik yang sering dikeluhkan Mahasiswa amik labuhan batu (Gomal Yuni Yanris ,2016) [7]

Tabel . 4 Hasil Iterasi

Jarak Data Ke Centroid					
	M1	M2	M3	Cluster	Jarak Terdekat
	3,00	2,45	0,00	3	0,00
	0,00	1,73	3,00	1	0,00
	1,41	1,00	2,24	2	1,00
	3,00	2,45	0,00	3	0,00
	2,24	1,41	1,41	2	1,41
	2,24	1,41	1,41	2	1,41
	0,00	1,73	3,00	1	0,00
	2,00	3,32	5,00	1	2,00
	3,16	2,24	1,00	3	1,00
	2,00	1,73	1,00	3	1,00
	2,00	1,73	1,00	3	1,00
	1,00	1,41	3,16	1	1,00
	1,00	2,45	3,16	1	1,00
					
	2,24	3,16	5,10	1	2,24
	3,00	2,45	0,00	3	0,00
	3,00	2,45	0,00	3	0,00
Total	250,57	218,88	260,39	205	146,35
Rata-rata	2,15	2,19	2,60	2,05	1,46

Tabel diatas merupakan hasil Iterasi pertama dari jarak terdekat dari Centroid sehingga dapat diketahui data data yang saling berhubungan antara cluster .

Hasil perhitungan setiap data ke setiap cluster Iterasi 1:

Asal Kota	Jumlah	Program Studi	Jumlah	Sekolah asal	Jumlah
1. Ngawi	33,33%	1. Teknik	53,33%	1. SMA	43,33%
2. Ponorogo	20,00%	Inf	10,00%	2. SMK	43,33%
3. lain-lain	46,67%	2. SI	33,33%	3. MA	13,33%
		3. TI			

Hasil perhitungan setiap data ke setiap cluster Iterasi 2 :

Asal Kota	Jumlah	Program Studi	Jumlah	Sekolah asal	Jumlah
1. Madiun	16,67%	T. Inf	13,33%	SMA	60,00%
2.Kab. Madiun	26,67%	SI	13,33%	SMK	26,67%
3. Magetan	40,00%	TI	40,00%	MA	10,00%
4. Ngawi	16,67%	T. Elektro	20,00%		

Hasil perhitungan setiap data ke setiap cluster Iterasi 3 :

Asal Kota	Jumlah	Program Studi	Jumlah	Sekolah asal	Jumlah
1. Madiun	86,67%	T. Informatika	116,67%	SMA	50,00%
2. Kab. Madiun	40,00%	SI	10,00%	SMK	10,00%
3. Magetan	6,67%	TI	6,67%	MA	6,67%

Setelah melalui penganalisaan pada perhitungan Iterasi pertama maka didapat bahwa pemilihan program studi pada fakultas Teknik para calon mahasiswa baru kebanyakan memilih prodi Informatika dan berasal dari Sekolah Menengah Atas sehingga dengan adanya hasil pengclusteran Teknik Informatika menduduki posisi paling atas . Berdasarkan tabel diatas maka untuk mengambil keputusan dalam peningkatan mutu untuk prodi Informatika dapat dipertahankan. Prodi yang banyak diminati adalah Prodi Teknik Informatika.

KESIMPULAN

Berdasarkan penelitian diatas penulis menyimpulkan sebagai berikut :

1. Pengelompokan berdasarkan pemilihan program studi menghasilkan mahasiswa baru untuk memilih jurusan yang diinginkan.
2. Penerapan Metode clustering merupakan salah satu langkah untuk menyelesaikan masalah tentang pengelompokan data siswa.
3. Sistem yang pengelompokan ini dapat mempermudah lembaga untuk mengambil keputusan dalam meningkatkan prestasi mahasiswa khususnya prodi Informatika di wilayah Madiun.

DAFTAR PUSTAKA

- [1] F. Yunita, “Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru (Studi Kasus : Universitas Islam Indragiri),” *J. Sist.*, vol. 7, no. September, pp. 238–249, 2018.
- [2] E. Ndruru and R. Limbong, “Implementasi Data Mining Dalam Pengelompokan Jurusan yang Diminati Siswa SMK Negeri 1 Lolowa’u menggunakan Metode Clustering | Ndruru | MEANS (Media Informasi Analisa dan Sistem),” *MEANS (Media Inf. Anal. dan Sist.*, vol. 3, no. 2, pp. 107–113, 2018, [Online]. Available: http://ejournal.ust.ac.id/index.php/Jurnal_Means/article/view/273/pdfdsdx11.
- [3] A. P. Windarto, “Penerapan Datamining Pada Ekspor Buah-Buahan Menurut Negara Tujuan Menggunakan K-Means Clustering Method,” *Techno.Com*, vol. 16, no. 4, pp. 348–357, 2017, doi: 10.33633/tc.v16i4.1447.
- [4] T. Suprawoto, “Klasifikasi Data Mahasiswa Menggunakan Metode K-Means Untuk Menunjang Pemilihan Strategi Pemasaran,” *JIKO (Jurnal Inform. dan Komputer)*, vol. 1, no. 1, pp. 12–18, 2016, doi: 10.26798/jiko.2016.v1i1.9.
- [5] A. Maulana and A. A. Fajrin, “Penerapan Data Mining Untuk Analisis Pola Pembelian Konsumen Dengan Algoritma Fp-Growth Pada Data Transaksi Penjualan Spare Part Motor,” *Klik - Kumpul. J. Ilmu Komput.*, vol. 5, no. 1, p. 27, 2018, doi: 10.20527/klik.v5i1.100.
- [6] F. Nasari and S. Darma, “Seminar Nasional Teknologi Informasi dan Multimedia 2015 PENERAPAN K-MEANS CLUSTERING PADA DATA PENERIMAAN MAHASISWA BARU (STUDI KASUS : UNIVERSITAS POTENSI UTAMA),” pp. 6–8, 2015.
- [7] G. J. Yanris, “Analisis Dan Implementasi Data Mining Dalam Menganalisa Kendala Akademik Yang Sering Dikeluhkan Mahasiswa Amik Labuhanbatu (Studi Kasus : Amik Labuhanbatu),” *J. Inform.*, vol. 4, no. 1, pp. 15–24, 2019, doi: 10.36987/informatika.v4i1.226