

Analisis Data Mahasiswa Menggunakan Algoritma K-Means Clustering Untuk Menunjang Strategi Promosi Pada Program Studi Manajemen Informatika Politeknik Negeri Sambas

Helen Septriani^{1*}, Fathushahib², Theresia Widji Astuti³, Erifa Syahnaz⁴

Manajemen Informatika, Jurusan Manajemen Informatika, Politeknik Negeri Sambas, Sambas, 79400, Indonesia. E-mail: p3m.politekniksambas@gmail.com

Abstract. *Data from Sambas State Polytechnic Informatics Management Study Program student every year produces abundant data, resulting in a buildup of data. Abundant data requires data processing to explore the information contained in the data. The aim of this research is to cluster student data in the Sambas State Polytechnic Informatics Management Study Program through a data mining process using the K-Means algoritma to form clusters. The data attributes that will be used are student data from the 2020-2023 class with 219 data items and the data source comes from the staff of the Sambas State Polytechnic Informatics Management Study Program.*

Keywords: *Data Mining, Algorithm, K-Means, Clustering, Promotion Strategy.*

Abstrak. Data mahasiswa Program Studi Manajemen Informatika Politeknik Negeri Sambas setiap tahunnya dan menghasilkan data yang berlimpah sehingga terjadi penumpukan data. Data yang berlimpah perlu dilakukan pengolahan data untuk menggali informasi yang terdapat didalam data tersebut. Tujuan penelitian ini untuk mengcluster data mahasiswa di Program Studi Manajemen Informatika Politeknik Negeri Sambas melalui proses data mining dengan menggunakan algoritma K-Means untuk pembentukan cluster. Atribut data yang akan digunakan adalah nama mahasiswa, alamat, jenis kelamin, jalur pendaftaran, dan kecamatan. Data yang digunakan adalah data mahasiswa angkatan 2020-2023 dengan data 219 items dan sumber data berasal dari bagian staff Prodi Manajemen Informatika Politeknik Negeri Sambas.

Kata Kunci: Data Mining, Algoritma, K-Means, Clustering, Strategi Promosi.

PENDAHULUAN

1. Latar Belakang

Politeknik Negeri Sambas merupakan salah satu perguruan tinggi negeri yang ada di Kalimantan Barat dan satu-satunya Politeknik Negeri Sambas di Kabupaten Sambas. Politeknik Negeri Sambas merupakan politeknik tingkat negeri yang lebih banyak menerapkan muatan pembelajaran praktikum dibandingkan muatan teori.

Penerimaan mahasiswa baru yang dilakukan setiap tahun menghasilkan data yang berlimpah berupa profil mahasiswa baru. Berdasarkan berlimpahnya data mahasiswa, terdapat informasi yang dapat di ketahui dengan cara melakukan pengolahan data. Pengolahan data mahasiswa perlu dilakukan untuk mengetahui informasi penting berupa pengetahuan baru. Informasi tersebut dapat digunakan sebagai penunjang dalam pengambilan keputusan bagi kampus.

Pengolahan data berukuran besar yang mempunyai jumlah record dan jumlah atribut yang cukup banyak tidak dapat dilakukan dengan mudah. Salah satu cara yang dapat diterapkan untuk mengolah data yang berukuran besar adalah dengan menerapkan penggunaan data mining. Teknologi data mining merupakan salah satu alat bantu untuk penggalian data pada basis data berukuran besar dengan spesifikasi kerumitan tinggi. Karena di dalam data mining informasi yang luas, dan dari informasi yang kita dapat bisa kita gunakan sebagai suatu keputusan atau menentukan sebuah kualitas dalam menentukan suatu keputusan. Adapun kumpulan dari sebuah data atau

informasi memiliki banyak potensi untuk dijadikan suatu informasi yang terdapat di dalam sebuah data.

Tujuan pengolahan data mahasiswa Program Studi Manajemen Informatika Politeknik Negeri Sambas yaitu menerapkan algoritma *K-Means Clustering* pada data penerimaan mahasiswa baru di program Studi Manajemen Informatika. Atribut data yang akan digunakan adalah nama mahasiswa, alamat, jenis kelamin, jalur pendaftaran, sekolah asal, dan kecamatannya. Data yang digunakan adalah data mahasiswa angkatan 2020-2023 dan sumber data berasal dari bagian akademik Program Studi Manajemen Informatika di Politeknik Negeri Sambas. Pengolahan data menggunakan teknik data mining dengan algoritma *K-Means clustering*.

1.1 Rumusan Masalah

Berdasarkan latar belakang masalah di atas, dapat dirumuskan sebuah permasalahan yaitu:

- a. Bagaimana karakteristik demografis pada Progra Studi Manajemen Informatika di Politeknik Negeri Sambas?
- b. Berapa jumlah cluster yang optimal yang dapat dihasilkan dari analisis data mahasiswa menggunakan algoritma K-Means clustering?
- c. Bagaimana hasil analisis data mahasiswa tersebut dapat digunakan untuk merancang strategi promosi yang lebih terarah dan relevan?

1.2 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dipaparkan, maka tujuan dari penelitian ini adalah sebagai berikut:

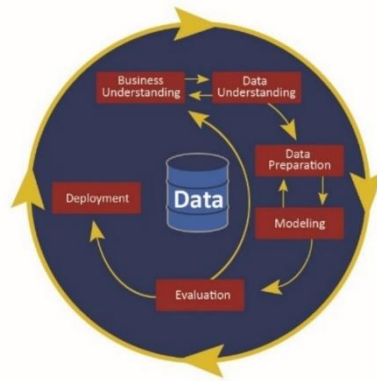
- a. Data mahasiswa bervariasi tergantung pada kampusnya. Disini yang sering diidentifikasi dalam analisis mahasiswa yaitu karakteristik demografis berupa kecamatannya.
- b. Untuk penentuan jumlah cluster yang optimal dalam analisis data mahasiswa menggunakan algoritma K-Means Clustering dapat dilakukan dengan berbagai metode, namun tidak ada pendekatan yang tunggal atau baku untuk menentukannya. Disini yang digunakan untuk menentukan jumlah cluster optimal menggunakan cluster Elbow Method. Metode ini melibatkan plot nilai inertia (within-cluster sum of squares) terhadap jumlah cluster yang berbeda. Pada grafik, titik dimana penurunan inertia mulai melambat (menyerupai siku) dapat dianggap sebagai cluster yang optimal.
- c. Jumlah cluster terpilih, hasil analisis tersebut dapat digunakan untuk merancang strategi promosi dengan cara menggunakan hasil clustering untuk membagi mahasiswa ke dalam kelompok-kelompok homogen berdasarkan karakteristik demografis.

METODE PENELITIAN

Jenis metode penelitian yang digunakan yaitu metode *mixed-method* metode ini adalah pendekatan penelitian yang menggabungkan metode kualitatif dan kuantitatif dalam satu studi penelitian. Pendekatan ini memungkinkan peneliti untuk memperoleh pemahaman yang lebih mendalam

dan komprehensif tentang suatu fenomena dengan memanfaatkan kekuatan dari kedua jenis metode tersebut.

Metode Penelitian yang digunakan dalam analisis ini adalah menggunakan metode CRISP-DM (*Cross Industry Standard Process for Data Mining*). Langkah-langkah pada analisis yang akan dilakukan dapat dilihat pada Gambar 1.



Gambar 1. Metode CRISP-DM

Data Mining memiliki 6 fase *CRISP-DM* (*Cross Industry Standard Process for Data Mining*).

- a. Pemahaman Bisnis (*Business Understanding*)
Proses memahami konteks, tujuan, operasi, dan tantangan yang dihadapi oleh sebuah bisnis atau organisasi.
- b. Pemahaman Data (*Data Understanding*)
Langkah penting dalam proyek data science dan analisis data yang bertujuan untuk mengenali, mengidentifikasi, dan memahami data yang akan digunakan.
- c. Pengolahan Data (*Data Preparation*)
Langkah penting dalam proses analisis data yang melibatkan pembersihan, transformasi, dan pengorganisasian data mentah menjadi bentuk yang sesuai untuk analisis lebih lanjut. Tahap ini memastikan bahwa data yang akan dianalisis bebas dari kesalahan, konsisten, dan siap digunakan dalam berbagai metode analisis.
- d. Pemodelan (*Modelling*)
Proses pembuatan model matematis, statistik, atau komputasional untuk merepresentasikan suatu sistem, proses, atau fenomena nyata. Dalam konteks data science dan machine learning, pemodelan melibatkan pembangunan algoritma atau model yang dapat memprediksi atau mengklarifikasikan data berdasarkan pola yang telah dikenali dari data historis.
- e. Evaluasi (*Evaluation*)
Langkah penting dalam proses analisis data dan pengembangan model yang bertujuan untuk menilai kinerja dan efektivitas model atau metode yang digunakan.
- f. Penyebaran (*Deployment*)

Tahap akhir dalam siklus pengembangan model data science dan machine learning, dimana model yang telah dilatih dan diuji ditempatkan ke dalam lingkungan produksi sehingga dapat digunakan untuk membuat prediksi atau keputusan dalam aplikasi dunia nyata.

HASIL DAN PEMBAHASAN

Data yang digunakan pada analisis ini adalah data mahasiswa Program Studi Manajemen Informatika Politeknik Negeri Sambas yang terdiri dari 33 atribut dan 219 record. Data awal disajikan pada Tabel 1 dibawah ini.

Tabel 1. Data Mahasiswa Tahun 2020-2023 Prodi Manajemen Informatika

	No	NIM	Nama	Tempat Lahir	Tanggal Lahir	Jalur Pendaftaran	Status	Jumlah Saudara	Jenis Kelamin	Alamat
0	1	3202002001	CLARISA FEBRIANTI	TEBAS	Sabtu, 16 Februari 2002	SNMPN	Single	4	P	Dusun sutera RT 01 Rw 01 Desa makrampai
1	2	3202002002	FAISAL ARYANSAH	Kab. Sambas	Selasa, 26 Februari 2002	SNMPN	Single	4	L	PARIT RAJA
2	3	3202002003	RENGGA ARIANDORA	SENDOYAN	Sabtu, 29 Desember 2001	SNMPN	Single	4	L	Sendoyan bungor
3	4	3202002004	TRIANA LARASATI	SEMPALAI	Minggu, 08 September 2002	SNMPN	Single	3	P	Sepuk tanjung
4	5	3202002005	DENDY GUNAWAN	Kab. Sambas	Sabtu, 17 November 2001	SNMPN	Single	5	L	Desa kartiasa dusun bindang rt01 rw01
...	...	...	...	...	...	...	...	...	...	...
214	215	3202202092	ACHMAD ROMADON	TP PERAPAKAN	2004-11-03 00:00:00	SELEKSI MANDIRI	NaN	2	L	Dusun Cempaka Rt 001/ Rw 005 Desa Mekar Sekuntum
215	216	3202202093	HADAD ALWI	BANGANG	2004-09-11 00:00:00	SELEKSI MANDIRI	NaN	3	L	Dusun Bangang Rt 001/ Rw 001 Desa Kubangga
216	217	3202202094	ALFIN ALDENDY	SAMBAS	2002-04-05 00:00:00	SELEKSI MANDIRI	NaN	2	L	Dusun Serdang Rt 001/ rw 001 Desa Perapakan
217	218	3202202095	IBADURRAHMAN	TEBAS	2004-09-08 00:00:00	SELEKSI MANDIRI	NaN	3	L	Dusun Asam Kanis Rt 019/ Rw 010 Desa Tebas Kuala
218	219	3202202096	AINUL RUBIAH	PEMANGKAT	28 OKTOBER 2003	SELEKSI MANDIRI	NaN	4	P	Jl. Moh. Hambal Gang Ilham no 5

Data yang sudah dikumpulkan kemudian dilakukan pengolahan data awal untuk mendapatkan data yang lebih akurat. Data hasil pengolahan data awal ditunjukkan pada Tabel 2.

Tabel 2. Hasil Pengolahan Data Awal

No	Nama	Tempat Lahir	Jalur Pendaftaran	Jenis Kelamin	Alamat	Kecamatan	Agama	Pendidikan Terakhir	
0	1	CLARISA FEBRIANTI	TEBAS	SNMPN	P	Dusun sutera RT 01 Rw 01 Desa makrampai	TEBAS	Islam	SMA
1	2	FAISAL ARYANSAH	Kab. Sambas	SNMPN	L	PARIT RAJA	SEJANGKUNG	Islam	SMA
2	3	RENGGAARIANDORA	SENDOYAN	SNMPN	L	Sendoyan bungor	SEJANGKUNG	Islam	SMA
3	4	TRIANA LARASATI	SEMPALAI	SNMPN	P	Sepuk tanjung	SEBAWI	Islam	SMA
4	5	DENDY GUNAWAN	Kab. Sambas	SNMPN	L	Desa kartiasa dusun bindang rt01 rw01	SAMBAS	Islam	SMA
...	...	...	...	...	...	...	...	...	...
214	215	ACHMAD ROMADON	TP PERAPAKAN	SELEKSI MANDIRI	L	Dusun Cempaka Rt 010/ Rw 005 Desa Mekar Sekuntum	TEBAS	0	SMK
215	216	HADAD ALWI	BANGANG	SELEKSI MANDIRI	L	Dusun Bangang Rt 001/ Rw 001 Desa Kubangga	TELUK KERAMAT	0	MA
216	217	ALFIN ALDENDY	SAMBAS	SELEKSI MANDIRI	L	Dusun Serdang Rt 001/ rw 001 Desa Perapakan	PEMANGKAT	0	SMK
217	218	IBADURRAHMAN	TEBAS	SELEKSI MANDIRI	L	Dusun Asam Kanis Rt 019/ Rw 010 Desa Tebas Kuala	TEBAS	0	MA
218	219	AINUL RUBIAH	PEMANGKAT	SELEKSI MANDIRI	P	Jl. Moh. Hambal Gang Ilham no 5 R001/ Rw 004 ...	PEMANGKAT	0	SMA

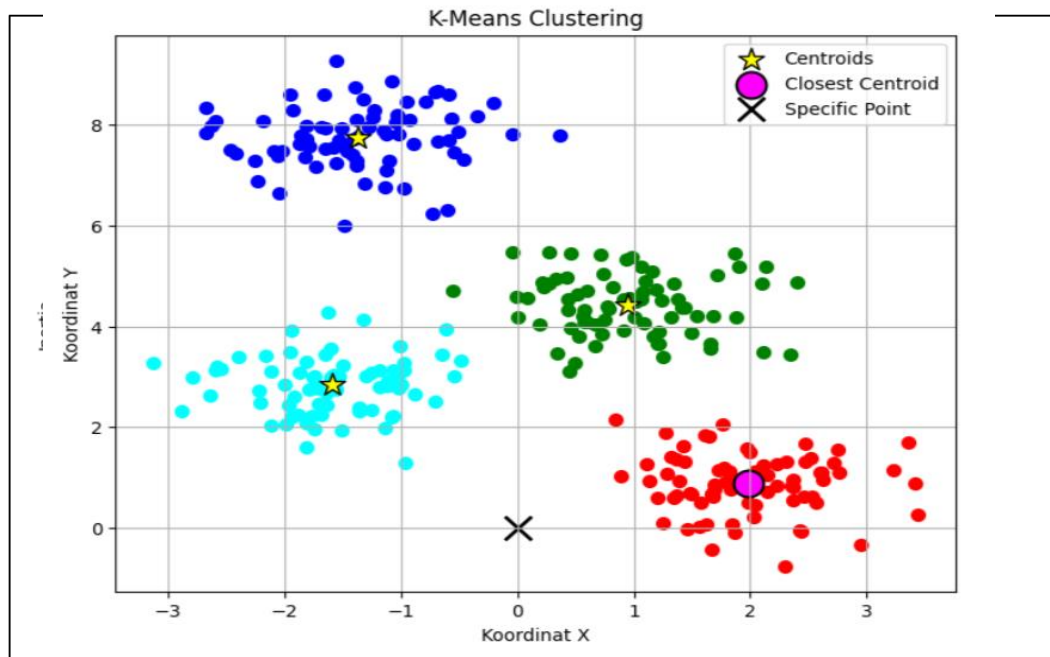
219 rows x 9 columns

Selanjutnya dilakukan proses clustering menggunakan data kecamatan, jumlah mahasiswa perkecamatan, latitude, dan longitude di setiap kecamatannya.

Tabel 3. Data Untuk proses Clustering

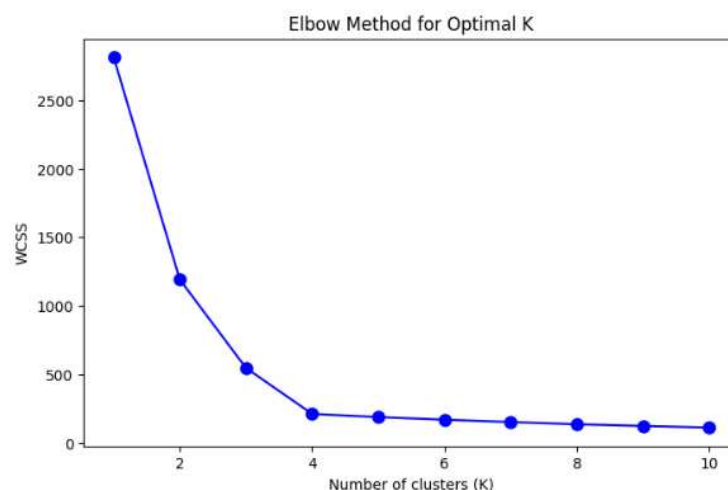
Kecamatan	Jumlah Data	Latitude	Longitude
Galing	3	1.5997404	109.354527
Jawai	6	1.3916063	109.0809868
Jawai Selatan	1	1.2369493	109.0215438
Paloh	10	1.8199703	109.4021154
Pemangkat	8	1.175333	108.970398
Sajad	9	1.3797652	109.4021154
Sajingan Besar	2	1.6898465	109.5211063
Sambas	58	1.362519	109.283153
Sebawi	13	1.271938	109.21179
Sejangkung	14	1.4700033	109.5211063
Selakau	4	1.056333	108.968498
Selakau Timur	1	1.0545813	109.1166566
Semparuk	11	1.1638103	109.0690976
Subah	6	1.1794272	109.4259114
Tangaran	3	1.528629	109.1166566
Tebas	24	1.223939	109.147598
Tekarang	8	1.2818773	109.1047663
Teluk Keramat	33	1.454493	109.21179
Jagoi Babang	2	1.3953336	109.7115472
Singawang Utara	1	0.9556428	109.0096562
Sanggau Ledo	1	1.1600876	109.5925139
Pontianak Barat	1	-0.0132044	109.3009955

Kemudian dilakukan metode optimasi menggunakan metode *elbow* untuk menentukan jumlah *cluster* terbaik. Berikut Gambar 1 Hasil Grafik 22 Kecamatan.



Gambar 1. Hasil Grafik Elbow Method 22 Kecamatan

Didapatkan titik yang paling optimal dilihat pada Gambar 2 pada $k=4$ grafik nya terlihat membentuk siku manusia dan jika sudah membentuk siku itulah titik yang paling optimal. Dapat dilihat pada Gambar 2.



Gambar 2. Metode Elbow for Optimal

Pada Gambar 3 visualisasi dari hasil K-Means Clustering dengan 4 cluster. Closest Centroid atau lingkaran magenta yaitu titik yang plaiing dekat dengan titik tertentu [0, 0]. Dapat dilihat pada Gambar 3.

Gambar 3. Visualisasi Hasil K-Means Clustering

KESIMPULAN

Berdasarkan metode Elbow, jumlah cluster yang optimal untuk analisis data mahasiswa Program Studi Manajemen Informatika menggunakan algoritma K-Means Clustering adalah 4 cluster. Ini berarti mahasiswa dapat secara efektif dikelompokkan kedalam 4 kelompok berbeda berdasarkan variabel yang dianalisis (seperti, Quarter / tanggal lahir, jenis kelamin, jalur pendaftaran dan jumlah data perkecamatan).

Hasil analisis data mahasiswa yang dihasilkan dari algoritma K-Means clustering dapat digunakan untuk merancang strategi promosi yang lebih terarah dan relevan untuk Program Studi Manajemen Informatika Politeknik Negeri Sambas, yaitu dengan melihat kecamatan mana yang lebih sedikit jumlah mahasiswa yang mendaftar di Prodi Manajemen Informatika.

UCAPAN TERIMA KASIH

Penelitian ini tidak akan mungkin terlaksana tanpa dukungan dan bantuan dari berbagai pihak. Oleh karena itu, dengan segala kerendahan hati, penulis ingin menyampaikan ucapan terima kasih kepada:

1. Politeknik Negeri Sambas yang telah memberikan dukungan finansial dan sarana yang diperlukan selama proses penelitian ini.
2. Bapak Fathushahib, S.Kom., M.Kom. yang telah terlibat secara langsung dalam pelaksanaa penelitian ini. Terima kasih atas dedikasi dan kontribusinya yang tak ternilai selama proses penelitian berlangsung.
3. Ibu Theresia Widji Astuti, S.Kom., M.Kom. dan Ibu Erifa Syahnaz, S.Kom., M.Kom. yang telah memberikan dukungan moral dan teknis. Kontribusi yang sangat penting dalam menyelesaikan penelitian ini tepat waktu.

PERNYATAAN KONTRIBUSI PENULIS

Dengan ini, penulis menyatakan bahwa kontribusi setiap penulis terhadap pembuatan karya tulis ini adalah sebagai berikut:

1. Helen Septriani sebagai kontributor utama yang bertanggung jawab atas konsep penelitian, analisis data dan penulisan draft utama.
2. Bapak Fathushahib, S.Kom., M.Kom. sebagai koresponden yang bertanggung jawab memberikan kontribusi dalam analisis statistik dan peninjauan literatur.
3. Ibu Theresia Widji Astuti, S.Kom., M.Kom. dan Ibu Erifa Syahnaz, S.Kom., M.Kom. sebagai anggota yang bertanggung jawab dalam revisi manuskrip.

DAFTAR PUSTAKA

- [1] Joanna Mita. (2019). Medicine Inventory Grouping Using Clustering Data Mining. Program Studi Teknik Informatika. Fakultas Teknologi Industri. Universitas Atma Jaya Yogyakarta.
- [2] Karina, M. Y. (2023). Penerapan Algoritma K-Means Untuk Penentuan Strategi Promosi(Studi Kasus: SMA PGRI 1 PURWAKARTA). Program studi teknik informatika S1.STT wastukancana purwakarta.
- [3] Lestari Widhi, (2019). Clustering Data Mahasiswa Menggunakan Algoritma K-Means Untuk Menunjang Strategi Promosi.STMIK Bina Bangsa Kendari.
- [4] Nanda, A. (2022). Penerapan Metode K-Means Clustering Dalam Menentukan Strategi Promosi Berdasarkan Data Penerimaan Mahasiswa Baru.fakultas teknologi dan informasi.universitas stikubank semarang indonesia.
- [5] Pita, S. (2024). Analisis Clustering Data Penduduk Miskin
- [6] Rima, D. R. (2022). Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Universitas Dian Nuswantoro. Jurusan Sistem Informasi. Fakultas Ilmu Komputer. Universitas Dian Nuswantoro.
- [7] Rosihan, A. Y. (2019). Konsep & Implementasi Pemograman PYTHON: Kasus BIG DATA. Perpustakaan Nasional: Katalog Dalam Terbitan (KDT). CV. LOKOMEDIA.
- [8] Sterry, N. A. (2018). Analisis Data Mahasiswa Menggunakan Algoritma K-Means Clustering Sbagai Dasar Pelaksana Promosi. Fakultas teknologi informatika. Universitas kristen satya wacana.
- [9] Teguh, I. (2022). Penerapan K-Means Clustering dan Cross-Industry Standart Process For Data Mining (CRISP-DM) Untuk Mengelompokkan Penjualan Kue. Program Studi Teknik Informatika. Sekolah Tinggi Teknologi Wastukancana