

Clusterization Of Productive Youth In Sirampog District Using The K-Means Method

Klasterisasi Remaja Produktif Di Kecamatan Sirampog Menggunakan Metode K-Means

M. ALI SUNAN

Universitas Muhammadiyah Brebes, Indonesia
Email: alisunan1605@gmail.com

Abstract

Adolescence is a transition from childhood to adulthood where maturity does not mean only physical maturity, but also psychological self-maturity, the age limit in question is that the body and soul are mature and can carry out marriage to realize the goals of marriage properly without ending in divorce. There are many clustering methods in data mining, one of which is the K-Means method which is included in the partition-based method, where this method divides data into a number of non-hierarchical groups (partitions) that try to partition data into clusters so that data that has the same characteristics will be included in the same cluster. Cluster evaluation in the form of a validity test of the Davies Bouldin Index (DBI) in clustering marriage readiness for adolescents of productive age in Sirampog District. The method used in this study is the experimental method. The object of research in this study is productive adolescents (20-29 years) in Sirampog District. The research was conducted from June 17, 2024 – August 18, 2024 (two months). The results of the validity test resulted in 4 clusters with details of Cluster 1 (cluster_0) as many as 8 people, cluster 2 (cluster_1) as many as 16 people, cluster 3 (cluster_2) as many as 17 people, cluster 1 (cluster_0) as many as 8 people with a total sample of 50 people. The research conducted proves that the performance of the K-Means Clustering Algorithm can be used to cluster adolescents of productive age ready to get married.

Keywords: K Means clustering, Davies Bouldin Index, Data Mining

Abstrak

Masa remaja adalah transisi dari masa kanak-kanak ke masa dewasa di mana kematangan tidak hanya berarti kematangan fisik, tetapi juga kematangan diri psikologis. Batasan usia yang dimaksud adalah tubuh dan jiwa telah matang dan mampu menjalani pernikahan untuk mewujudkan tujuan pernikahan dengan baik tanpa berakhir dengan perceraian. Terdapat banyak metode pengelompokan dalam penambangan data, salah satunya adalah metode K-Means yang termasuk dalam metode berbasis partisi, di mana metode ini membagi data menjadi sejumlah kelompok non-hierarkis (partisi) yang mencoba membagi data ke dalam kluster sehingga data yang memiliki karakteristik yang sama akan termasuk dalam kluster yang sama. Evaluasi kluster dalam bentuk uji validitas Indeks Davies Bouldin (DBI) dalam pengelompokan kesiapan pernikahan remaja usia produktif di Kabupaten Sirampog. Metode yang digunakan dalam penelitian ini adalah metode eksperimental. Objek penelitian dalam penelitian ini adalah remaja produktif (20-29 tahun) di Kabupaten Sirampog. Penelitian ini dilakukan dari tanggal 17 Juni 2024 – 18 Agustus 2024 (dua bulan). Hasil uji validitas menghasilkan 4 kluster dengan rincian Kluster 1 (cluster_0) sebanyak 8 orang, kluster 2 (cluster_1) sebanyak 16 orang, kluster 3 (cluster_2) sebanyak 17 orang, kluster 1 (cluster_0) sebanyak 8 orang dengan total sampel 50 orang. Penelitian ini membuktikan bahwa kinerja Algoritma K-Means Clustering dapat digunakan untuk mengelompokkan remaja usia produktif yang siap menikah.

Kata kunci: K Means clustering, Indeks Davies Bouldin, Data Mining

1. PENDAHULUAN

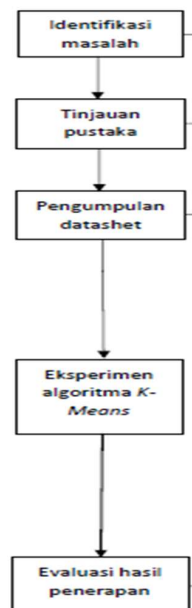
Perkembangan pengetahuan dan teknologi memanfaatkan data yang besar tersebut menjadi suatu pengetahuan baru yang dikenal dengan teknik data mining [1]. Data mining adalah proses pencarian pola yang penting dari data yang ada didalam *database* yang besar [2]. Tujuan dari data mining sendiri yaitu menemukan pola tersembunyi, hubungan atau wawasan yang tidak dapat ditemukan secara langsung melalui analisis tradisional [3]. Menurut [4] data mining dapat digunakan untuk mengelola data sehingga dapat menghasilkan informasi baru yang bermanfaat. Pada penelitian [5] bahwa pengambilan keputusan diwaktu yang akan datang dapat menggunakan dengan data mining.

Banyak metode *clustering* data mining, salah satunya adalah metode *K-Means* yang termasuk dalam metode berbasis partisi, dimana metode ini membagi data kedalam sejumlah kelompok [6]. Metode *K-Means clustering* merupakan salah satu metode pengelompokan non *hirarki* (sekatan) yang berusaha mempartisi data kedalam *cluster*/kelompok sehingga data yang memiliki karakteristik yang sama akan dimasukkan kedalam satu *cluster* yang sama [7]. Dimana data yang memiliki *representative* persamaan nilai dalam satu kelompok dan data yang memiliki perbedaan kelompok yang lain memungkinkan pengelompokan data yang berbeda yang memiliki tingkat variasi kecil [8].

Masa remaja yaitu masa transisi dari masa kanak-kanak menuju dewasa dengan terjadinya pertumbuhan dan perkembangan yang pesat dalam diri yang mencakup perkembangan perubahan fisik, kognif, emosional dan sosial [9]. Pada tahapan masa remaja, proses bimbingan orang tua sangat penting dalam perkembangan diri mereka [10].

2. METODE PENELITIAN

Metode yang digunakan pada penelitian ini yaitu metode eksperimen. Menurut [11] metode eksperimen merupakan metode pemberian kesempatan kepada anak didik perorangan atau kelompok untuk dilatih melakukan suatu proses atau percobaan. Pendekatan eksperimental bertujuan untuk menetapkan hubungan sebab-akibat antara variabel-variabel yang dipelajari. Metode yang digunakan berupa eksperimen algoritma *K-Means* untuk menentukan klasifikasi remaja produktif di Kecamatan Sirampog.



Gambar 2. 1 Tahapan Penelitian

2.1 Identifikasi masalah

Penelitian ini berkaitan dengan pemetaan remaja usia produktif di Kabupaten Brebes khususnya Kecamatan Sirampog menggunakan algoritma *K-Means*, diawali dengan tahapan identifikasi masalah

2.2 Studi Pustaka

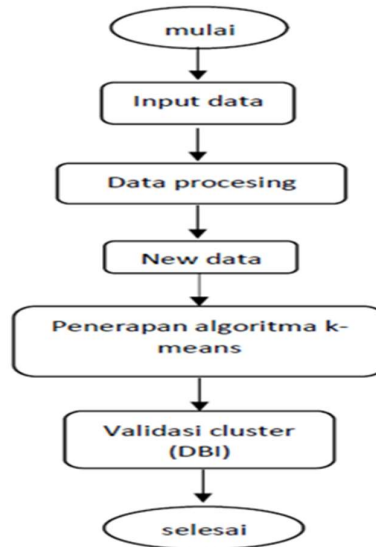
Studi Pustaka bertujuan untuk mengidentifikasi serta mengevaluasi penerapan algoritma *K-Means* yang berisi *review* dari penelitian-penelitian terdahulu serta landasan teori.

2.3 Pengumpulan Dataset

Dataset yang dipakai pada penelitian ini adalah remaja produktif Kecamatan Sirampog. Pengumpulan dataset menggunakan data *Google form* yang disebar dengan atribut-atribut

2.4 Eksperimen Algoritma *K-Means*

Penelitian ini menggunakan algoritma *K-Means* untuk pengelompokan remaja produktif di Kecamatan Sirampog, dengan tahapan dapat dilihat pada gambar 3.1.



Gambar 2. 2 Tahapan Eksperimen

2.4.1. Input data

Tahapan utama yang dilakukan pada penerapan algoritma *K-Means* adalah input data yang telah didapat dari proses pengumpulan data. Dataset yang digunakan berupa data hasil pengumpulan *Google Form*.

2.4.2. Data *processing*

Tahapan ini dilakukan dengan mentransformasikan data untuk mengubah struktur data, format data, dan nilai data menjadi dataset yang sesuai untuk melakukan implementasi proses algoritma yang sudah dirancang.

2.4.3. New data

Tahapan ini merupakan hasil dari pengolahan dataset awal yang telah melalui tahap data preprocessing sehingga menghasilkan dataset baru. New dataset terdiri dari data awal yang telah disesuaikan dengan kebutuhan sehingga dapat dilakukan pemrosesan data.

2.4.4. Penerapan algoritma *K-Means*

Tahapan ini merupakan algoritma *K-Means* dalam membuat pemetaan remaja usia produktif di Kabupaten Brebes sehingga memberikan hasil nilai cluster untuk menentukan angka kesiapan menikah di Kabupaten Brebes.

2.4.5. Validasi cluster (DBI)

Berdasarkan hasil *cluster* data yang diolah menggunakan algoritma *K-Means* dilakukan pengujian nilai validasi dengan *Davies Bouldin Index* (DBI) pada *RapidMiner*. Pengujian kerja algoritma yang lebih relevan dalam pemetaan data remaja produktif di Kabupaten Brebes

3. HASIL DAN PEMBAHASAN

Hasil penerapan algoritma *K-Means* untuk memetakan remaja usia produktif di Kecamatan Sirampog dengan metode eksperimen memiliki tahapan sebagai berikut :

3.1 Pengumpulan Dataset

Dataset yang dipakai pada penelitian ini adalah remaja produktif Kecamatan Sirampog. Pengumpulan dataset menggunakan data *Google form* yang disebar dengan atribut-atribut

3.2 Data Processing

Data set yang sudah diinput ke dalam *Rapidminer* memiliki value yang lengkap, sehingga tidak perlu melakukan perbaikan terhadap data tidak lengkap (*missing value*). Berdasarkan 50 *record* data, terdapat 6 atribut dan atribut nama sebagai label. Kemudian atribut jenis kelamin memiliki nilai *min* 1 dan *max* 2,

atribut umur *min* 1 *max* 4, atribut pendidikan *min* 2 dan *max* 4, atribut pekerjaan *min* 0 *max* 3, serta atribut penghasilan *min* 0 dan *max* 4.

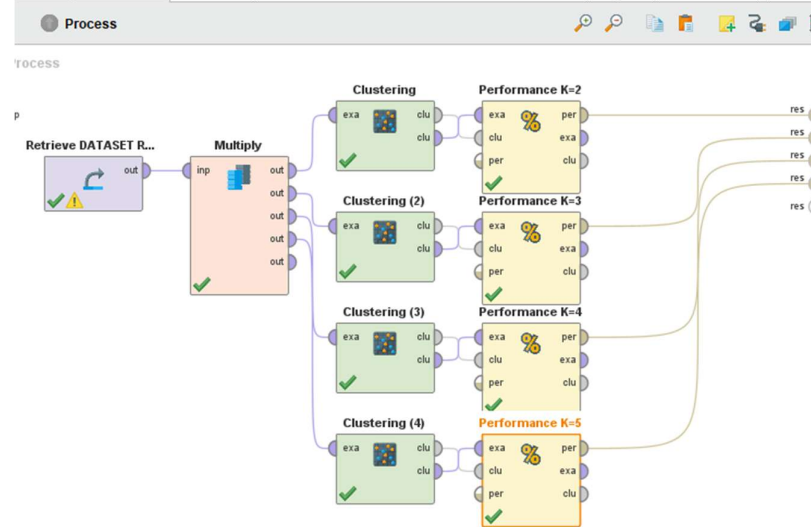
3.3 Eksperimen

3.3.1 Input Dataset

Data set yang dikumpulkan menggunakan *Google form*, kemudian diolah untuk dikategorikan sesuai kategori yang telah ditentukan. *Data Preprocessing*

3.3.2 Penerapan Metode K-means

Model yang diterapkan untuk penggunaan dan pengujian algoritma *K-means Clustering* dalam mengklasterisasi Remaja Produktif di Kecamatan Sirampog dapat dilihat pada gambar 3.1..



Gambar 3. 1 Penerapan Metode K Means pada Rapidminer

3.3.3 Validasi Cluster

Validasi *cluster* dengan menggunakan 4 operator *performance* yang menggunakan *main criterion Davies Bouldin*.

Tabel 3. 1 Centroid yang digunakan

ATRIBUT	JK	U	PD	PK	PH
C1	1.125	3.375	2.500	1.125	1.500
C2	1.625	1.812	3.375	1.062	1.375
C3	1.471	2.529	3.706	2.882	2.824
C4	1.667	1.111	3	0	0

Centroid digunakan untuk menguji parameter kedekatan masing-masing cluster. *Centroid* yang dihasilkan menggunakan teori jarak *Euclidean*. *Cluster* awal yang diusulkan adalah 4 kluster seperti *Centroid* yang terlihat pada tabel 4.2. Hasil validasi dari masing-masing cluster yang diuji yaitu sebagai berikut :

```

Performance Vector (Performance K=2)
Result not stored in repository.

PerformanceVector:
Avg. within centroid distance: -2.149
Avg. within centroid distance_cluster_0: -2.127
Avg. within centroid distance_cluster_1: -2.162
Davies Bouldin: -0.979

```

Gambar 3.2 Hasil *Performance* K=2

Pada gambar 4.8, dapat dilihat bahwa rata-rata jarak dengan Centroid adalah -2.149. Sementara jarak Centroid dengan Cluster 1 adalah -2.127 dan jarak Centroid dengan Cluster 2 adalah -2.162. Dengan hasil tersebut, *Performance* yang membagi 2 klaster memiliki DBI sebesar -0.979.

```
Performance Vector (Performance K=3)
Result not stored in repository.

PerformanceVector:
Avg. within centroid distance: -1.622
Avg. within centroid distance_cluster_0: -1.456
Avg. within centroid distance_cluster_1: -2.009
Avg. within centroid distance_cluster_2: -1.306
Davies Bouldin: -1.097
```

Gambar 3.3 Hasil *Performance* K=3

Pada gambar 4.9, dapat dilihat bahwa Rata-rata jarak dengan Centroid adalah -1.622. Sementara jarak Centroid dengan Cluster 1 adalah -1.456 dan jarak Centroid dengan Cluster 2 adalah -2.009. jarak Centroid dengan Cluster 3 adalah -1.306. Dengan hasil tersebut, *Performance* yang membagi 3 klaster ini memiliki DBI sebesar -1.097.

```
Performance Vector (Performance K=4)
Result not stored in repository.

PerformanceVector:
Avg. within centroid distance: -1.227
Avg. within centroid distance_cluster_0: -0.953
Avg. within centroid distance_cluster_1: -1.164
Avg. within centroid distance_cluster_2: -1.896
Avg. within centroid distance_cluster_3: -0.321
Davies Bouldin: -0.977
```

Gambar 4. 1 Hasil *Performance* K=4

Pada gambar 4.10, dapat dilihat bahwa Rata-rata jarak dengan Centroid adalah -1.227. Sementara jarak Centroid dengan Cluster 1 adalah -0.953, jarak Centroid dengan Cluster 2 adalah -1.164, jarak Centroid dengan Cluster 3 adalah -1.096, jarak Centroid dengan Cluster 4 adalah -0.321. Dengan hasil tersebut, *Performance* yang membagi 4 klaster ini memiliki DBI sebesar -0.977.

```
Performance Vector (Performance K=5)
Result not stored in repository.

PerformanceVector:
Avg. within centroid distance: -1.034
Avg. within centroid distance_cluster_0: -0.321
Avg. within centroid distance_cluster_1: -1.510
Avg. within centroid distance_cluster_2: -1.164
Avg. within centroid distance_cluster_3: -0.953
Avg. within centroid distance_cluster_4: -1.200
Davies Bouldin: -1.172
```

Gambar 3.4 Hasil *Performance* K=5

Pada gambar 4.11, dapat dilihat bahwa Rata-rata jarak dengan Centroid adalah -1.034. Sementara jarak Centroid dengan Cluster 1 adalah -0.321, jarak Centroid dengan Cluster 2 adalah -1.510, jarak Centroid dengan Cluster 3 adalah -1.164, jarak Centroid dengan Cluster 4 adalah -0.953, jarak Centroid dengan Cluster 5 adalah -1.200. Dengan hasil tersebut, *performance* yang membagi 5 klaster ini memiliki DBI sebesar -1.172.

3.4 Hasil Eksperimen

Berdasarkan Hasil penerapan algoritma *K-means* untuk mengklasterisasi remaja produktif di kecamatan sirampog didapatkan hasil bahwa remaja produktif di kecamatan sirampog dapat diklasterisasi

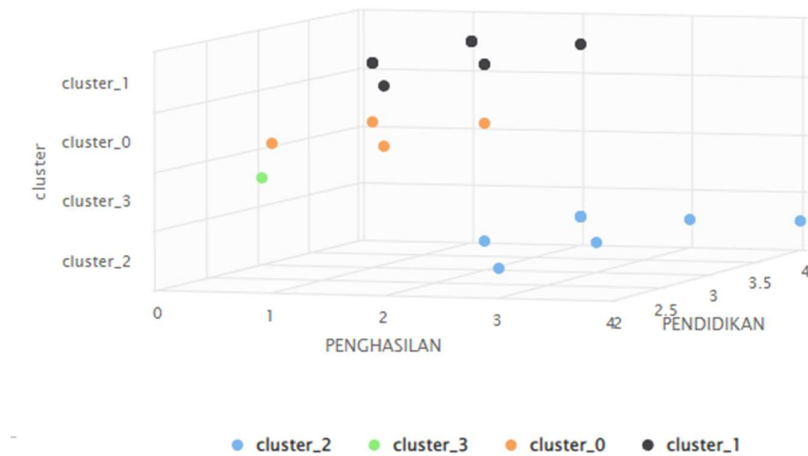
menjadi 4 kluster. Hal tersebut merupakan hasil uji *performance* yang telah dilakukan dihasilkan bahwa *performance* K=4 yang membagi merupakan yang terbaik karena memiliki nilai DBI paling kecil yakni -0.977. Hasil rekapitulasi uji *performance* ditunjukkan pada tabel 4.3

Tabel 3.2 Hasil Uji *Performance*

<i>Performance</i>	Nilai DBI
K = 2	-0.979
K = 3	-1.097
K = 4	-0.977
K = 5	-1.172

3.1 Kategori Klasterisasi

Berdasarkan hasil uji performa yang telah dilakukan ditetapkan 4 kluster remaja produktif di kecamatan Sirampog yang divisualisasikan pada gambar 4.12.



Gambar 3.5 Visualisasi Hasil Klasterisasi

Klaster 1 (cluster_0) sebanyak 8 orang, klaster 2 (cluster_1) sebanyak 16 orang, klaster 3 (cluster_2) sebanyak 17 orang, klaster 1 (cluster_0) sebanyak 8 orang. Pengelompokan klaster dapat dilihat pada tabel 4.4 dengan sampel sebanyak 50 orang remaja di kecamatan Sirampog.

Tabel 3.3 Hasil Pengelompokan K = 4 Algoritma K means

Cluster 0	Cluster 1	Cluster 2	Cluster 3
K7	K8	K1	K2
K11	K9	K4	K3
K19	K16	K12	K5
K21	K20	K15	K6
K28	K22	K17	K10
K38	K23	K18	K13
K47	K29	K24	K14
K48	K31	K25	K30
	K32	K26	K40
	K33	K27	
	K35	K34	
	K36	K37	
	K39	K41	
	K44	K42	
	K45	K43	
	K50	K46	
		K49	

Pengelompokan remaja produktif di kecamatan Sirampog merupakan atribut nama yang digunakan sebagai Id dengan kode kriteria K1 sampai dengan K50 sejumlah sampel yaitu sebanyak 50 orang.

Berdasarkan proses yang sudah dilakukan menghasilkan Klaster 1 (cluster_0) sebanyak 8 orang, klaster 2 (cluster_1) sebanyak 16 orang, klaster 3 (cluster_2) sebanyak 17 orang, klaster 1 (cluster_0) sebanyak 8 orang dengan sampel sebanyak 50 orang remaja di kecamatan Sirampog. Pengklasteran dapat dijadikan bahan untuk pemerintah dalam membuat program untuk pemberdayaan remaja usia produktif menikah.

4. KESIMPULAN

Penelitian ini menerapkan algoritma *K-Means* sebagai usulan untuk mengklaster remaja usia produktif di Kecamatan Sirampog. Algoritma pemodelan yang digunakan merupakan komponen dari metode *clustering* pada *Data Mining* yang lazim digunakan untuk mengelompokkan data. Berdasarkan hasil penelitian yang telah dilakukan, klasterisasi remaja produktif di Kecamatan Sirampog dapat dibagi menjadi 4 Klaster. Hasil uji *validitas* pada $K=3$ menggunakan algoritma *K-Means* sebesar -0.977, yang merupakan nilai terkecil dibandingkan dengan $K=1$ yang sebesar -0.979, $K=2$ yang sebesar -1.-097 dan $K=4$ yang sebesar -1.172.

UCAPAN TERIMA KASIH

Terima kasih kepada Universitas Muhammadiyah Brebes, Indonesia, yang telah mengizinkan kami untuk mengumpulkan data agar penelitian ini dapat berjalan dengan baik.

REFERENSI

- [1] Hidayat, «Klasifikasi Data Jamaah Umroh Menggunakan Metode K-Means Clustering,» *Jurnal Sistem Informasi dan Teknologi*, p. 19024, 2022.
- [2] F. O. O. & S. Sembiring, «Implementasi Metode K-Means Dalam Pengklasteran Daerah Pungutan Liar Di Kabupaten Sukabumi (Studi Kasus : Dinas Kependudukan Dan Pencatatan Sipil),» *Jurnal Tekno Insentif*, pp. 40-47, 2020.
- [3] R. N. Baji Syadewo, « Klasifikasi Penerimaan Dana Bantuan Pada Dusun Jati Bening,» *JATI (Jurnal Mahasiswa Teknik Informatika)*, Vol. %1 av %2 7 (2),, p. 122 0–1226, 2023.
- [4] Ariyanto, «Data Mining Menggunakan Algoritma K-Means untuk Klasifikasi Penyakit Infeksi Saluran Pernafasan Akut,» *Jurnal Sistem Informasi dan Teknologi*, vol. 4, pp. 13-18, 2022.
- [5] H. Sudarto, «Implementasi Data Mining Pada Pengaturan Data Invoice Distributor Menggunakan Algoritma Fp Growth,» *Computer Science Industrial Engineering (COMASIE)*, vol. 9, nr. 2, 2023.
- [6] Bahtiar, «Pemetaan Penduduk Penerima Bantuan Sosial Desa Waru Jaya Menggunakan Algoritma K-Means Clustering,» *Scientia Sacra: Jurnal Sains*, vol. III, nr. 2, p. 29–39, 2023.
- [7] «Lesmana, B. (2021). Pengelompokan Pengiriman Hasil Kelapa Sawit Berdasarkan Tonase dan Kualitas Menggunakan Metode Clustering (Studi Kasus: KUD Bumi Pusaka),» *Seminar Nasional Informatika (SENATIKA)*, pp. 154-165, 2021.
- [8] S. Sulistiyawati, «Implementasi Algoritma K-means Clustring dalam Penentuan Siswa Kelas Unggulan,» *Jurnal Tekno Kompak*, vol. 2, pp. 15-25, 2021.
- [9] I. G. A. D. S. K. I. K. Rusdayanti, «Penerapan asas kesukarelaan dalam konseling kasus pacaran beresiko pada remaja,» *Jurnal EDUCATIO: Jurnal Pendidikan Indonesia*, vol. 9, nr. 1, 2023.
- [10] H. & U. M. S. S. Damhila, «Efektivitas Pelatihan Kebersyukuran Meningkatkan Kesejahteraan Subjektif Remaja Broken Home,» *Disputare-Jurnal Psikologi*, vol. 10, nr. 2, pp. 64-74, 2023.
- [11] F. A. L. U. N. D. M. O. & L. Z. A. Aufa, «Proses Peningkatan Hasil Pembelajaran IPA melalui Media Pembelajaran dan Metode Eksperimen di SD IT Miftahul Jannah Bandar Selamat, Kec. Medan Tembung,» *Journal on Education*, vol. 5, nr. 4, 2023.