

Analisis Klasifikasi Metode X-Means Pada Minat dan Bakat Anak Dimasa Pandemi

Purwa Hasan Putra¹, Alfiansyah Hasibuan², Effenril Agung Marpaung³

^{1,2,3}Sistem Komputer, Faculty of Science and Technology, Universitas Pembangunan Panca Budi, Jl. Jenderal Gatot Subroto, KM 4,5 Sei Sikambang 20122 Medan, Sumatera Utara Indonesia

Email: purwahasanutra@dosen.pancabudi.ac.id

ABSTRAK

Algoritma X-Means merupakan algoritma yang digunakan untuk pengelompokan data. Algoritma x means merupakan pengembangan dari k-means. X-means clustering digunakan untuk menyelesaikan salah satunya kelemahan utama dari K-means clustering, yaitu perlunya pengetahuan sebelumnya tentang jumlah cluster (K). Dalam metode ini, nilai sebenarnya dari K diperkirakan dalam suatu yang tidak diawasi cara dan hanya berdasarkan set data itu sendiri. Adapun hasil penelitian dengan menggunakan algoritma X-Means dengan evaluasi Davies-Bouldin Index penentuan jumlah Centroid cluster dilakukan dengan memodifikasi metode X-Means. Dalam pengelompokan data ini dilakukan pengelompokan pada setiap data siswa dari variabel-variabel yang telah dikumpulkan. Bakat dan minat setiap siswa akan dicocokkan dengan perguruan tinggi dan jurusan apa yang diminati dari setiap siswa.

Kata Kunci: Data Mining, Minat dan Bakat Anak

ABSTRACT

The X-Means algorithm is an algorithm used for grouping data. The X-means algorithm is a development of k-means. X-means clustering is used to solve one of the main weaknesses of K-means clustering, namely the need for prior knowledge about the number of clusters (K). In this method, the true value of K is estimated in an unsupervised manner and based solely on the data set itself. The results of the research using the X-Means algorithm with the evaluation of the Davies-Bouldin Index. Determining the number of Centroid clusters is done by modifying the X-Means method. In this data grouping, clustering is carried out on each student data from the variables that have been collected. will be matched to the college and what major each student is interested in.

Keywords: Data Mining, Children's Interests and Talents

Pendahuluan

Pengelompokan dapat menggunakan clustering untuk mengelompokkan data yang didasari pada kemiripan antar data, sehingga data dengan kemiripan paling dekat berada dalam satu cluster sedangkan data yang berbeda dalam kelompok lainnya. Proses dalam mengelompokkan data ke beberapa cluster atau pengelompokan sehingga data pada satu cluster memiliki tingkat kemiripan yang maksimum dan antar cluster memiliki kemiripan yang minimum disebut dengan Clustering [1], [2].

Menurut Shaleh Abdul Rahman dalam bukunya psikologi suatu pengantar dalam perspektif islam, menjelaskan bahwa minat adalah suatu kecenderungan untuk memberikan perhatian dan

bertindak terhadap orang, aktivitas atau situasi yang menjadi objek dari minat tersebut dengan disertai perasaan atau gembira. 1 Bakat Menurut William B. Michael Bakat merupakan kapasitas pada diri seseorang dalam melakukan tugasnya dan melakukan dengan pengaruh dan latihan yang dijalannya [3], [4].

Adapun tujuan dari *clustering* adalah agar objek-objek (data) dalam suatu kelompok yang sama (terkait) satu sama lain dan berbeda (tidak terkait) objek-objek dalam kelompok lain. Semakin besar kesamaan (homogenitas) dalam suatu kelompok dan semakin besar perbedaan antar kelompok, semakin baik atau lebih jelas pengelompokan. Salah satu algoritma yang dapat digunakan dalam pengelompokan adalah X-Means [5]–[7].

Metode Penelitian

Pada panduan ini akan dijelaskan tentang penulisan *heading*. Jika *heading* anda melebihi satu, gunakan level kedua heading seperti di bawah ini.

Machine Learning

Machine learning memungkinkan manusia untuk memprogram komputer sehingga mesin dapat mengenali pola atau belajar dari apa yang dimasukkan ke dalamnya. Konsep pembelajaran adalah proses mengubah informasi menjadi pengetahuan. Algoritma machine learning adalah data pelatihan, yang mewakili pengalaman, dan hasilnya adalah beberapa keahlian, yang biasanya berupa program komputer lain yang dapat melakukan beberapa tugas.

Machine learning memungkinkan dalam klasifikasi data, aplikasi ini mengenali pola dalam data baik dengan pelatihan atau tanpa pelatihan. Di klasifikasi data disebut clustering dalam machine learning. Beberapa contoh algoritma pengelompokan termasuk K-Means, Farthest-First Maksimalisasi-Ekspektasi (EM), dan lainnya[8]–[10].

X-Means

Algoritma X-Means dikembangkan oleh Dan Pelleg dan Andre Moore pada tahun 2000. Dalam algoritma ini jumlah cluster dihitung secara dinamis menggunakan batas atas dan bawah yang disediakan oleh pengguna. Algoritma ini terdiri dari dua langkah yang diulang sampai selesai.

1. Tingkatkan-Params, pada langkah ini menerapkan algoritma k-means pada awalnya untuk k cluster hingga konvergensi. Dimana k sama dengan batas bawah yang disediakan oleh pengguna.
2. Perbaiki Struktur, langkah perbaikan struktur ini dimulai dengan memecah setiap pusat cluster menjadi dua anak dalam arah yang berlawanan di sepanjang vektor yang dipilih secara acak. Setelah itu menjalankan k-means secara lokal di dalam setiap cluster untuk dua cluster. Keputusan masing-masing pusat cluster sendiri dengan membandingkan nilai-nilai BIC.
3. Jika $K \geq k_{max}$ (batas atas) berhenti dan laporkan ke model penilaian terbaik yang ditemukan selama penarian, jika tidak pergi ke langkah 1.

X-Means berarti mengambil keuntungan dari Informasi Bayesian Cri terion (BIC) untuk mengontrol proses pemisahan cluster. Dengan kata lain, jika memecah satu cluster menjadi dua clusters meningkatkan skor BIC, kemudian memiliki dua kelompok lebih mungkin daripada satu

cluster. Dalam makalah ini, kami sarankan untuk menggunakan Panjang Deskripsi Berisik Minimum (MNDL) sebagai kriteria pemisahan klaster, yang mengarah untuk prediksi yang lebih tepat untuk jumlah cluster[11]–[13].

X-means clustering digunakan untuk menyelesaikan salah satunya kelemahan utama dari K-means clustering, yaitu perlunya pengetahuan sebelumnya tentang jumlah cluster (K). Dalam metode ini, nilai sebenarnya dari K diperkirakan dalam suatu yang tidak diawasi cara dan hanya berdasarkan set data itu sendiri[14]–[16].



Gambar 1. Langkah-Langkah Umum dalam Pengelompokan X-Means

K_{max} dan K_{min} sebagai batas atas dan bawah untuk nilai yang mungkin dari X. Pada langkah pertama X-means pengelompokan, mengetahui bahwa saat ini $X = X_{min}$, X-berarti menemukan struktur awal dan centroid. Di langkah selanjutnya, setiap cluster dalam struktur yang diperkirakan diperlakukan sebagai induk cluster, yang dapat dibagi menjadi dua kelompok[17], [18].

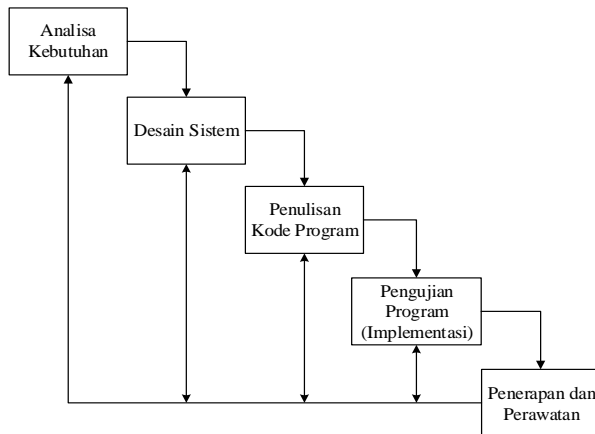
Algoritma ini bisa terlalu lambat karena perlu menjalankan kembali K berarti untuk setiap pemisahan klaster. Untuk mengatasi masalah ini, menerapkan kd-tree dari kumpulan data disarankan dalam, yang secara alami mengurangi jumlah permintaan tetangga terdekat untuk K-means[19], [20].

Hasil dan Pembahasan

Dari hasil dan pembahasan di sini diuraikan metodologi dan kerangka kerja penelitian digunakan

untuk menggambarkan langkah-langkah yang akan diterapkan dalam melakukan penelitian.

Kerangka Kerja Penelitian



Gambar 2. Kerangka Kerja Penelitian

Berdasarkan gambar 2 penulis dapat menjelaskan beberapa kerangka kerja yang akan dilakukan dalam penelitian ini, yaitu sebagai berikut :

1. **Mempelajari Literatur**
 Pada penelitian ini dipelajari literatur yang berhubungan dengan permasalahan. Kemudian literatur yang dipelajari diseleksi untuk dapat ditentukan literatur mana yang akan digunakan dalam penelitian. Sumber literatur didapatkan dari perpustakaan, jurnal, artikel dan konsep-konsep lain yang mendukung dalam menyelesaikan sistem yang akan dibangun termasuk referensi.
2. **Pengumpulan Data**
 Dalam melakukan penelitian ini, pengumpulan data dan informasi pada tahap ini dilakukan untuk mengetahui mengenai sistem yang diteliti. Dari data dan informasi yang dikumpulkan akan didapat data untuk pendukung penelitian serta pengumpulan data dilakukan untuk mengetahui kebutuhan dari pengguna. Metode yang digunakan penulis untuk pengumpulan data adalah sebagai berikut :
 - a. **Observasi**
 Observasi berguna untuk melakukan pengumpulan data dan observasi dengan langsung terjun kelapangan pada pihak-pihak yang terkait dalam menyelesaikan penelitian ini dimana informasi dan materi akan diperoleh sebagai bahan dari rancang bangun sistem.
 - b. **Wawancara**
 Melakukan wawancara pada pihak yang berkaitan dengan alur permasalahan.

Wawancara ini dilakukan untuk mendapatkan bahan penulisan dan penjelasan pengamatan yang dilakukan.

3. **Analisa Kebutuhan**
 Analisis dapat didefinisikan sebagai penguraian dari suatu sistem informasi yang utuh kedalam bagian-bagian komponennya. Analisa kebutuhan ini bertujuan untuk mengetahui apa saja yang dibutuhkan dalam perancangan sistem yang di bangun dan mengetahui kebutuhan-kebutuhan pendukung dari perancangan sistem.
4. **Desain dan Perancangan Sistem**
 Kegiatan desain sistem dilakukan untuk sebagai awal dari perancangan sistem yang akan dibangun sesuai kebutuhan. Dan pada tahap ini dilakukan perancangan antar muka terhadap sistem yang akan dibuat.
5. **Implementasi Sistem**
 Implementasi sistem dilakukan sesuai desain dan rancangan antar muka aplikasi yang akan dibangun. Pada tahap ini melakukan pengkodean atau pembuatan program sehingga sistem yang dirancang dapat digunakan oleh pengguna.
6. **Pengujian Sistem**
 Pengujian sistem dilakukan untuk mengetahui uji kelayakan sistem yang telah dibangun sesuai yang diharapkan dan dengan dilakukannya pengujian dapat mengetahui kelemahan serta kelebihan dari sistem yang dirancang sehingga dapat dilakukan perbaikan pada tahap selanjutnya.
7. **Maintenance**
 Pemeliharaan Sistem.

Penerapan X-Means

Adapun penelitian ini, menggunakan algoritma X-Means sebagai metode klasifikasi data set iris dan menggunakan algoritma euclidean distance dan manhattan distance sebagai metode untuk menghitung jarak antara data tes dan data uji pada algoritma X-Means. Dan akan dianalisis dengan algoritma X-Means menggunakan metode jarak Euclidean Distance dengan perbandingan metode jarak Manhattan Distance.

- a. Tentukan kelompokkan data yang dekat dengan centroid. Adapun ditunjukkan sebagai berikut.

Tabel 1. Cluster 1

Cluster 1	
Bidang olahraga	No.3
Bidang kedokteran	No.2
Bidang teknik	No.1

Tabel 2. Cluster 2

Cluster 2	
-----------	--

Bidang hewani	No.10
Bidang seni, desain dan media	No.8
Bidang Pendidikan	No.7
Bidang olahraga	No.9
Bidang agama dan filsafat	No.6
Bidang bahasa	No.4
Bidang ekonomi	No.5

$$= \frac{4.7 + 4.9 + 5.1}{3} = 4.9$$

$$= \frac{3.2 + 3 + 3.5}{3} = 3.233$$

$$= \frac{1.3 + 1.4 + 1.4}{3} = 1.366$$

$$= \frac{0.2 + 0.2 + 0.2}{3} = 0.2$$

b. Hitunglah mean x dan mean y dari cluster yang terbentuk. Adapun ditunjukkan sebagai berikut.

$$= \frac{x_3 + x_2 + x_1}{\text{jumlah x}}$$

Berikut merupakan hasil dari perhitungan untuk mencari nilai mean x dan mean y berdasarkan klasifikasi data titik centroid cluster 1. Ditunjukkan pada tabel 3.

Tabel 3. Nilai Mean X dan Y Cluster 1

Cluster 1		Mean X1	Mean X2	Mean Y1	Mean Y2
Bidang olahraga	No.3	4.9	3.233333333	1.366666667	0.2
Bidang kedokteran	No.2				
Bidang teknik	No.1				

$$= \frac{y_{10} + y_8 + y_7 + y_9 + y_6 + y_4 + y_5}{\text{jumlah y}}$$

$$= \frac{6.3 + 5.8 + 6.3 + 7.1 + 6.9 + 7 + 6.4}{7} = 6.542$$

$$= \frac{2.9 + 2.7 + 3.3 + 3 + 3.1 + 3.2 + 3.2}{7} = 3.057$$

$$= \frac{5.6 + 5.1 + 6 + 5.9 + 4.9 + 4.7 + 4.5}{7} = 5.242$$

$$= \frac{1.8 + 1.9 + 2.5 + 2.1 + 1.5 + 1.4 + 1.5}{7} = 4.708$$

Berikut merupakan hasil dari perhitungan untuk mencari nilai mean x dan mean y berdasarkan klasifikasi data titik centroid cluster 2. Ditunjukkan pada tabel 4.

Table 4. Nilai Mean X dan Y Cluster 2

Cluster 2		Mean X1	Mean X2	Mean Y1	Mean Y2
Bidang hewani	No.10				
Bidang seni, desain dan media	No.8				
Bidang Pendidikan	No.7				
Bidang olahraga	No.9	6.542	3.057	5.242	1.8142
Bidang agama dan filsafat	No.6				
Bidang bahasa	No.4				
Bidang ekonomi	No.5				

Tabel 5. Kelompok Cluster Setiap Data

No	Nama Item	Kelompok Cluster
1	Aulya	Cluster 1
2	Siti Mardiyah	Cluster 1
3	Heldi Irawan	Cluster 1
4	Shandy Setiawan	Cluster 2
5	Tiara Azzahra	Cluster 2
6	Asbiyadi	Cluster 2
7	Iris Virginica	Cluster 2
8	Mutia Azizah	Cluster 2

9	Lusi	Cluster 2
10	M.Ashar F	Cluster 2

Setelah proses cluster, dari tabel 3 dapat dilihat pengelompokan cluster setiap data berdasarkan titik pusat cluster. Adapun berikut merupakan hasil dari perhitungan iterasi ke-1, dimana Data ke 2 dan 8 menjadi titik pusat cluster. Untuk iterasi selanjutnya dihitung dengan tahap yang sama sampai ditemukannya konvergen setiap data yang di klasifikasi.

Kesimpulan

Adapun kesimpulan dari Analisis Klasifikasi Minat dan Bakat Anak Dimasa Pandemi yang penulis bangun sebagai berikut :

1. Sistem pendukung keputusan yang dibangun dapat membantu dan memberikan kemudahan dalam proses klasifikasi minat dan bakat anak.
2. *Setting* variabel penilaian dilakukan melalui sistem. Jadi, jika ada perubahan nilai maksimal dan minimal variabel, admin tinggal mengupdate melalui sistem tanpa harus merubah programnya.
3. Berdasarkan penelitian yang dilakukan, dapat di ketahui bahwa dengan kumpulan nilai ujian yang di lakukan siswa, dapat di ketahui seberapa minat siswa dengan mata pelajaran matematika. Hasil clustering dengan menggunakan algoritma k-means memperlihatkan bahawa 45 siswa memiliki minat tinggi dengan mata pelajaran matematika, 48 siswa memiliki minat sedang dan 29 siswa memiliki minat rendah pada mata pelajaran matematika.

Daftar Pustaka

- [1] N. Arminarahmah, A. D. GS, G. W. Bhawika, M. P. Dewi, and A. Wanto, "Mapping the Spread of Covid-19 in Asia Using Data Mining X-Means Algorithms," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1071, no. 1, p. 012018, 2021, doi: 10.1088/1757-899x/1071/1/012018.
- [2] M. Mughnyanti, S. Efendi, and M. Zarlis, "Analysis of determining centroid clustering x-means algorithm with davies-bouldin index evaluation," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 725, no. 1, pp. 0–6, 2020, doi: 10.1088/1757-899X/725/1/012128.
- [3] R. Adhitama, A. Burhanuddin, and R. Ananda, "Penentuan Jumlah Cluster Ideal Smk Di Jawa Tengah Dengan Metode X-Means Clustering Dan K-Means Clusterin," *JIKO (Jurnal Inform. dan Komputer)*, vol. 3, no. 1, pp. 1–5, 2020, doi: 10.33387/jiko.v3i1.1635.
- [4] R. K. Dhanaraj *et al.*, "Random Forest Bagging and X-Means Clustered Antipattern Detection from SQL Query Log for Accessing Secure Mobile Data," *Wirel. Commun. Mob. Comput.*, vol. 2021, 2021, doi: 10.1155/2021/2730246.
- [5] A. Radwan *et al.*, "X-means clustering for wireless sensor networks," *J. Robot. Netw. Artif. Life*, vol. 7, no. 2, pp. 111–115, 2020, doi: 10.2991/jrnal.k.200528.008.
- [6] N. Zentrato, H. W. Dhany, N. A. Siagian, and F. Izhari, "Bigdata Clustering using X-means method with Euclidean Distance," *J. Phys. Conf. Ser.*, vol. 1566, no. 1, pp. 0–6, 2020, doi: 10.1088/1742-6596/1566/1/012103.
- [7] S. Anwar, N. D. Nuris, and Y. A. Wijaya, "Pengelompokkan Tingkat Pemahaman Kurikulum Berbasis KKNI Menggunakan Metode X-Means Clustering," *J. Inform. J. Pengemb. IT Poltek Tegal*, vol. 04, no. 2, pp. 187–190, 2019, doi: 10.30591/jpit.v4i2-2.1869.
- [8] S. Chaikul, S. Phithakitnukoon, and C. Ratti, "Xplaces: Segmenting Physical Space Through Wi-Fi Traces Using Eigendecomposition and X-Means," *IEEE Access*, vol. 10, pp. 17377–17388, 2022, doi: 10.1109/ACCESS.2022.3150006.
- [9] A. Wijayanto, "Penggunaan X-Means Clustering Method untuk Mengelompokkan Potensi Sekolah Menengah Unggul di Kabupaten Banyumas," *J. Informatics, Inf. Syst. Softw. Eng. Appl.*, vol. 2, no. 1, pp. 80–88, 2019, doi: 10.20895/inista.v2i1.99.
- [10] A. Radwan, N. H. Kamarudin, M. I. Solihin, H. Leong, and M. Rizon, "Implementation of the x-means clustering algorithm for wireless sensor networks," *Proc. Int. Conf. Artif. Life Robot.*, vol. 2020, no. January, pp. 333–337, 2020, doi: 10.5954/ICAROB.2020.OS10-4.
- [11] S. Monica, F. Natalia, and S. Sudirman, "Clustering Tourism Object in Bali Province Using K-Means and X-Means Clustering Algorithm," *Proc. - 20th Int. Conf. High Perform. Comput. Commun. 16th Int. Conf. Smart City 4th Int. Conf. Data Sci. Syst. HPCC/SmartCity/DSS 2018*, no. 1, pp. 1462–1467, 2019, doi: 10.1109/HPCC/SmartCity/DSS.2018.00241.
- [12] F. Noorbebhahani and S. Mansoori, "A new semi-supervised method for network traffic classification based on X-means clustering and label propagation," *2018 8th Int. Conf. Comput. Knowl. Eng. ICCKE 2018*, no. Iccke, pp. 120–125, 2018, doi: 10.1109/ICCKE.2018.8566608.
- [13] J. Ge *et al.*, "LPX: Overlapping community detection based on X-means and label propagation algorithm in attributed networks," *Comput. Intell.*, vol. 37, no. 1, pp. 484–510, 2021, doi: 10.1111/coin.12420.
- [14] S. Mudgil, "UNSUPERVISED DATA SENTIMENT ANALYSIS USING K-MEANS AND THE X- UNSUPERVISED DATA SENTIMENT ANALYSIS USING K-MEANS AND THE X- MEANS ALGORITHM," no. June, 2022.
- [15] S. K. Hamid, W. A. Teniwut, R. M. K. Teniwut, and M. Renhoran, "Outliers Detection on Fisheries Commodity Transaction from Local Market in Tual City based on the x-means Clustering," *J. Phys. Conf. Ser.*, vol. 1424, no. 1, pp. 0–7, 2019, doi: 10.1088/1742-6596/1424/1/012017.
- [16] M. Anoop and P. Sripriya, "Focused

- information criterion based partitioned iterative X-means dice correlation clustering for big Geo-social data,” *J. Crit. Rev.*, vol. 7, no. 6, pp. 54–62, 2020, doi: 10.31838/jcr.07.06.13.
- [17] A. Radwan, N. H. Kamarudin, M. I. Solihin, and H. Leong, “Slitting K-means clusters to X-means clusters for prolonging wireless sensor networks lifetime,” *AIP Conf. Proc.*, vol. 2306, no. December, 2020, doi: 10.1063/5.0032429.
- [18] H. Jeong and E. R. Lee, “Modified multi-sense skip-gram using weighted context and x-means,” vol. 34, no. 3, pp. 389–399, 2021.
- [19] M. Ahmed, R. Seraj, and S. M. S. Islam, “The k-means algorithm: A comprehensive survey and performance evaluation,” *Electron.*, vol. 9, no. 8, pp. 1–12, 2020, doi: 10.3390/electronics9081295.
- [20] B. S. B and C. Plant, *KMN - Removing Noise from K-Means*, vol. 1. Springer International Publishing, 2018. doi: 10.1007/978-3-319-98539-8.