

Algoritma *K-Means Clustering* Sebagai Penentuan Beasiswa PIP pada SMPN 9 Kota Blitar

Rizki Dwi Romadhona^{a,1,*}, Arta Ainur Rofiq^{a,2}

^a Universitas Islam Balitar, Jl. Majapahit No. 2-4, Blitar, Indonesia

¹ rizkidwiromadhona@unisbablitar.ac.id; ² arthaainurrofiq@gmail.com

* Penulis Koresponden

INFO ARTIKEL

Histori Artikel

29-Mei-2023

06-Juni-2023

14-Juni-2023

Kata Kunci

Beasiswa PIP

K-Means Clustering

Euclidean distance

ABSTRAK

Struktur dalam penelitian ini melibatkan pengambilan data perolehan beasiswa PIP dengan berbagai indikator yang telah ditetapkan (penghasilan orangtua, pekerjaan orangtua, kepemilikan kendaraan, KIP, KKS, tempat tinggal, jumlah saudara, PKH) sebagai penerima beasiswa. Dilaksanakan di salah satu sekolah negeri SMPN 9 Kota Blitar. Dimana data diolah dengan algoritma unsupervised learning yaitu *K-Means Clustering* dengan perhitungan *euclidean distance* dengan jarak nilai yang terdekat daripada *centroid* yang telah ditentukan yang kemudian untuk menentukan penerima beasiswa PIP. Uji sampel yang dilakukan kepada siswa sebanyak 289 yang dapat diimplementasikan secara baik dan cluster berjalan dengan sesuai ketentuan yang telah ditetapkan sebagai indikator 10 atribut (X1-X10).

Ini adalah artikel akses terbuka di bawah lisensi [CC-BY-SA](#).



1. Pendahuluan

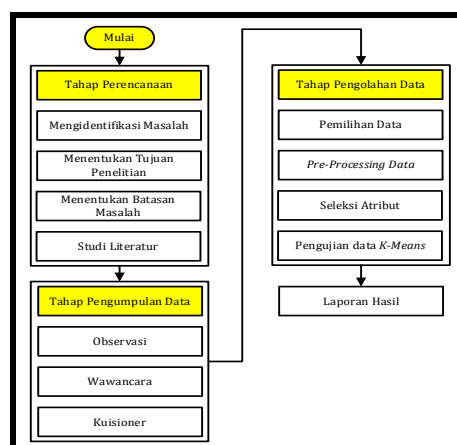
Pendidikan yang menjadi tanggung jawab semua aspek mulai dari dasar meliputi keluarga hingga ke pemerintah ikut andil dan berkontribusi dalam memajukan dan mengembangkan pendidikan untuk warga negara agar memiliki karakteristik, keunikan, berkehidupan sosial yang sehat dan berakal. Dapat dikatakan bahwa pendidikan adalah kunci dalam membentuk sumber daya manusia yang berkualitas. Dalam hal ini pemerintah telah melaksanakan program beasiswa (Program Indonesia Pintar) PIP untuk pendidikan dasar hingga tamat SMA namun sayangnya program beasiswa seringkali salah target dalam menentukan siswa yang berhak mendapat beasiswa karena itu dalam menentukan kelayakan penerima KIP agar sesuai dengan program pemerintah yang diadakan adalah menentukan siapa yang benar-benar layak mendapatkan KIP, agar program PIP yang di jalankan oleh pemerintah itu sesuai dengan sasaran dan kriteria dalam pemberiannya.

Dalam menyelesaikan masalah ini dengan metode *K-Means Clustering* dapat membantu menentukan golongan calon penerima KIP dan golongan yang tidak menerima KIP. Dengan meng *cluster*-kan variabel-variabel data yang di ambil dari data siswa yang ada. Variabel yang di gunakan adalah variabel tentang penghasilan orangtua, pekerjaan orangtua, jumlah saudara yang dimiliki dan status yatim/piatu ataukah masih memiliki orangtua lengkap yang menggunakan metode algoritma *K-Means Clustering*.

Terdapat beberapa penelitian yang terkait dalam pengembangan *clustering* menggunakan *K-Means Clustering* dengan seperti [1] yang mengungkap menentukan penerima beasiswa Yayasan untuk mahasiswa dengan metode data mining Algoritma *K-Means Clustering*. Kemudian ada juga penelitian yang menyinggung *clustering* dari [2] menurut beliau *clustering* dapat menentukan prediksi penerima BLT dan desa. Namun dari penelitian dahulu *clustering* yang dilakukan adalah belum *realtime* dan hanya mengambil data dan diolah sebagai mana data mining yang telah diadaptkan dari data sebelumnya.

Clustering atau klasterisasi adalah metode pengelompokan data. Menurut [3] *clustering* adalah sebuah proses untuk mengelompokkan data ke dalam beberapa *cluster* atau kelompok sehingga data dalam satu *cluster* memiliki tingkat kemiripan yang maksimum dan data antar *cluster* memiliki kemiripan yang minimum. *K-Means* merupakan salah satu algoritma yang bersifat *unsupervised learning*. *K-Means* memiliki fungsi untuk mengelompokkan data kedalam data *cluster*. Algoritma ini dapat menerima data tanpa ada label kategori. *K-Means Clustering* Algoritma juga merupakan metode *non-hierarchy* [4]. *Clustering* pada permasalahan pendidikan juga dapat dilakukan dalam penentuan penerima beasiswa. Pada penelitian sebelumnya, terdapat rekomendasi penentuan mahasiswa penerima beasiswa dengan menerapkan metode *Naive Bayes* [5]. Metode *Naive Bayes* merupakan metode probabilistik pengklasifikasian dilakukan melalui *training set* sejumlah data secara efisien. Akan tetapi, pada penelitian tersebut, hasil tingkat akurasi data algoritma *Naive Bayes* angka diperoleh kurang maksimal [6]. Sedangkan penelitian yang membahas tentang klasifikasi siswa yang berhak menerima beasiswa di UN PGRI Kediri diperoleh hasil bahwa algoritma *K-Means Clustering* dapat membantu dalam mengklasifikasi siswa yang berhak menerima beasiswa, siswa yang dipertimbangkan menerima, dan siswa yang tidak berhak menerima dengan kriteria yang telah ditentukan [7]. Algoritma *K-Means Clustering* mampu membantu proses menentukan siswa yang mendapatkan beasiswa secara otomatis, sehingga proses pelaporan menjadi lebih efektif dan efisien [8]. Metode *K-Means* berusaha mengelompokkan data yang ada kedalam beberapa kelompok, dimana data dalam satu kelompok mempunyai karakteristik yang sama satu sama lainnya dan mempunyai karakteristik yang berbeda dengan data yang ada di dalam kelompok yang lain [9]. Maka kita akan mengetahui hasil dari klasterisasi beasiswa KIP Kuliah mahasiswa menggunakan metode *K-Means Clustering*. Berbeda dengan penelitian yang dilakukan oleh [10] pada penelitian ini mengambil permasalahan penentuan penerima beasiswa KIP di SMP N 9 Kota Blitar.

2. Metode penelitian



Gambar 1. Tahapan Penelitian

Tahapan metodologi atau alur penelitian yang digunakan adalah secara garis besar dibagi menjadi 3 blok yaitu: tahap perencanaan, tahap pengumpulan data, dan tahap pengolahan data.

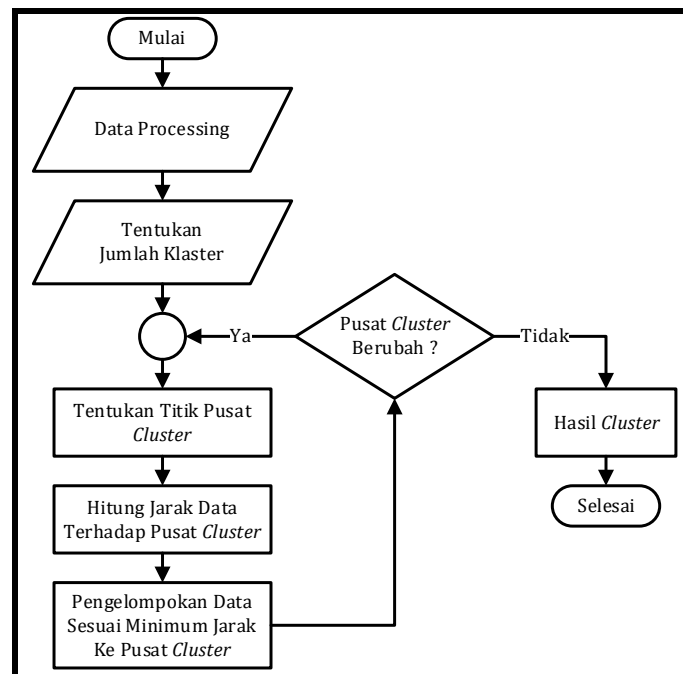
2.1 Lokasi dan Waktu Penelitian

Penelitian ini berlokasi di salah satu sekolah menengah pertama di kota Blitar yaitu UPT Satuan Pendidikan SMPN 9 Blitar, yang beralamat di Jalan Widias No 29 Kecamatan Sukorejo Kota Blitar. Waktu penelitian dilakukan selama 3 bulan Februari 2023 hingga bulan Mei 2023.

2.2 Metode Pengumpulan Data

Pengumpulan data dilakukan dengan cara kegiatan observasi, wawancara kepada pihak pemangku kepentingan yang berkaitan dengan perolehan dana hibah atau beasiswa PIP, dan pengisian kuesioner yang dilakukan oleh wali murid dari siswa SMPN 9 Kota Blitar yang telah dibagikan oleh pihak sekolah dengan menjawab sejumlah pertanyaan.

2.3 Perancangan Algoritma *K-means Clustering*



Gambar 2. Flowchart Algoritma *K-Means Clustering*

Data mining yang telah didapat dari proses pengumpulan data, kemudian diolah dengan menggunakan algoritma *K-Means Clustering* dengan perhitungan *Euclidean distance*.

$$d(x_i, \mu_j) = \sqrt{\sum (x_i - \mu_j)^2} \quad (1)$$

Keterangan:

$d(x_i, \mu_j)$ = Jarak di antara x dan j

j = nilai pembolehubah i bagi x

x_i = nilai pembolehubah i bagi j

Sesuai dengan atribut dan ketentuan yang telah ditetapkan $K=2$. Dan atribut sebanyak $X=10$ mulai dari X_1 hingga X_{10} yang mewakili indikasi untuk siswa yang berhak menerima beasiswa PIP.

2.4 Pemilihan Atribut

Tabel 1. Pemilihan Atribut

No.	Data Awal	Atribut
1.	Jenis tempat tinggal	X1
2.	Penghasilan Orang Tua	X2
3.	Pekerjaan Orang Tua	X3
4.	Jumlah Saudara	X4
5.	Kepemilikan Transportasi	X5
6.	Transportasi ke Sekolah	X6
7.	Memiliki KIS	X7
8.	Memiliki KKS	X8
9.	Memiliki PKH	X9
10.	Jenis Surat Keterangan	X10

Seleksi ini untuk menentukan atribut mana saja yang digunakan dalam suatu penelitian, maka dari itu diperlukan tahap seleksi atribut agar tahu atribut-atribut mana saja yang relevan dan memiliki hubungan antar sesama atribut.

3. Hasil dan Analisis

Pada hasil dan analisis meliputi hasil yang sudah dilakukan yaitu mengolah data menggunakan metode algoritma *K-Means clustering* dengan perhitungan nilai jarak terdekat *Euclidean distance* dimana data yang sudah diperoleh melalui observasi, wawancara, dan pengisian kuesioner melalui web yang sudah tersedia secara *realtime*.

3.1 Proses pembersihan dan transformasi data

Pada proses pembersihan/*cleaning* data adalah proses untuk menghilangkan data yang kosong, data yang tidak dibutuhkan atau tidak relevan (*noise*) dan data (*outlier*) yang tidak memiliki keterkaitan baik secara langsung ataupun tidak langsung terhadap penelitian. Pada penelitian ini semua data yang digunakan adalah *numeric* jadi data yang berbentuk *polynomial* harus ditransformasikan kedalam bentuk *numeric*.

3.2 Hasil *K-Means Clustering*

Hasil dari *k-means clustering* dibagi dari jumlah *k* yang telah ditentukan dimana terdapat 2 dari *k*, *clustering* 1 atau C1 yaitu cluster yang layak menerima beasiswa PIP sedangkan untuk *cluster* 2 atau C2 untuk siswa yang tidak layak menerima beasiswa PIP. Setelah penentuan *k* menentukan *centroid* yang akan dijadikan patokan (acuan) nilai beasiswa yang layak menerima PIP

Tabel 2. *Centroid* Awal

Centroid	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10
C1	2	1	2	5	2	1	3	3	1	1
C2	2	5	6	4	4	3	3	3	3	7

Centroid awal merupakan nilai acuan yang digunakan untuk pengelompokan *clustering* C1 dan C2 menyesuaikan dengan ketentuan yang sudah dilakukan dalam transformasi data.

Tabel 3. *Centroid* Baru

Centroid	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10
C1	2,03	3,07	2,95	5,13	2,88	1,91	2,86	2,56	1,25	1,22
C2	2,39	4,89	5,99	4,74	4,05	3,85	2,96	2,77	2,76	5,79

Centroid baru terbentuk ketika sudah melakukan iterasi atau percobaan pertama yang menghasilkan nilai-nilai baru dalam *clustering* C1 maupun C2. Berikut perbandingan hasil dari C1 dan C2 pada Tabel 4 dibawah ini.

Tabel 4. Perbandingan iterasi ke-1 dan ke-2

No.	C1	C2	Cluster	C1	C2	Cluster
130.	6,32	6,24	2	4,77	4,96	1
131.	2,65	8,49	1	2,04	7,70	1
132.	10,82	3,16	2	8,94	2,84	2
133.	2,65	7,48	1	0,61	6,59	1
134.	9,64	2,00	2	7,86	1,90	2
135.	3,87	6,93	1	2,14	5,95	1
136.	6,63	5,00	2	4,46	3,51	2
137.	3,87	6,93	1	2,14	5,95	1
138.	2,65	8,49	1	2,04	7,70	1
139.	3,00	8,37	1	2,08	7,62	1
140.	2,24	8,12	1	1,45	7,42	1
141.	9,43	1,41	2	7,55	1,93	2
142.	9,11	2,00	2	7,39	2,74	2
143.	3,16	7,14	1	1,21	6,20	1
144.	10,10	2,24	2	8,16	2,20	2
145.	11,00	3,74	2	9,15	3,24	2
146.	0,00	9,00	1	2,68	8,35	1
147.	5,39	4,69	2	3,21	3,58	1
148.	2,65	8,49	1	2,04	7,70	1
149.	6,32	6,24	2	2,13	4,67	1

Memperlihatkan hasil dari *clustering* iterasi ke 1 dan ke 2 mengalami perubahan data dan nilai setelah nilai *centroid* baru juga telah terbentuk. Di dalam iterasi ke 2 terdapat perbedaan nilai yang berubah (pada warna merah) dimana perubahan *clustering* yang sebelumnya masuk *cluster* 2 menjadi *cluster* 1. Proses iterasi ini berlanjut hingga iterasi ke-n sampai nilai *clustering* tidak berubah terhadap iterasi sebelumnya.

4. Kesimpulan

Pemanfaatan algoritma *K-Means* dalam penentuan beasiswa PIP dapat digunakan dan memberikan keputusan terhadap pemberi kebijakan secara nyata dan adil sesuai dengan indikator atau atribut yang telah ditentukan. Berdasarkan *clustering* yang dilakukan data perolehan beasiswa PIP di salah satu sekolah SMPN 9 Blitar atau masuk ke dalam *Cluster* 1 sebanyak 198 siswa. Sehingga dengan informasi yang didapat sebagai pembuat keputusan dapat memperhatikan unsur-unsur yang mendetail dan sebagai bahan pertimbangan untuk pengajuan proses beasiswa PIP pada siswa SMPN 9 Blitar.

Peluang penelitian yang dapat dilakukan adalah menganalisa jumlah *clustering* untuk kemudian diprediksi dari pola menggunakan algoritma *K-Means* dan dapat mempertajam keputusan karena terdapat pola sebelum keputusan dibuat.

Pengakuan dan Penghargaan

Terimakasih kepada pihak SMPN 9 Blitar atas kerjasama dan prasana untuk keberlangsungan penelitian, farid dan arta sebagai penulis kedua dan ketiga dalam kontribusi terhadap penelitian ini.

Referensi

- [1] Devi Tri Yuliana, M. Ivan Ariful Fathoni, Naning Kurniawati. (2022). Penentuan Penerima Kartu Indonesia Pintar KIP Kuliah dengan Menggunakan Metode *K-Means Clustering*.
- [2] Rini Gustini, RZ. Abdul Aziz. (2019). Pengembangan Model Pengambilan Keputusan Penerima Kartu Indonesia Pintar (KIP) Dengan Metode *K-Means* Dan *Average Linkage Clustering* (Studi Kasus : SMA Negeri 1 Kotagajah).
- [3] Yosep Filki. (2022). Algoritma *K-Means Clustering* dalam Memprediksi Penerima Bantuan Langsung Tunai (BLT) Dana Desa.
- [4] Darlinda, Joy Nashar Utamajaya. (2022). Sistem Pendukung Keputusan Penerima Beasiswa Program Indonesia Pintar Menggunakan Metode Algoritma *K-Means Clustering*.
- [5] Hartama, D., & Anjelita, M. (2022). *Analysis of Silhouette Coefficient Evaluation with Euclidean Distance in the Clustering Method (Case Study: Number of Public Schools in Indonesia)*. Jurnal Mantik, 6(3), 3667-3677.
- [6] Lubis, A. R., & Lubis, M. (2020). *Optimization of distance formula in K-Nearest Neighbor method*. Bulletin of Electrical Engineering and Informatics, 9(1), 326-338.
- [7] Karo, I. M. K., Khosuri, A., & Setiawan, R. (2021, October). *Effects of Distance Measurement Methods in K-Nearest Neighbor Algorithm to Select Indonesia Smart Card Recipient*. In *2021 International Conference on Data Science and Its Applications (ICoDSA)* (pp. 209-214). IEEE.
- [8] Fammaldo, E., & Hakim, L. (2018). Penerapan Algoritma *K-Means Clustering* Untuk Pengelompokan Tingkat Kesejahteraan Keluarga Untuk Program Kartu Indonesia Pintar. Jurnal Ilmiah Teknologi Infomasi Terapan, 5(1), 23-31.
- [9] Sompa, M., & Ishak, R. (2022). *Clustering* Tingkat Ekonomi Mahasiswa Calon Penerima Kartu Indonesia Pintar (KIP) Kuliah Metode *K-Means*. Jurnal Ilmiah Ilmu Komputer Banthayo Lo Komputer, 1(2), 65-71.
- [10] Nuraeni, F., Kurniadi, D., & Dermawan, G. F. (2023). Pemetaan Karakteristik Mahasiswa Penerima Kartu Indonesia Pintar Kuliah (KIP-K) menggunakan Algoritma *K-Means++*. Jurnal Sisfokom (Sistem Informasi Dan Komputer), 11(3), 437-443.



Rizki Dwi Romadhona lahir di Malang, 30 Januari 1996. Memperoleh gelar Sarjana Sains Terapan dari Politeknik Negeri Malang pada 2018 dan Magister Terapan Teknik dari Politeknik Negeri Malang pada 2021. Saat ini bekerja di Universitas Islam Balitar Fakultas Teknologi dan Informasi sebagai dosen dan peneliti dengan bidang keilmuan *Internet of Things* dan Sistem Kendali Cerdas, *Computer Science Mobile app* dan *Cloud Computing*.

Alamat Email: rizkidwiromadhona@unisbablitar.ac.id