



IMPLEMENTASI ALGORITMA DBSCAN UNTUK *CLUSTERING* SELEKSI PENENTUAN MAHASISWA YANG BERHAK MENERIMA BEASISWA YAYASAN

Dwina Pri Indini¹, Sartika R. Siburian², Nurhasanah³, Dito Putro Utomo^{4*}, Mesran⁵

^{1,2,3,4,5}Program Studi Teknik Informatika, Universitas Budi Darma, Medan

e-mail: ¹dwinapriindini@gmail.com, ²sartikasiburian00@gmail.com,

³nurhasanahhana322@gmail.com, ^{4,*}ditoputro12@gmail.com,

⁵mesran.skompom@gmail.com

*Corresponding Author

Abstrak

Beasiswa yayasan merupakan beasiswa yang diberikan kepada mahasiswa untuk membantu masalah finansial sehingga mahasiswa dapat melanjutkan pendidikan tidak terhalang dengan keadaan ekonomi. Proses seleksi pemberian beasiswa sering terjadi kesalahan dikarenakan sistem seleksi dilakukan secara manual sehingga tidak efektif dan efisien. Data mining dapat digunakan dalam pengelompokan mahasiswa yang berhak menerima beasiswa yayasan. Dalam pengelompokan tersebut harus memenuhi kriteria IPK, Prestasi Non-Akademik, Jumlah Tanggungan Orang Tua dan Penghasilan Orang Tua. Maka dari itu dibutuhkan penerapan data mining untuk mengatasi masalah tersebut menggunakan algoritma DBSCAN yang dapat melakukan pengelompokan lebih akurat dengan kepadatan yang berbeda. Sehingga menghasilkan dua cluster yang dapat dibandingkan, dengan demikian mahasiswa yang berhak menerima bantuan yayasan terletak pada pada *cluster* I yaitu alternatif D, F, G, H, I dan J.

Kata kunci—Beasiswa, Yayasan, Mahasiswa, Data Mining, DBSCAN

Abstract

Foundation scholarships are scholarships given to students to help with financial problems so that students can continue their education without being hindered by economic conditions. The selection process for scholarships often occurs because the selection system is done manually so it is not effective and efficient. Data mining can be used in grouping students who are entitled to receive foundation scholarships. The grouping must meet the criteria for GPA, Non-Academic Achievement, Number of Dependent Parents and Parents' Income. Therefore, it is necessary to apply data mining to overcome these problems using the DBSCAN algorithm which can perform more accurate grouping with different densities. So as to produce two comparable clusters, thus students who are entitled to receive foundation assistance are located in cluster I, namely alternatives D, F, G, H, I and J.

Keywords—Scholarship, Foundation, Student, Data Mining, DBSCAN



I. PENDAHULUAN

Pendidikan merupakan sesuatu yang penting untuk diperhatikan saat ini karena pendidikan dapat membentuk kepribadian yang baik sesuai standar normatif melalui pendidikan. Beasiswa merupakan salah satu bentuk dukungan finansial atau bantuan dana yang dialokasikan kepada individu untuk melanjutkan pendidikannya agar dapat dilanjutkan dengan tujuan untuk meringankan beban selama menempuh pendidikan, terutama dalam keadaan sulit sehingga untuk mempermudah penyelesaian kegiatan belajar[1]. Beasiswa yayasan merupakan salah satu jenis beasiswa yang disediakan oleh pihak Universitas Budi Darma, beasiswa ini diajukan ke Yayasan yang akan diberikan kepada mahasiswa yang belum mendapatkan beasiswa. Keputusan pemberi beasiswa terdapat pada yayasan untuk memastikan bahwa beasiswa tersebut diberikan kepada orang yang tepat.

Proses seleksi pemberian beasiswa sering menjadi permasalahan dikarenakan sering terjadinya kesalahan dalam proses pengajuan dalam merekomendasikan mahasiswa yang tidak seharusnya menerima beasiswa dikarenakan jumlah mahasiswa yang banyak dan proses dilakukan secara manual, maka proses dinyatakan tidak efektif dan efisien karena membutuhkan waktu yang lama. Dalam proses *clustering* seleksi registrasi beasiswa yayasan mempunyai persyaratan-persyaratan yang wajib dipenuhi mahasiswa sebelum menerimanya. Kriteria persyaratannya antara lain yaitu IPK, Prestasi Non-Akademik, Jumlah Tanggungan Orang Tua dan Pekerjaan Orang tua. Karena itu diperlukan suatu sistem yang bisa memfasilitasi untuk melakukan pengelompokan data secara komputerisasi pada proses *clustering* seleksi terhadap data mahasiswa dalam penentuan mahasiswa yang berhak menerima beasiswa yayasan dengan tujuan

memperoleh data dan hasil yang lebih akurat dan tepat berdasarkan nilai-nilai yang dimiliki calon penerima beasiswa yang tersedia dalam penerimaan beasiswa yayasan yaitu data mining.

Data mining diperlukan untuk mengatasi dan menyelesaikan masalah tersebut, data mining dapat diartikan sebagai kegiatan yang berkaitan dengan pengumpulan data yang bertujuan untuk mendapatkan informasi, pola atau hubungan dalam *big data* yang bisa dimanfaatkan sebagai alternatif untuk mengambil sebuah keputusan[2]. Penelitian ini menggunakan algoritma *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN) merupakan algoritma yang dapat mendeteksi *outlier* atau *noise* dan menghasilkan *cluster* lebih akurat dan baik untuk data dalam jumlah besar serta tidak perlu menentukan jumlah *cluster* di awal. Algoritma ini lebih efektif dan lebih baik dalam menentukan parameter daripada algoritma *Dynamic Method Density Based Spatial Clustering of Application with Noise* (DMDBSCAN) dan memiliki kemampuan yang berbeda dari algoritma K-Means dan K-Medoids[3].

Berdasarkan penelitian sebelumnya yang berkaitan dengan kesamaan algoritma, penulis menjadikannya sebagai rujukan dalam penelitian ini. Penelitian yang dilakukan oleh Nana Nurhaliza dan Mustakim tahun 2021 membahas mengenai algoritma DBSCAN dalam mengelompokkan data kasus Covid-19 di Dunia yang dilakukan sebanyak 22 percobaan, maka dihasilkan 3 *cluster* dengan nilai *Eps* 0.2 dan *MinPts* 3[4]. Penelitian yang dilakukan Raisa Rizky, dkk tahun 2021 membahas pengelompokan data gempa bumi dengan menerapkan algoritma DBSCAN, hasil yang diperoleh dari penelitian yaitu 3 *cluster* dengan nilai *Eps* 0.28 dan *MinPts* 3 sebagai wilayah yang beresiko mengalami gempa bumi

berlandaskan karakteristik parameter gempa bumi dan menyimpulkan bahwa DBSCAN mampu mengelompokkan dengan cukup baik dan dapat dijadikan sebagai bahan pertimbangan dalam mengambil keputusan [5].

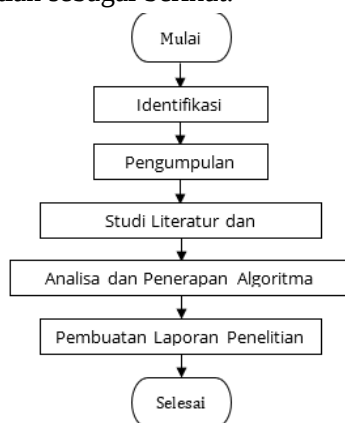
II. METODOLOGI PENELITIAN

2.1 Data Mining

Data mining merupakan teknik *machine learning* yang menggunakan teknik dan algoritma tertentu untuk secara otomatis menganalisis dan mengekstrak pengetahuan untuk menemukan pola dan informasi menarik yang ingin diketahui. Data-mining dianggap sebagai cabang integral dari ilmu komputer yang dapat digunakan untuk kumpulan data besar untuk tujuan mengekstraksi pengetahuan yang berarti secara otomatis atau semi-otomatis. Data mining ada sejak 1990-an untuk menemukan pola dan informasi yang dapat digunakan untuk mengelompokkan menjadi satu atau lebih *cluster* [6], [7].

2.2 Tahapan Penelitian

Beberapa urutan tahapan untuk menghasilkan data yang diperlukan dalam penelitian sebagai berikut.



Gambar 1. Tahapan Penelitian

2.3 DBSCAN

DBSCAN merupakan algoritma *clustering* yang dapat menemukan area

berdasarkan kerapatan yang tinggi untuk membentuk *cluster*. *Cluster* pada algoritma DBSCAN didefinisikan sebagai area yang mempunyai sampel yang padat atau kerapatan yang tinggi dan dibedakan dari area kerapatan yang rendah (noise) [8], [9]. Urutan algoritma DBSCAN memiliki 5 langkah secara umum yaitu [10]:

1. Tentukan parameter *Eps* dan *MinPts*.
2. Tentukan titik awal (*p*) secara *random*.
3. Lakukan perhitungan jarak *p* dengan seluruh titik yang lain.
Perhitungan jarak dengan rumus *Euclidean Distance*.
$$\text{Jarak} = \sqrt{(x - x_p)^2 + (y - y_p)^2}$$
4. Ambil semua titik yang *density-reachable* dari *p* yang memenuhi *Eps* dan *MinPts*. Jika *p* adalah *core point* maka *cluster* terpenuhi.
5. Ulangi langkah 3 dan 4 sampai semua poin telah diproses. Jika *p* adalah *border point* dan tidak ada yang memiliki hubungan *density-reachable* dengan *p* maka selanjutnya akan memasuki titik lainnya dalam database.



III. HASIL DAN PEMBAHASAN

3.1 Analisa Masalah

Dalam membentuk *cluster* untuk seleksi mahasiswa yang berhak menerima beasiswa yayasan maka dalam penelitian ini menggunakan data-data mahasiswa yang akan mengikuti seleksi calon mahasiswa penerima beasiswa yayasan pada universitas Budi Darma. Pada penelitian ini hanya menggunakan 10 data mahasiswa angkatan tahun 2019. Berikut merupakan tabel hasil data *Preprocessing* data mahasiswa.

Tabel 1. Hasil Data *Preprocessing*

No	Titik	IPK	Prestasi Non-akademik	Jumlah Tanggungan Orang Tua	Penghasilan Orang Tua (Rupiah)
1	A	10	2	5	2
2	B	9	2	4	5
3	C	7	2	6	3
4	D	5	1	3	1
5	E	8	2	4	4
6	F	6	1	2	1
7	G	4	1	3	2
8	H	2	1	1	2
9	I	1	1	2	2
10	J	3	1	3	1

3.2 Penerapan Algoritma DBSCAN

Langkah-langkah yang dibutuhkan dalam seleksi penentuan mahasiswa yang berhak menerima beasiswa yayasan menggunakan algoritma DBSCAN sebagai berikut.

1. Tentukan parameter *Eps* dan *MinPts*.

$Eps = 4$ dan $MinPts = 4$.

Iterasi I

Pada iterasi I titik pusat awal ditentukan secara acak dengan demikian titik pusat adalah titik D dengan $w_p=5$, $x_p=1$, $y_p=3$, $z_p=1$. Berikut merupakan perhitungan jarak setiap titik dengan titik pusat.

AD

=

$$\sqrt{(10-5)^2 + (2-1)^2 + (5-3)^2 + (2-1)^2} = 5.567764363$$

BD

=

$$\sqrt{(9-5)^2 + (2-1)^2 + (4-3)^2 + (5-1)^2} = 5.830951895$$

CD

=

$$\sqrt{(7-5)^2 + (2-1)^2 + (6-3)^2 + (3-1)^2} = 4.242640687$$

DD

=

$$\sqrt{(5-5)^2 + (1-1)^2 + (3-3)^2 + (1-1)^2} = 0$$

Setelah proses perhitungan maka dapat dilihat dalam bentuk tabel.

Tabel 2. Data Jarak Iterasi I

Jarak	Hasil
AD	5.567764363
BD	5.830951895
CD	4.242640687
DD	0
ED	4.472135955
FD	1.414213562
GD	1.414213562
HD	3.741657387
ID	4.242640687
JD	2

Density reachable diperoleh jika titik termasuk dalam radius dengan ketentuan nilai jarak kurang atau sama dengan *Eps* yang sudah ditetapkan. Berikut tabel titik terseleksi dengan ketentuan nilai $Eps = 4$.

Tabel 3. Titik Terseleksi dengan Ketentuan Nilai *Eps* Iterasi I

Titik	IPK	Prestasi Non-akademik	Jumlah Tanggungan Orang Tua	Penghasilan Orang Tua (Rupiah)	Jarak Ke Titik D
D	5	1	3	1	0
F	6	1	2	1	1.414213562
G	4	1	3	2	1.414213562
H	2	1	1	2	3.741657387
J	3	1	3	1	2

Neighborhood Core Object tercapai lantaran total *e-neighborhood* telah melengkapi syarat karena total titik lebih dari sama dengan $MinPts = 4$. Titik selanjutnya dipilih berdasarkan jarak terjauh dari *core object* yaitu titik H.

Iterasi II

Pada iterasi II titik pusat didapat dari iterasi I yaitu titik H dengan $w_p=2$, $x_p=1$, $y_p=1$, $z_p=2$. Berikut tabel hasil perhitungan jarak setiap titik dengan titik pusat pada iterasi II.

**Tabel 4.** Data Jarak Iterasi II

Jarak	Hasil
AH	9
BH	8.246211251
CH	7.211102551
DH	3.741657387
EH	7.071067812
FH	4.242640687
GH	2.828427125
HH	0
IH	1.414213562
JH	2.449489743

Tabel 5. Titik Terseleksi dengan Ketentuan Nilai Eps iterasi II

Titik	IP K	Prestasi Non-akademik	Jumlah Tanggungan Orang Tua	Penghasilan Orang Tua (Rupiah)	Jarak Ke Titik I
D	5	1	3	1	3.741657387
G	4	1	3	2	2.828427125
H	2	1	1	2	0
I	1	1	2	2	1.414213562
J	3	1	3	1	2.449489743

Dapat dilihat titik D adalah titik terjauh akan tetapi tidak dapat dipilih karena sudah pernah dikunjungi dan titik terjauh selanjutnya yaitu G dan J tidak dapat dipilih karena termasuk dalam *border point* di *core object* pada iterasi sebelumnya, maka titik yang dapat dipilih yaitu titik I.

Iterasi III

Pada iterasi III titik pusat didapat dari iterasi II yaitu titik I dengan $w_p=1$, $x_p=1$, $y_p=2$, $z_p=2$. Berikut tabel hasil perhitungan jarak setiap titik dengan titik pusat pada iterasi III.

Tabel 6. Data Jarak Iterasi III

Jarak	Hasil
AI	9.539392014
BI	8.831760866
CI	7.348469228
DI	4.242640687
EI	7.615773106
FI	5.099019514

GI	3.162277660
HI	1.414213562
II	0
JI	2.449489743

Tabel 7. Titik Terseleksi dengan Ketentuan Nilai Eps iterasi III

Titik	IP K	Prestasi Non-akademik	Jumlah Tanggungan Orang Tua	Penghasilan Orang Tua (Rupiah)	Jarak Ke Titik H
G	4	1	3	2	3.162277660
H	2	1	1	2	1.414213562
I	1	1	2	2	0
J	3	1	3	1	2.449489743

Dari hasil tersebut tidak ada lagi titik yang dapat dijadikan *core object*. Maka akan ditentukan titik lainnya dalam database.

Iterasi IV

Pada iterasi IV titik pusat ditentukan lagi secara acak yaitu titik A dengan $w_p=10$, $x_p=2$, $y_p=5$, $z_p=2$. Berikut tabel hasil perhitungan jarak setiap titik dengan titik pusat pada iterasi IV.

Tabel 8. Data Jarak Iterasi IV

Jarak	Hasil
AA	0
BA	3.316624790
CA	3.316624790
DA	5.567764363
EA	3
FA	5.196152423
GA	6.403124237
HA	9
IA	9.539392014
JA	7.416198487

Tabel 9. Titik Terseleksi dengan Ketentuan Nilai Eps iterasi IV

Titik	IP K	Prestasi Non-akademik	Jumlah Tanggungan Orang Tua	Penghasilan Orang Tua (Rupiah)	Jarak Ke Titik H
A	10	2	5	2	0
B	9	2	4	5	3.316624790
C	7	2	6	3	3.316624790
E	8	2	4	4	3

**Iterasi V**

Pada iterasi V titik pusat didapat dari iterasi IV yaitu titik B dengan $w_p=9$, $x_p=2$, $y_p=4$, $z_p=5$. Berikut tabel hasil perhitungan jarak setiap titik dengan titik pusat pada iterasi V.

Tabel 10. Data Jarak Iterasi V

Jarak	Hasil
AB	3.316624790
BB	0
CB	3.464101615
DB	5.830951895
EB	1.414213562
FB	5.477225575
GB	6
HB	8.246211251
IB	8.831760866
JB	7.348469228

Tabel 11. Titik Terseleksi dengan Ketentuan Nilai Eps iterasi V

Titik	IP K	Prestasi Non-akademik	Jumlah Tanggungan Orang Tua	Penghasilan Orang Tua (Rupiah)	Jarak Ke Titik H
A	10	2	5	2	3.316624790
B	9	2	4	5	0
C	7	2	6	3	3.464101615
E	8	2	4	4	1.414213562

Dari hasil tersebut karena *density reachable* yang didapat sama seperti iterasi sebelumnya, maka tidak ada titik yang dapat dipilih kembali dan proses selesai. Berikut tabel hasil iterasi I sampai V.

Tabel 12. Data Iterasi I sampai iterasi III

Iterasi I	Iterasi II	Iterasi III	Iterasi IV	Iterasi V
D	D	G	A	A
F	G	H	B	B
G	H	I	C	C
H	I	J	E	E
J	J			

Berdasarkan tabel dapat dilihat seluruh *core point* yaitu D, H, I, A, B. Titik D, H, dan I saling berkaitan maka dianggap satu *cluster*. Syarat dapat dinyatakan saling berkaitan yaitu ketika poin inti dari awal sampai akhir memiliki kaitan dengan

density-connected. Tidak ada data noise karena seluruh titik masuk kedalam bagian *cluster*. Dari seluruh perhitungan diatas dengan menggunakan algoritma DBSCAN maka dihasilkan *cluster* seperti berikut.

Tabel 13. Data Cluster

Cluster I	Cluster II
D	A
F	B
G	C
H	E
I	
J	

Dari kedua cluster dapat dibandingkan, maka dapat ditentukan mahasiswa yang berhak menerima beasiswa yayasan berada pada *cluster* I yaitu alternatif D, F, G, H, I dan J.



IV. KESIMPULAN

Berlandaskan pada penelitian yang dilakukan maka dapat disimpulkan algoritma DBSCAN dapat dioperasikan dalam pembentukan *cluster* dari titik-titik terpilih dengan tingkat kerapatan yang tinggi, sehingga menghasilkan dua *cluster* yang dapat dibandingkan dalam menentukan mahasiswa yang berhak menerima beasiswa yayasan berada pada *cluster* I yaitu alternatif D, F, G, H, I dan J. Dengan demikian hasil yang diperoleh dapat digunakan sebagai rekomendasi untuk memudahkan proses seleksi agar lebih mudah dan cepat.

V. SARAN

Metode DBSCAN dapat dipergunakan untuk proses penyelesaian dengan data yang lebih besar untuk proses clustering.

**VI. DAFTAR PUSTAKA**

- [1] M. I. Bachtiar, H. Suyono, and M. F. E. Purnomo, "METHOD COMPARISON IN THE DECISION SUPPORT SYSTEM," vol. 11, no. 2, pp. 75–82, 2021.
- [2] B. G. Sudarsono and S. P. Lestari, "Clustering Penerima Beasiswa Yayasan Untuk Mahasiswa Menggunakan Metode K-Means," *J. Media Inform. Budidarma*, vol. 5, no. 1, p. 258, 2021, doi: 10.30865/mib.v5i1.2670.
- [3] D. Algorithm, "Journal of Physics and Its Applications Clustering of Seismicity in the Indonesian Region for the 2018-2020 Period using the," vol. 4, no. 1, pp. 1–6, 2021.
- [4] M. Nana Nurhaliza, "Clustering of Data Covid-19 Cases in the World Using DBSCAN Algorithms Pengelompokan Data Kasus Covid-19 di Dunia Menggunakan Algoritma," *Indones. J. Inform. Res. Softw. Eng.*, vol. 1, no. 1, pp. 1–8, 2021.
- [5] R. R. A. Rahman and A. W. Wijayanto, "Pengelompokan Data Gempa Bumi Menggunakan Grouping Earthquakes Data Using," pp. 31–38, 2021.
- [6] J. Informatika, S. A. Rahmah, J. Antares, and U. Dharmawangsa, "Beasiswa Yayasan Menggunakan K-Means," vol. 13, no. 2, pp. 25–30, 2021.
- [7] Z. Chen, P. Xu, F. Feng, Y. Qiao, and W. Luo, "Data mining algorithm and framework for identifying HVAC control strategies in large commercial buildings," *Build. Simul.*, vol. 14, no. 1, pp. 63–74, 2021, doi: 10.1007/s12273-019-0599-0.
- [8] D. Fitrianah, "Implementasi Algoritma DBScan dalam Pemngambilan Data Menggunakan Scatterplot," vol. 6, no. 2, pp. 91–98, 2021.
- [9] R. Mahendra, F. Azmi, and C. Setianingsih, "Klasterisasi Pada Data Penggunaan Listrik Di Gedung Telkom University Menggunakan Algoritma Density-Based Spatial Clustering of Application With Noise (Dbscan) Clustering on Electricity Usage At Telkom University Building Using Density-Based Spatial Clust," vol. 8, no. 6, pp. 12014–12023, 2021.
- [10] R. Adha, N. Nurhaliza, U. Sholeha, and M. Mustakim, "Perbandingan Algoritma DBSCAN dan K-Means Clustering untuk Pengelompokan Kasus Covid-19 di Dunia," *SITEKIN J. Sains, Teknol. dan Ind.*, vol. 18, no. 2, pp. 206–211, 2021.