

**PENGELOMPOKAN PROVINSI DI INDONESIA PERIODE
2020/2021 DENGAN METODE *DENSITY BASED SPATIAL
CLUSTERING APPLICATION WITH NOISE (DBSCAN)***

(Studi Kasus: Data Perekonomian Provinsi di Indonesia Periode 2020/2021)

SKRIPSI

Diajukan untuk Memenuhi Sebagian dari Syarat Memperoleh Gelar Sarjana Program
Studi Matematika



Oleh:

Pitria Sari

NIM.1800258

**PROGRAM STUDI MATEMATIKA
DEPARTEMEN PENDIDIKAN MATEMATIKA
FAKULTAS PENDIDIKAN MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS PENDIDIKAN INDONESIA**

2022

Pitria Sari, 2022

*PENGELOMPOKAN PROVINSI DI INDONESIA PERIODE 2020/2021 DENGAN METODE DENSITY BASED
SPATIAL CLUSTERING APPLICATION WITH NOISE (DBSCAN)*

(Studi Kasus: Data Perekonomian Provinsi di Indonesia Periode 2020/2021)

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

**PENGELOMPOKAN PROVINSI DI INDONESIA PERIODE
2020/2021 DENGAN METODE *DENSITY BASED SPATIAL
CLUSTERING APPLICATION WITH NOISE (DBSCAN)***

(Studi Kasus: Data Perekonomian Provinsi di Indonesia Periode 2020/2021)

LEMBAR HAK CIPTA

Oleh

Pitria Sari

Nim 1800258

Sebuah skripsi yang diajukan untuk memenuhi salah satu syarat memperoleh gelar
Sarjana Matematika pada Fakultas Pendidikan Matematika
dan Ilmu Pengetahuan Alam

© Pitria Sari 2021

Universitas Pendidikan Indonesia

Hak cipta dilindungi undang-undang. Skripsi ini tidak boleh diperbanyak sebagian atau
seluruhnya dengan dicetak ulang, fotokopi atau cara lainnya tanpa izin dari penulis.

LEMBAR PENGESAHAN

PITRIA SARI

**PENGELOMPOKAN PROVINSI DI INDONESIA PERIODE
2020/2021 DENGAN METODE *DENSITY BASED SPATIAL
CLUSTERING APPLICATION WITH NOISE (DBSCAN)***

(Studi Kasus: Data Perekonomian Provinsi di Indonesia Periode 2020/2021)

Disetujui dan disahkan oleh pembimbing:

Pembimbing I



Dewi Rachmatin, S.Si., M.Si

NIP. 196909291994122001

Pembimbing II

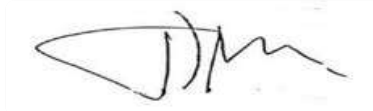


Dr. Dadan Dasari, M.Si

NIP. 196407171991021001

Mengetahui,

Ketua Departemen Pendidikan Matematika



Dr. H. Dadang Juandi, M.Si

NIP.196401171992021001

KATA PENGANTAR

Puji Syukur penulis panjatkan kepada Tuhan Yang Maha Esa yang telah melimpahkan rahmat dan karuniaNya kepada penulis, sehingga dapat Menyusun skripsi dengan judul “Pengelompokan Provinsi di Indonesia Periode 2020/2021 dengan Metode *Density Based Spatial Clustering Application with Noise* (DBSCAN)”. Skripsi ini sebagai tugas yang diajukan untuk memenuhi syarat memperoleh gelar sarjana matematika di Universitas Pendidikan Indonesia.

Skripsi ini terdiri dari lima bab. Bab I berisikan pendahuluan yang terdiri dari latar belakang masalah, rumusan masalah, tujuan penelitian, dan manfaat penelitian. Bab II berupa kajian Pustaka yang mendasari penelitian ini. Bab III berisi metodologi penelitian, variabel penelitian, pendekatan penelitian dan tahapan penelitian. Bab IV disajikan hasil penelitian serta pembahasannya. Bab V menguraikan kesimpulan dan saran.

Penulis menyadari sepenuhnya bahwa skripsi ini masih terdapat banyak kekurangan, namun penulis telah berusaha semaksimal mungkin. Oleh karena itu penulis mengharapkan saran dan kritikan yang sifatnya membangun. Demikian, semoga skripsi ini dapat bermanfaat bagi para penulis sendiri, para pembaca dan dunia Pendidikan.

Bandung, April 2022
Penulis

Pitria Sari

UCAPAN TERIMA KASIH

Penulis menyadari bahwa penyelesaian skripsi ini berkat bantuan, bimbingan, arahan dan motivasi dari berbagai pihak. Untuk itu penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada yang terhormat:

1. Dewi Rachmatin, S.Si., M.Si., selaku dosen Pembimbing 1, dan Pembimbing Akademis yang di tengah-tengah kesibukannya, telah memberikan bimbingan, arahan dan kritis terhadap berbagai permasalahan, sehingga penyusunan skripsi ini berjalan dengan baik.
2. Dr. Dadan Dasari, M.Si., selaku Pembimbing II, yang telah memberikan bimbingan, arahan dan saran kepada penulis sehingga penulis dapat menyelesaikan skripsi ini.
3. Dr. H. Dadang Djuandi, M.Si., selaku Ketua Departemen Pendidikan Matematika FPMIPA UPI Bandung.
4. Dr. Cece Kustiawan, S.Si., M.Si., selaku Ketua Prodi Matematika FPMIPA UPI Bandung.
5. Bapak dan Ibu Dosen Prodi Matematika UPI Bandung yang telah memberikan ilmu dan pengetahuan kepada penulis selama mengikuti perkuliahan kurang lebih 4 tahun ini.
6. Ayahanda Samsudin dan Ibunda Herlita Krisna teercinta atas segala bantuan, bimbingan, serta doa restu yang diberikan kepada penulis selama ini.
8. Kakak tercinta Fella Ulandari dan keluarga tersayang yang selalu memberikan arahan, semangat serta doa kepada penulis.
9. Teman-teman tercinta Sarmila Nurhasanah, Syifaul, Kiki, Shafa, Demvi, Tigin, Rian, Hilmi, dan Kiagus Muhammad Naufal yang telah membantu dan memberikan semangat setiap harinya.
10. Rekan-rekan Mahasiswa prodi Matematika Angkatan 2018, atas segala doa dan dukungannya.

11. Serta semua pihak yang telah membantu penulisan skripsi ini.

Semoga amal baik semua pihak mendapatkan balasan yang berlipat ganda dari Allah SWT. Aamiin.

Bandung, April 2022

Penulis,

Pitria Sari

**PENGELOMPOKAN PROVINSI DI INDONESIA PERIODE
2020/2021 DENGAN METODE *DENSITY BASED SPATIAL
CLUSTERING APPLICATION WITH NOISE (DBSCAN)***

(Studi Kasus: Data Perekonomian Provinsi di Indonesia Periode 2020/2021)

ABSTRAK

Analisis kluster merupakan suatu metode pengelompokan data dengan mempertimbangkan sebuah pendekatan yang berarti untuk mencari kesamaan dalam data dan membuat data yang sama menjadi suatu kelompok. Sehingga, data dalam suatu kluster memiliki tingkat kemiripan yang besar sedangkan data antar-kluster memiliki tingkat kemiripan yang kecil. Secara garis besar analisis kluster terdiri atas analisis hierarki (*hierarchial clustering*) dan analisis partisi (*partional clustering*). Jain (1988) telah mengembangkan algoritma *clustering* dengan pendekatan berdasarkan kepekatan (*density-based approach*). Algoritma *density based-approach* mampu mengidentifikasi kluster dalam bentuk apapun. Namun, algoritma tersebut memerlukan ruang penyimpanan yang besar dan membutuhkan waktu pemrosesan yang lama, sehingga Ester dkk. (1996) mengusulkan algoritma DBSCAN untuk mengatasi kelemahan tersebut. DBSCAN merupakan suatu metode *clustering* yang menentukan kluster berdasarkan kepadatan di mana kepadatan di dalam kluster akan lebih besar dibandingkan kepadatan di luar kluster. Tujuan penelitian ini adalah mengelompokkan perekonomian provinsi di Indonesia periode 2020/2021 dengan metode *density based spatial clustering application with noise*. Dengan menggunakan data perekonomian Indonesia tahun 2020/2021 didapatkan 2 kluster optimal dan 9 *noise*. Dari hasil *noise* didapatkan provinsi dengan karakteristik perekonomian yang sangat baik (ekstrem tinggi) dan sangat buruk (ekstrem rendah). Provinsi yang tergolong ekstrem rendah adalah provinsi Papua. Sehingga provinsi Papua menjadi prioritas pemulihan ekonomi paska pandemi.

Kata kunci: Analisis Kluster, DBSCAN, Indeks Validitas, Data Perekonomian.

v

**PROVINCIAL GROUPING IN INDONESIA 2020/2021 PERIOD
USING THE DENSITY BASED SPATIAL CLUSTERING
APPLICATION WITH NOISE (DBSCAN) METHOD**

(Case Study: Provincial Economic Data in Indonesia for the 2020/2021 Period)

ABSTRACT

Cluster analysis is a method of grouping data by considering a meaningful approach to look for similarities in the data and make the same data into a group. Thus, data in a cluster has a high degree of similarity while data between clusters has a small degree of similarity. Broadly speaking, cluster analysis consists of hierarchical clustering and partition analysis. Jain (1988) has developed a clustering algorithm with a density-based approach. Density based-approach algorithm is able to identify clusters in any form. However, the algorithm requires large storage space and requires a long processing time, so Ester et al. (1996) proposed the DBSCAN algorithm to overcome these weaknesses. DBSCAN is a clustering method that determines clusters based on density where the density inside the cluster will be greater than the density outside the cluster. The purpose of this study is to classify the provincial economy in Indonesia for the 2020/2021 period using the density based sp atial clustering application with noise method. By using data on the Indonesian economy for 2020/2021, 2 optimal clusters and 9 noise were obtained. From the noise results, we get provinces with very good (extremely high) and very bad (extremely low) economic characteristics. The province that is classified as extreme low is the province of Papua. Thus, the author recommends the government to make a province classified as extreme low, namely Papua province, as a priority for post-pandemic economic recovery.

Keyword: Cluster Analysis, DBSCAN, Validity Index, Economic Data.

DAFTAR ISI

LEMBAR PENGESAHAN	i
KATA PENGANTAR	ii
UCAPAN TERIMA KASIH.....	iii
ABSTRAK	v
ABSTRACT	vi
DAFTAR ISI.....	vii
DAFTAR TABEL	ix
DAFTAR GAMBAR	xi
BAB I	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	5
1.3 Tujuan Penelitian.....	5
1.4 Manfaat Penelitian.....	5
1.5 Sistematika Penulisan.....	5
BAB II.....	8
2.1 Analisis Klaster	8
2.2 DBSCAN.....	9
2.3 Standarisasi Data	10
2.4 ALGORITMA DBSCAN.....	11
2.5 Jarak Euclidian	12
2.6 Jarak Manhattan	13
2.7 Algoritma <i>K-Nearest Neighbour</i>	13
2.8 Indeks Validitas	14
2.9 Program R.....	16
2.10 Perekonomian Indonesia	17
2.11 Penelitian Terkait	19
BAB III	21
3.1 Jenis dan Sumber Data	21
3.2 Pendekatan dan Variabel Penelitian	22

3.3 Kerangka Berpikir	22
3.4 Metode Analisis Data	23
3.5 Tahapan Penelitian	23
BAB IV	27
4.1 Pemrosesan Data	27
4.2 Standarisasi Data	28
4.3 Menentukan Epsilon dan MinPts	31
4.4 Proses Pengolahan Data	41
4.5 Klaster yang Terbentuk	42
BAB V	53
5.1 Simpulan.....	53
5.2 Saran	53
DAFTAR PUSTAKA	55
Lampiran Gambar A	57
Lampiran Gambar B.....	59
Lampiran Tabel A	61
Lampiran Tabel B	62
Lampiran Tabel C	63
Lampiran Tabel D	64

DAFTAR TABEL

Tabel 3. 1 Daftar Nama 34 Provinsi yang digunakan dalam Penelitian.....	21
Tabel 4. 1 Perhitungan Standarisasi Data Tahun 2020	30
Tabel 4. 2 Perhitungan Standarisasi Data Tahun 2021	31
Tabel 4. 3 Eksperimen Jarak Euclid Tahun 2020 dengan MinPts=2	33
Tabel 4. 4 Eksperimen Jarak Euclid Tahun 2020 dengan MinPts=3	33
Tabel 4. 5 Eksperimen Jarak Euclid Tahun 2020 dengan MinPts=4	33
Tabel 4. 6 Eksperimen Jarak Euclid Tahun 2020 dengan MinPts=5	34
Tabel 4. 7 Eksperimen Jarak Euclid Tahun 2020 dengan MinPts=6	34
Tabel 4. 8 Eksperimen Jarak Euclid Tahun 2020 dengan MinPts=7	34
Tabel 4. 9 Eksperimen Jarak Manhattan Tahun 2020 dengan MinPts=2	35
Tabel 4. 10 Eksperimen Jarak Manhattan Tahun 2020 dengan MinPts=3	35
Tabel 4. 11 Eksperimen Jarak Manhattan Tahun 2020 dengan MinPts=4	36
Tabel 4. 12 Eksperimen Jarak Manhattan Tahun 2020 dengan MinPts=5	36
Tabel 4. 13 Eksperimen Jarak Manhattan Tahun 2020 dengan MinPts=6	36
Tabel 4. 14 Eksperimen Jarak Manhattan Tahun 2020 dengan MinPts=7	37
Tabel 4. 15 Eksperimen Jarak Euclid Tahun 2021 dengan MinPts=2	37
Tabel 4. 16 Eksperimen Jarak Euclid Tahun 2021 dengan MinPts=3	38
Tabel 4. 17 Eksperimen Jarak Euclid Tahun 2021 dengan MinPts=4	38
Tabel 4. 18 Eksperimen Jarak Euclid Tahun 2021 dengan MinPts=5	38
Tabel 4. 19 Eksperimen Jarak Euclid Tahun 2021 dengan MinPts=6	39
Tabel 4. 20 Eksperimen Jarak Euclid Tahun 2021 dengan MinPts=7	39
Tabel 4. 21 Eksperimen Jarak Manhattan Tahun 2021 dengan MinPts=2	39
Tabel 4. 22 Eksperimen Jarak Manhattan Tahun 2021 dengan MinPts=3	40
Tabel 4. 23 Eksperimen Jarak Manhattan Tahun 2021 dengan MinPts=4	40
Tabel 4. 24 Eksperimen Jarak Manhattan Tahun 2021 dengan MinPts=5	40
Tabel 4. 25 Eksperimen Jarak Manhattan Tahun 2021 dengan MinPts=6	41
Tabel 4. 26 Eksperimen Jarak Manhattan Tahun 2021 dengan MinPts=7	41
Tabel 4. 27 Klaster Tahun 2020 Jarak Euclidean	42

Tabel 4. 28 Nilai Variabel Noise Jarak Euclid Tahun 2020	43
Tabel 4. 29 Klaster Tahun 2020 Jarak Manhattan Eps=1,4	43
Tabel 4. 30 Nilai Variabel Noise Jarak Manhattan 2020	44
Tabel 4. 31 Tabel klaster Tahun 2020 dengan Jarak Manhattan Eps= 1,75	44
Tabel 4. 32 Nilai Variabel Noise Tahun 2020 Jarak Manhattan.....	45
Tabel 4. 33 Klaster Tahun 2021 Jarak Euclid Eps 0,85	46
Tabel 4. 34 Nilai Variabel Noise Jarak Euclid Tahun 2021	47
Tabel 4. 35 Klaster Tahun 2021 Jarak Manhattan Eps 1,49	47
Tabel 4. 36 Nilai Variabel Noise Jarak Manhattan Tahun 2021	48
Tabel 4. 37 Karakteristik Kelompok Klaster	49
Tabel 4. 38 Pengelompokan Provinsi di Indonesia tahun 2020	49
Tabel 4. 39 Pengelompokan Provinsi di Indonesia Tahun 2021	50

DAFTAR GAMBAR

Gambar 2. 1 Plot KNN.....	14
Gambar 3. 1 Kerangka Berpikir.....	22
Gambar 3. 2 Alur Penelitian.....	23
Gambar 3. 3 Flowchart DBSCAN	26
Gambar 4. 1 Data Perekonomian Tahun 2020	27
Gambar 4. 2 Data Perekonomian Tahun 2021	28
Gambar 4. 3 Plot KNN pada MinPts=2	32

DAFTAR PUSTAKA

- Babenko, V., et al. (2021). Classical machine learning methods in economics research: Macro and micro level examples. *WSEAS Transactions on Business and Economics*, 18, 209–217. <https://doi.org/10.37394/23207.2021.18.22>
- Chen, Y., et al. (2019). KNN-BLOCK DBSCAN: Fast clustering for large-scale data. *IEEE transactions on systems, man, and cybernetics: systems*, 51(6), 3939-3953.
- Ester, M., et al. (1996). Density-Based Clustering Methods. *Comprehensive Chemometrics*, 2, 635–654. <https://doi.org/10.1016/B978-044452701-1.00067-3>
- Garcia-Dias, R., et al. (2019). Clustering analysis. *Machine Learning: Methods and Applications to Brain Disorders*, 227–247. <https://doi.org/10.1016/B978-0-12-815739-8.00013-4>
- Gu, J. (2021). Comparative analysis based on clustering algorithms. *Journal of Physics: Conference Series*, 1994(1). <https://doi.org/10.1088/1742-6596/1994/1/012024>
- Hashler, M., et al. (2019, Oktober 22). *Package 'dbscan'*. Retrieved from CRAN: <http://cran.cnr.berkeley.edu/web/packages/dbscan/dbscan.pdf>
- Hennig, C. (2003) Clusters, outliers and regression: fixed point clusters, *Journal of Multivariate Analysis* 86, 183-212.
- Hennig, C. (2005) Fuzzy and Crisp Mahalanobis Fixed Point Clusters, in Baier, D., Decker, R., and Schmidt-Thieme, L. (eds.): *Data Analysis and Decision Support*. Springer, Heidelberg, 47-56.
- Hennig, C. (2010) Methods for merging Gaussian mixture components, *Advances in Data Analysis and Classification*, 4, 3-34.
- Jain, A. K. 1988. *Algorithms for Clustering Data*. Prentice Hall.
- Kaufman, L., & peter j.rouss. (2005). *Finding Group In Data*.
- Pusdiktasari, Z. F., et al. (2021). *Pengelompokan Provinsi di Indonesia dengan Ekonomi Terdampak Covid-19 Menggunakan Analisis Cluster*. 5(1), 117–129.
- R Foundation. (2020, Januari 17). *What is R?* Retrieved from R: <https://www.rproject.org/about.html>
- Rendon, E., et al. (2011). Internal versus External cluster validation indexes .

INTERNATIONAL JOURNAL OF COMPUTERS AND COMMUNICATIONS,
27-34.

Schubert, E., et al. (2017). DBSCAN revisited, revisited: Why and how you should (still) use DBSCAN. *ACM Transactions on Database Systems*, 42(3).
<https://doi.org/10.1145/3068335>