

ANALISIS DATA SEBARAN BANDWIDTH MENGGUNAKAN ALGORITMA DBSCAN UNTUK MENENTUKAN TINGKAT KEBUTUHAN BANDWIDTH DI KABUPATEN PURWAKARTA

¹⁾ Teguh Iman Hermanto, ²⁾ Yusuf Muhyidin

^{1,2)} Teknik Informatika, STT Wastukencana Purwakarta

^{1,2)} Jl. Cikopak No. 53, Sadang, Purwakarta - Jawa Barat - Indonesia

E-mail : teguh@stt-wastukencana.ac.id

ABSTRAK

Berdasarkan data yang tercatat pada tahun 2018 terdapat 43 organisasi perangkat daerah di kabupaten Purwakarta yang sudah mendapatkan bandwidth internet. Namun saat ini jumlah pembagian bandwidth dan tingkat kebutuhan belum dapat dikelompokkan. Tujuan dari penelitian ini untuk menentukan tingkat kebutuhan bandwidth di Purwakarta dengan cara melakukan analisis data mining terhadap data yang ada menggunakan algoritma DBSCAN sehingga akan terbentuk cluster yang dibagi berdasarkan tingkat kebutuhan. Pada penelitian ini metode analisis yang digunakan yaitu SEMMA dengan tahapan meliputi Data Selection, Pre-processing / cleaning, Transformation, Data Mining dan Assess / Evaluation. Hasil dari analisis menggunakan nilai $minpts = 5$ dan nilai $epsilon = 3$. Cluster yang terbentuk yaitu sebanyak 2 cluster, cluster 1 merupakan tingkat kebutuhan bandwidth rendah dan cluster 2 merupakan tingkat kebutuhan bandwidth sedang, dan 1 Noise merupakan tingkat kebutuhan bandwidth yang terlalu tinggi. Dari hasil cluster yang terbentuk dapat digunakan untuk pemerataan kebutuhan *bandwidth* pada organisasi perangkat daerah di kabupaten Purwakarta.

Kata Kunci: analisis, *Data Mining*, SEMMA, DBSCAN.

ABSTRACT

Based on data recorded in 2018 there are 43 regional apparatus organizations in Purwakarta regency that have gained internet bandwidth. However, at this time the amount of bandwidth sharing and the level of needs cannot be grouped yet. The purpose of this study is to determine the level of bandwidth requirements in Purwakarta by analyzing data mining of existing data using the DBSCAN algorithm so that a cluster will be formed which is divided based on the level of need. In this study the analytical method used is SEMMA with stages including Data Selection, Pre-processing / cleaning, Transformation, Data Mining and Assess / Evaluation. The results of the analysis use the value of $minpts = 5$ and $epsilon$ value = 3. The clusters formed are as many as 2 clusters, cluster 1 is the level of low bandwidth requirements and cluster 2 is the level of medium bandwidth requirements, and 1 Noise is a reminder that bandwidth requirements are too high. From the results of the cluster formed can be used to equalize the bandwidth requirements of regional device organizations in Purwakarta district.

Keyword: Analysis, *Data Mining*, SEMMA, DBSCAN.

PENDAHULUAN

Dinas Komunikasi dan Informatika (Diskominfo) Kabupaten Purwakarta, terdiri dari 4 (empat) bidang yaitu Bidang Aplikasi dan Informatika (Aptika), Bidang Teknologi Informasi (Jaringan), Bidang Statistik dan Persandian dan Bidang Informasi dan Komunikasi Publik. Dinas Komunikasi dan Informatika Kabupaten Purwakarta merupakan salah satu instansi pemerintah yang mempunyai tugas membantu Bupati melaksanakan urusan pemerintahan bidang komunikasi dan

informatika, bidang persandian, dan bidang statistik yang menjadi kewenangan Kecamatan dan tugas pembantuan yang ditugaskan kepada Kecamatan.

Berdasarkan data yang tercatat hingga saat ini terdapat 43 sebaran *bandwidth internet* di kabupaten Purwakarta yang tersebar di setiap organisasi perangkat daerah. Jumlah ini diambil dari data sebaran *bandwidth internet* tahun 2018 dan belum termasuk organisasi perangkat daerah yang belum mendapatkan akses untuk *bandwidth internet*. Untuk penambahan kecepatan

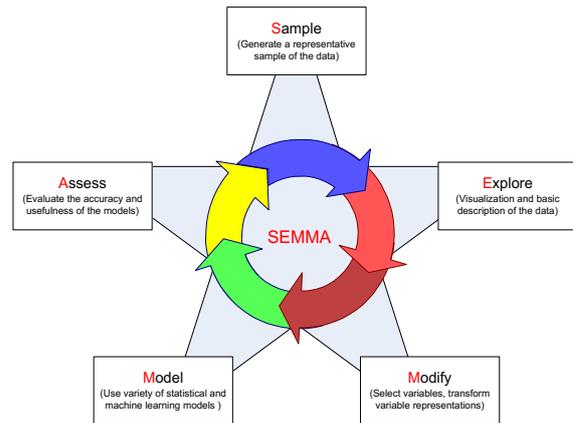
maupun *bandwidth internet* itu disesuaikan dengan anggaran yang ada di Dinas Komunikasi dan Informatika.

Dalam penelitian ini penulis melakukan proses *Clustering* dilakukan dengan menggunakan metode *Density Based Spatial Clustering of Application With Noise*. *Density Based Spatial Clustering of Application With Noise* merupakan metode yang dikembangkan berdasarkan *density* (kepadatan) tertentu. Metode ini menganggap *cluster* sebagai suatu area yang berisi objek-objek yang padat atau sesak, yang dipisahkan oleh area yang memiliki kepadatan rendah (merepresentasikan *Noise*).

Density Based Spatial Clustering of Application With Noise adalah algoritma menumbuhkan area-area dengan kepadatan yang cukup tinggi ke dalam *cluster-cluster* and menemukan *cluster-cluster* dalam bentuk yang sembarang dalam suatu database spatial yang memuat *Noise* [1].

Data mining merupakan suatu konsep yang umumnya digunakan untuk menemukan pengetahuan yang tersembunyi dalam basis data [2]. *Data mining* dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan yaitu Deskripsi, Estimasi, Prediksi, Klasifikasi, Pengklasteran, dan Asosiasi [3].

SEMMA merupakan singkatan dari *Sample, Explore, Modify, Model, Assess*. Singkatan tersebut mengacu pada proses dalam melakukan sebuah objek *data mining*. Pada umumnya penerapannya, SAS institute membagi siklus SEMMA menjadi 5(lima) tahapan untuk proses *data mining* [4].



Gambar 1. Tahapan dalam SEMMA

METODE

Pada penelitian ini penulis menggunakan metode SEMMA dalam pengolahan data menggunakan Teknik data mining sebagai berikut (a) Tahap pengumpulan data (*Sample*), (b) Tahap deskripsi data (*Explore*), (c) Tahap transformasi data (*Modify*), (d) Tahap pemodelan data (*Model*), (e) Tahap evaluasi data (*Asses*) [5] :

- a. Tahap Pengumpulan data (*Sample*)
Pada tahap ini Penulis mengumpulkan data dengan meminta data sebaran *Bandwidth* yang berjumlah 43 organisasi perangkat daerah Purwakarta.
- b. Tahap deskripsi data (*Explore*)
Dari data yang terkumpul, dilakukan pengekplorasian dengan menjelaskan data secara lengkap mulai dari data nama perangkat daerah sesuai dengan atribut-atribut seperti koordinat latitude, koordinat longitude, sebaran *bandwidth*, kapasitas metro, dan jumlah pengguna. Pada tahap ini juga dilakukan Data Selection dan Data Pre-processing/ Data Cleaning.
- c. Tahap transformasi data (*Modify*)
Tahap ini melakukan modifikasi terhadap data dengan menciptakan, memilih, dan mengubah variable untuk focus pada proses pemilihan model.

- d. Tahap pemodelan data (*Model*)
Tahap ini terdiri dari pemodelan data dengan suatu perangkat lunak untuk mencari kombinasi data yang dapat memprediksi hasil yang diinginkan.
- e. Tahap evaluasi data (*Asses*)
Tahap ini terdiri dari penilaian data dengan mengevaluasi kegunaan dan keandalan dari temuan pada proses *data mining*

HASIL

A. *Sample*

Sumber data utama yang digunakan dalam penelitian ini berasal dari data sebaran *Bandwidth internet* Dinas Komunikasi dan Informatik Purwakarta yang berjumlah 43 opd. Data tersebut nantinya akan diproses untuk dapat menghasilkan tingkat kebutuhan *bandwidth* yang sesuai dengan sistem *data mining* yang digunakan. Contoh data sebaran yang didapatkan dapat dilihat pada table 1:

Tabel 1. Contoh data sebaran

id	Nama OPD	Alamat Lokasi	Latitude	Longitude	Sebaran bw	Kapasitas metro	Alamat ip	Jumlah Pengguna	Keterangan
1	Sekretariat Daerah Purwakarta	Jl. Gandanegara No.25	- 6.557.002	107.442.270	20	20		10	-
2	Sekretariat DPRD Purwakarta	Jalan Pemuda Blok Haji Makbul No. 26,	- 6.564.471	107.432.753	6	6		3	-
3	Inspektorat Daerah Purwakarta	Jl. Veteran No.147,	- 6.527.871	107.446.852	6	6		3	-
4	Dinas Pendidikan Purwakarta	Jl. Surawinata No.30A,	- 6.545.882	107.445.513	7	10		5	-

B. *Explore*

Setelah melakukan pengumpulan data dilakukanlah penyeleksian atribut kepada data awal yang didapatkan, penyeleksian ini bertujuan untuk memilih dan memilah atribut apa saja yang harus diambil atau dibuang. Adapun teknik yang penulis gunakan dalam seleksi data ini yaitu *feature selection* yang terdiri dari tahap *Eliminate*, *engineer*, dan *validate* [6].

1. Eliminate

Pada tahap ini penulis melakukan pemilihan atribut yang akan digunakan seperti id, nama OPD, Latitude, Longitude, sebaran_bw, kapasitas metro, dan jumlah pengguna.

2. Engineer

Setelah proses *eliminate* selanjutnya data sumber tersebut dibentuk kembali menjadi data yang siap divalidasi. Hasil dari tahap *engineer* dapat dilihat pada table 2

Tabel 2. Pengolahan Data

id	Nama OPD	Latitude	Longitude	Sebaran bw	Kap. metro	Jumlah pengguna
1	Sekretariat Daerah Purwakarta	- 6.557.002	107.442.270	20	20	10
2	Sekretariat DPRD	-	107.432.753	6	6	3

	Purwakarta	6.564.471				
3	Inspektorat Daerah Purwakarta	- 6.527.871	107.446.852	6	6	3
4	Dinas Pendidikan Purawakarta	- 6.545.882	107.445.513	10	10	5

C. Modify

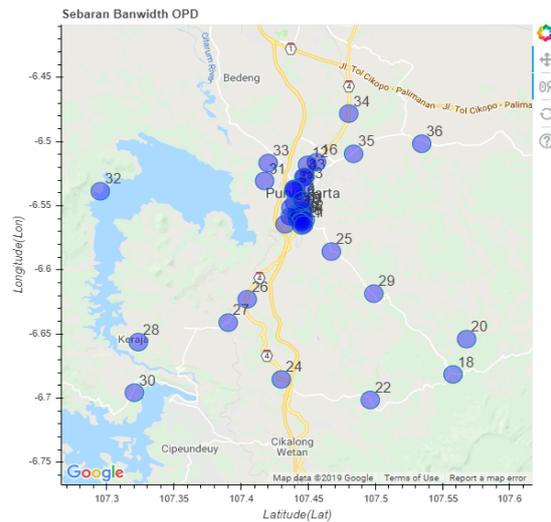
Tahap ini melakukan modifikasi terhadap data dengan menciptakan, memilih, dan mengubah variable untuk focus pada proses pemilihan model. Pada tahap ini atribut yang awalnya *Numeric* menjadi tipe kategori dengan tujuan untuk memudahkan dalam pemrosesan *data mining* [7]. Dan atribut yang dikelompokan dan dinisialisasi adalah atribut Jumlah Pengguna. Untuk hasil pengelompokan atribut Jumlah Pengguna dapat dilihat pada Tabel 3.

Tabel 3. Transformasi Numeric menjadi kategori

No	Jml_pengguna	Inisialisasi
1	1-5	Rendah
2	6-10	Sedang
3	11-15	Tinggi

Selain Jumlah Pengguna penulis juga mengubah data koordinat kedalam bentuk

pemetaan *maps* sesuai dengan data koordinat latitude dan longitude. Untuk hasil transformasi data koordinat dapat dilihat pada gambar 2.



Gambar 2. Transformasi Data ke Map

Untuk hasil keseluruhan dari proses transformasi dapat dilihat pada tabel 3.6 yang menjelaskan rincian dari pemetaan pada map

Tabel 3. Transformasi Tabel dari map

id	Nama OPD	Latitude	Longitude	Sebaran bw	Kap. metro	Jumlah pengguna
1	Sekretariat Daerah Purwakarta	- 6.557.002	107.442.270	20	20	2
2	Sekretariat DPRD Purwakarta	- 6.564.471	107.432.753	6	6	1
3	Inspektorat Daerah Purwakarta	- 6.527.871	107.446.852	6	6	1
4	Dinas Pendidikan Purawakarta	- 6.545.882	107.445.513	10	10	1

D. Model Analisis menggunakan Algoritma DBSCAN

Analisa perhitungan algoritma DBSCAN menggunakan 43 data organisasi perangkat daerah yang ada di purwakarta, untuk perhitungan manualnya dapat dilihat dari iterasi yang dibuat.

Iterasi 1 adalah dengan mengitung jarak masing masing point atau titik terhadap titik pusat 2. Berikut adalah perhitungan jarak masing-masing titik dengan *core point* untuk iterasi pertama, pada soal diatas dengan menggunakan rumus jarak [8].

- a. MinPts : 5
- b. Eps : 3
- c. Pusat Titik 2 Latitude=107,432753
Longitude=-6,564471 Sebaran=6;
Kapasitas=6 pengguna=1.

$$\sqrt{(-6,564471 - 6,564471)^2 + (107,432753 - 107,432753)^2 + (6 - 6)^2 + (6 - 6)^2 + (1 - 1)^2} = 0$$

E. Analisis Data Mining

Berdasarkan hasil yang didapatkan dari visualisasi, didapatkan sebuah pengetahuan sebagai berikut :

- 1. Total *cluster* pada sebaran *bandwidth internet* organisasi perangkat daerah yang terbentuk adalah 2 dengan total *Noise* sebanyak 7 dari 43 data sebaran *bandwidth* yang diolah dengan menggunakan algoritma DBSCAN
- 2. Untuk *cluster* 1 terdiri dari 15 organisasi perangkat daerah, untuk hasil pengelompokkan dapat dilihat pada tabel 4:

Tabel 4. Data OPD Cluster 1

No	Nama OPD	Cluster
1	Sekretariat DPRD Purwakarta	1
2	Inspektorat Daerah Purwakarta	1
3	Dinas PU Bina Marga dan	1

	Pengairan Purwakarta	
4	Dinas Tata Ruang dan Permukiman	1
5	Dinas Sosial Pemberdayaan Perempuan Dan Perlindungan Anak Purwakarta	1
6	Dinas Polisi Pamong Praja	1
7	Dinas Pemadam Kebakaran dan Penanggulangan Bencana	1
8	Dinas Pemberdayaan Masyarakat Dan Desa	1
9	Dinas Perhubungan Kabupaten Purwakarta	1
10	Dinas Koperasi UKM, Perdagangan, dan Perindustrian	1
11	Dinas Pemuda Olahraga Pariwisata Dan Kebudayaan	1
12	Dinas Kearsipan dan Perpustakaan	1
13	Badan Lingkungan Hidup Purwakarta	1
14	Kecamatan Cibatu	1
15	Radio Pro 89 FM	1

Untuk *cluster* 2 terdiri dari 21 organisasi perangkat daerah, untuk hasil pengelompokkan dapat dilihat pada tabel 5 :

Tabel 5. Data OPD Cluster 2

No	Nama OPD	Cluster
1	Kecamatan Wanayasa	2
2	Kecamatan Kiarapedes	2
3	Kecamatan Bojong	2
4	Kantor Kecamatan Purwakarta	2
5	Kecamatan Darangdan	2
6	Kecamatan Pasawahan Purwakarta	2
7	Kecamatan Sukatani	2
8	Kecamatan Plered	2
9	Kecamatan Tegalwaru	2
10	Kecamatan Pondok Salam	2

11	Kecamatan Maniis	2
12	Kecamatan Jatiluhur	2
13	Kecamatan Sukasari	2
14	Kecamatan Babakancikao	2
15	Kecamatan Bungursari	2
16	Kecamatan Campaka	2
17	Kelurahan Purwamekar	2
18	Kelurahan Cipaisan	2
19	Kelurahan Sindangkasih	2
20	Puskesmas Kota Purwakarta	2
21	Kantor Kesbangpol	2

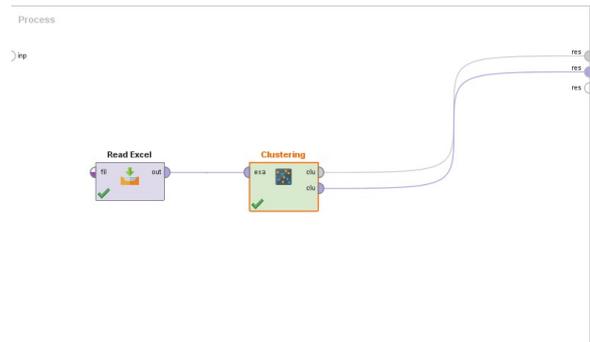
Untuk *Noise* terdiri dari 7 organisasi perangkat daerah, untuk hasil pengelompokan dapat dilihat pada tabel 6 :

Tabel 6. Data yang masuk *Noise*

No	Nama OPD	Ket
1	Sekretariat Daerah Purwakarta	<i>Noise</i>
2	Dinas Pendidikan Purawakarta	<i>Noise</i>
3	Dinas Kependudukan dan Pencatatan Sipil	<i>Noise</i>
4	Dinas Komunikasi dan Informatika	<i>Noise</i>
5	Dinas Perencanaan Pembangunan, Penelitian dan Pembangunan Daerah	<i>Noise</i>
6	Badan Keuangan dan Aset Daerah	<i>Noise</i>
7	LPSE Kabupaten Purwakarta	<i>Noise</i>

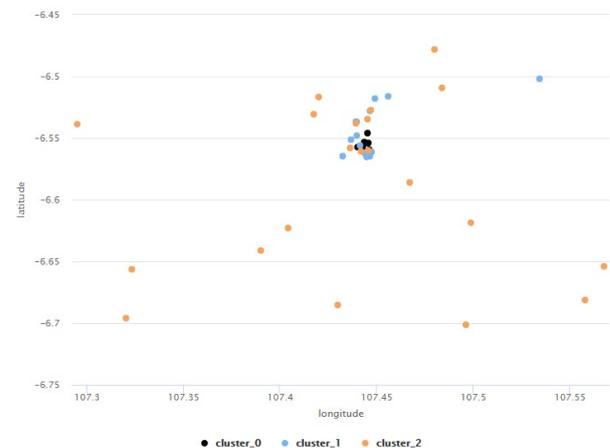
F. Visualisasi Hasil Data Mining

Penggalian data pada *rapidminer* dilakukan dengan mengimport data *excel* yang sudah disiapkan ke dalam *rapidminer*, setelah data dimasukan pada *read excel* kemudian dihubungkan dengan operator *modeling Density Based Spatial Clustering with Noise* dengan nilai epsilon 3 dan minpoint 5. Pemodelan cluster dapat dilihat pad gambar 3.



Gambar 3. Alur Pemodelan *Cluster* Pada Rapidminer

Kemudian hubungkan operator *model clustering* ke res dengan cara menarik garis dari *clu(cluster)* uke res. Setelah semua nilai ditentukan maka Algoritma DBSCAN dapat bekerja. Untuk hasil visualisasi dari *Clustering* algoritma DBSCAN dapat dilihat pada gambar 4 :



Gambar 4. Ploting Hasil DBSCAN Menggunakan Rapidminer

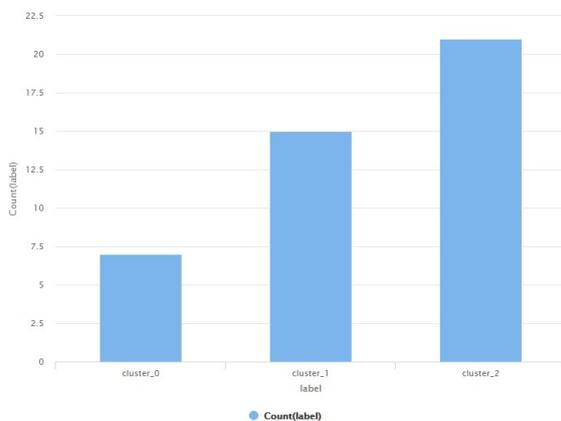
G. Assess

Berdasarkan Hasil Analisa algoritma *Density Based Spatial Clustering of Application With Noise* dengan menggunakan nilai minpts = 5 dan nilai epsilon = 3 *cluster* yang terbentuk yaitu :

1. *Cluster* 1 terdapat 15 organisasi perangkat daerah
2. *Cluster* 2 terdapat 21 organisasi perangkat daerah

3. *Noise* terdapat 7 organisasi perangkat daerah

Hasil pembagian *cluster* dari RapidMiner sama dengan perhitungan manual yang dilakukan di Microsoft excel sehingga data yang digunakan valid, dari perhitungan yang sudah dilakukan dapat disimpulkan bahwa *cluster* 1 yaitu *cluster* dengan tingkat kebutuhan *bandwidth* rendah, *cluster* 2 memiliki tingkat kebutuhan sedang dan *Noise* itu merupakan kebutuhan *bandwidth* yang terlalu tinggi, sehingga belum adanya pemerataan pembagian *bandwidth* di organisasi perangkat daerah Purwakarta [9]. Statistik data *cluster* yang terbentuk dapat dilihat pada gambar 5.



Gambar 5. Statistik Data *Cluster* Yang Terbentuk

Kesimpulan dari hasil *Assess/Evaluasi* diatas dapat dilihat sebagai berikut :

1. *Cluster* 1 dengan tingkat kebutuhan *bandwidth* rendah berkisar 2 sampai 5 MBps dan perlu ditambahkan kecepatan agar bias dilakukan pemerataan pada sebaran *bandwidth*.
2. *Cluster* 2 dengan tingkat kebutuhan sedang berkisar 6 sampai 9 MBps dan sudah memenuhi kriteria standar *Bandwidth* yang dibagikan.
3. *Noise* dengan tingkat kebutuhan *bandwidth* yang tinggi dikarenakan ada beberapa

organisasi perangkat daerah purwakarta yang memiliki kebutuhan yang sangat tinggi.

Dari ketiga *cluster* yang terbentuk perlu dilakukan pemerataan *bandwidth* agar tidak ada organisasi perangkat daerah yang mendapatkan sebaran *bandwidth* yang terlalu tinggi dan terlalu rendah [10]. Hasil dari proses clustering ini dapat memberikan kontribusi untuk pemerintah kabupaten Purwakarta dalam melakukan pemerataan sebaran *bandwidth* pada masing-masing organisasi perangkat daerah.

KESIMPULAN

Kesimpulan yang didapatkan dari analisis *data mining* ini adalah pemilihan parameter *minpts* dan *epsilon* sangat berpengaruh dalam proses pengolahan data menggunakan algoritma *Density Based Spatial Clustering of Application With Noise*. dengan menentukan *minpts* dan *epsilon* dapat mempengaruhi proses kerja algoritma untuk menentukan jumlah *cluster* dan *noise* yang terbentuk. Hasil *cluster* menunjukkan perbedaan kebutuhan *bandwidth* pada masing-masing perangkat daerah sehingga dapat dilakukan pemerataan *bandwidth* yang lebih optimal.

DAFTAR PUSTAKA

- [1] Sander, J., Ester, M., Kriegel, H.-P., & Xu, X. (1998). Density-based clustering in spatial databases: The algorithm GDBSCAN and its applications. *Int. Journal, Kluwer Academic Publishers*.
- [2] Turban, E., Aronson, J. ., Liang, T. -. & McCarthy, T. -. (2006). *Decision Support Systems*, Pearson Education. *Decision Support Systems, Pearson Education*.
- [3] Larose, D. T. (2005). *Discovering Knowledge in Data: An Introduction to Data Mining*. New Jersey: John Wiley & Sons, Inc.
- [4] Azevedo, Santos, A., & F, M. (2008). *KDD, SEMMA AND CRISP-DM: A PARALLEL OVERVIEW*. IADIS.

- [5] Nagpal, P. ., & Mann, P. . (2011). Comparative study if density of based clustering algorithms. *American Jurnal of Applications*
- [6] Prasetyo, E. (2012). *Data Mining Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta: ANDI
- [7] Arsih, N., Hajarisman, N., & Darwis, S. (n.d.). *Metode Pengclusteran Berbasis Densitas Menggunakan Algoritma DBSCAN*. Bandung: Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung.
- [8] Anindya Santika Devi, N.M., Gede Darma Putra, I.K. dan Sukarsa, I.M., 2015. Implementasi Metode Clustering DBSCAN pada Proses Pengambilan Keputusan. *Lontar Komputer*, 6(3), pp.185–191
- [9] Imanuel Ndaumanu, R., Kusrini dan Arief, M.R., 2014. Analisis Prediksi Tingkat Pengunduran Diri Mahasiswa dengan Metode K-Nearest Neighbor. *JatISI*, 1(1).
- [10] Kusumo, D.S., Bijaksana, M.A. and Darmantoro, D., 2016. DATA MINING DENGAN ALGORITMA APRIORI PADA RDBMS ORACLE. *TEKTRIKA - Jurnal Penelitian dan Pengembangan Telekomunikasi, Kendali, Komputer, Elektrik, dan Elektronika*, 8(1), pp.1–5