

Clustering Pada Data Mining Untuk Mengetahui Potensi Penyebaran Penyakit DBD Menggunakan Metode Algoritma K-Means dan Metode Perhitungan Jarak Euclidean Distance

Muhammad Hariyanto¹⁾, Rizky Tahara Shita²⁾

¹⁾Teknik Informatika, Fakultas Teknologi Informasi, Universitas Budi Luhur

^{1,2)}Jl. Raya Ciledug, Petukangan Utara, Kebayoran Lama, Jakarta Selatan 12260

E-mail : muhammadhariyanto201@gmail.com¹⁾, rizky.tahara@gmail.com²⁾

Abstrak

Demam Berdarah merupakan penyakit menular yang terdapat di wilayah tropis maupun subtropis. Semakin meningkatnya kepadatan penduduk di Kota Tangerang Selatan pada tahun 2016 berdasarkan Badan Pusat Statistik Kota Tangerang Selatan jumlah penduduk mencapai 1.504.835 jiwa yang tersebar di 52 kelurahan. Melihat daerah penyebarannya perlu dibuat sebuah pengelompokan data penyebaran agar dapat memperoleh pusat titik penyebaran. Implementasi data mining menggunakan algoritma K-Means sangat membantu dalam pengelompokan daerah sporadis, potensi dan endemis. Pengelompokan C1 berada pada titik pusat cluster Kelurahan Rawa Buntu, C2 berada pada titik pusat cluster Kelurahan Pondok Ranji, dan C3 berada pada titik pusat cluster Kelurahan Serua Indah. Hasil pengelompokan C1 ada 19 Kelurahan, C2 ada 13 Kelurahan, dan C3 ada 20 Kelurahan dari total 52 Kelurahan. Akan menjadi bahan untuk melakukan penyehatan lingkungan sesuai dengan kelompok yang akan dikerjakan oleh Dinas Kesehatan Kota Tangerang Selatan.

Kata Kunci : Data Mining, Clustering, K-Means, Demam Berdarah

1. PENDAHULUAN

Penyakit Demam Berdarah masih menjadi permasalahan kesehatan masyarakat. Demam Berdarah merupakan salah satu jenis penyakit menular yang terdapat di wilayah tropis maupun subtropis, penyakit demam berdarah dapat menular karena gigitan nyamuk *Aedes Aegypti*. Dampak dari demam berdarah dapat membuat suhu tubuh penderita menjadi sangat tinggi dan pada umumnya disertai demam, mual/muntah, sakit kepala, dan nyeri perut. Dinas kesehatan mempunyai tugas pokok untuk membantu penyelenggaraan kegiatan penyehatan lingkungan.

Berdasarkan Badan Pusat Statistik Kota Tangerang Selatan pada tahun 2016 berjumlah 1.504.835 jiwa dan tersebar di 52 Kelurahan. Meningkatnya penderita penyakit demam berdarah dikarenakan kurangnya kesadaran masyarakat terhadap kesehatan lingkungan. Timbul masalah yang menyulitkan Dinas Kesehatan untuk mengelompokan daerah yang akan menghasilkan titik-titik pusat penyebaran penderita demam berdarah, maka dilakukan pengelompokan data berdasarkan jumlah penduduk, jumlah penderita. Agar program kerja Dinas Kesehatan Kota Tangerang Selatan bisa dengan mudah melakukan budi daya tanaman anti nyamuk, program 3M (Menguras, Menutup, Mengubur) yang bertujuan untuk menurunkan angka penderita penyakit demam berdarah dan melakukan penyuluhan untuk memberi informasi tentang bahaya penyakit demam berdarah.

Tujuan penulisan penelitian ini adalah untuk membangun sebuah sistem yang dapat mengelompokan data daerah penyebaran penyakit

demam berdarah, dengan mengimplementasikan algoritma K-Means Clustering maka akan menghasilkan titik-titik pusat penyebaran penyakit demam berdarah, dengan itu maka dapat memudahkan pihak Dinas Kesehatan Kota Tangerang Selatan dalam melaksanakan program kerja yang telah direncanakan untuk mencegah terjadinya penyebaran penyakit Demam Berdarah yang semakin meningkat.

Berdasarkan pembahasan diatas dapat dirumuskan permasalahan “Bagaimana cara implementasi data mining untuk pengelompokan daerah penyebaran penyakit demam berdarah menggunakan algoritma k-means ?”

Batasan masalah pada penelitian kali ini telah ditetapkan agar tidak menyimpang dari pokok pembahasan dan demi terciptanya tujuan penelitian yang telah direncanakan, adapun batasan masalah pada penelitian kali ini diantaranya, penelitian hanya mencakup 52 kelurahan yang berada di kota Tangerang Selatan, data penduduk yang digunakan berdasarkan Badan Statistik Kota Tangerang Selatan pada tahun 2016, dan penelitian ini hanya membantu Dinas Kesehatan dalam mengelompokan daerah potensi penyebaran penyakit demam berdarah.

Metodologi yang digunakan pada penelitian kali ini menggunakan metode penelitian Studi Literatur dan Observasi, sedangkan metode pengembangan perangkat lunak pada penelitian kali ini menggunakan metode Waterfall yang terdiri dari 4(empat) tahapan diantaranya, Tahap analisis, Tahap desain, Tahap implementasi, dan Tahap perawatan.

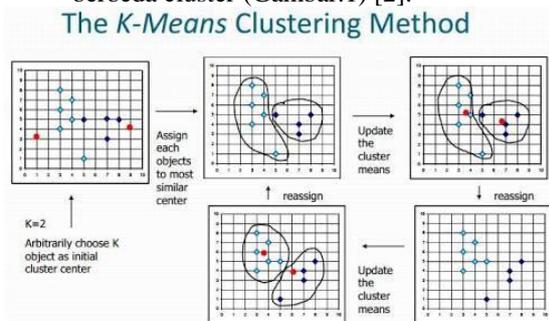
2. LANDASAN TEORI

Data mining merupakan suatu istilah yang digunakan untuk mendapatkan pengetahuan yang tersembunyi dari kumpulan data yang berukuran sangat besar. Tujuan utama data mining adalah untuk menemukan, menggali, atau menambang pengetahuan dari data atau informasi yang kita miliki [1].

Data mining merupakan bidang dari segala bidang keilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, database, dan visualisasi untuk pengenalan permasalahan pengambilan informasi dari database yang besar. Untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari database yang besar tersebut dapat menggunakan beberapa teknik diantaranya, teknik statistik, matematika, kecerdasan tiruan, dan machine learning [1].

2.1 Clustering

Clustering atau Klasterisasi adalah suatu alat bantu pada data mining yang bertujuan mengelompokkan objek-objek kedalam cluster-cluster. Cluster adalah sekelompok atau sekumpulan objek-objek data yang mempunyai kemiripan karakteristik satu sama lain dalam cluster yang sama dan berbeda karakteristik terhadap objek-objek yang berbeda cluster (Gambar.1) [2].



Gambar 1 : Metode Clustering

2.2 Algoritma K-Means Clustering

Metode K-Means pertama kali diperkenalkan oleh MacQueen JB pada tahun 1976 [1].

Metode ini adalah salah satu metode non hierarchi yang umum digunakan. Metode ini termasuk dalam teknik partisi yang membagi atau memisahkan objek ke kelompok daerah bagian yang terpisah. Pada K-Means, setiap objek harus masuk dalam kelompok tertentu, tetapi dalam satu tahapan proses tertentu, objek yang sudah masuk dalam satu kelompok, pada satu tahapan berikutnya akan pindah ke kelompok lain [5].

Hasil cluster dengan metode K-Means sangat bergantung pada nilai pusat kelompok awal yang diberikan. Pemberian nilai awal yang berbeda bisa menghasilkan kelompok

yang berbeda. Ada beberapa cara memberi nilai awal misalnya dengan mengambil sampel awal dari objek, lalu mencari nilai pusatnya, memberi nilai awal secara random, menentukan nilai awalnya atau menggunakan hasil dari kelompok hirarki dengan jumlah kelompok yang sesuai [5].

K-Means adalah suatu metode penganalisaan data atau metode Data Mining yang melakukan proses pemodelan tanpa supervisi (unsupervised) dan merupakan salah satu metode yang melakukan pengelompokan data dengan sistem partisi. Metode K-Means berusaha mengelompokkan data yang ada kedalam beberapa kelompok, dimana data dalam satu kelompok mempunyai karakteristik yang sama satu dengan yang lainnya dan mempunyai karakteristik yang berbeda dengan data yang berada pada kelompok yang lain. Dengan kata lain, metode ini berusaha untuk meminimalkan variasi antar data yang berada didalam suatu kelompok dan memaksimalkan variasi dengan data yang berada pada kelompok lainnya [5].

2.3 Metode Analisa K-Means Clustering

Proses clustering dimulai dengan mengidentifikasi data yang akan di kluster, X_{ij} ($i=1, \dots, n; j=1, \dots, m$) dengan n adalah jumlah data yang akan di kluster dan m adalah jumlah variabel. Pada awal iterasi, pusat setiap cluster ditetapkan secara random,

$$C_{kj} \quad (k = 1, \dots, k; j = 1, \dots, m).$$

Kemudian dihitung jarak antara setiap data dengan setiap pusat cluster. Untuk melakukan perhitungan jarak data ke- i (x_i) pada pusat cluster ke- k (c_k), diberi nama (d_{ik}), dapat digunakan rumus formula Eulidean Distance seperti yang terlihat pada rumus persamaan (1), berikut :

$$(1) \dots \quad d_{ij} = \sqrt{\sum_{k=1}^m X_{ij} - C_{ij}^2}$$

Dimana :

d_{ij} : Jarak objek antar nilai data dan nilai pusat cluster

m : Jumlah dimensi data

X_{ij} : Nilai data dari dimensi ke- k

X_{jk} : Nilai pusat cluster dari dimensi ke- k

Untuk menghitung centroid baru, dapat menggunakan rumus persamaan (2), sebagai berikut:

$$(2) \dots \quad C = \frac{\sum m}{n}$$

Dimana :

- C : Centroid data
- m : Anggota data yang termasuk kedalam centroid tertentu
- n : Jumlah data yang menjadi anggota centroid tertentu

Penambahan langkah perhitungan yaitu Rasio antara BCV (Between Cluster Variation) dengan WCV (Within Cluster Variation) sebagai berikut [6] :

Menghitung BCV :

$$C1 = (3;2,59;1), C2 = (1;3,46;2), C3 = (1;3,02;3)$$

$$d(c1, c2) = \sqrt{(3-1)^2 + (2,59-3,46)^2 + (1-2)^2} = 2.40$$

$$d(c1, c3) = \sqrt{(3-1)^2 + (2,59-3,02)^2 + (1-3)^2} = 2.86$$

$$d(c2, c3) = \sqrt{(1-1)^2 + (3,46-3,02)^2 + (1-3)^2} = 1.09$$

$$BCV = d(c1, c2) + d(c1, c3) + d(c2, c3) = 2.40 + 2.86 + 1.09 = 6.35.$$

Maka nilai BCV (Between Cluster Variation) nya adalah 6,35.

Rumus menghitung WCV :

$$WCV = \sum (\text{jarak pusat tiap cluster paling minimum})^2$$

Untuk menghitung Rasio dapat menggunakan rumus persamaan (3) :

$$(3)... Rasio = \frac{BCV}{WCV}$$

3. RANCANGAN SISTEM DAN APLIKASI

3.1. Membaca Berkas Dokumen Uji

Seperti halnya pemilihan berkas dokumen sample, pemilihan berkas dokumen uji juga menggunakan Browse File untuk memilih dokumen yang diinginkan. Pada saat berkas telah dipilih, program akan melakukan pengambilan nilai atribut yang dibutuhkan sesuai dengan pengaturan program yang telah ditentukan sebelumnya. Banyaknya nilai atribut yang diambil dari data uji harus sama banyaknya dengan nilai atribut yang dimiliki data sample.

3.2. Pengujian K-Means Clustering

Pengujian K-Means dilakukan dengan menghitung jarak seluruh data sample terhadap satu data uji. Pada penelitian ini, perhitungan jarak akan dilakukan dengan rumus persamaan (1) Eulidean Distance.

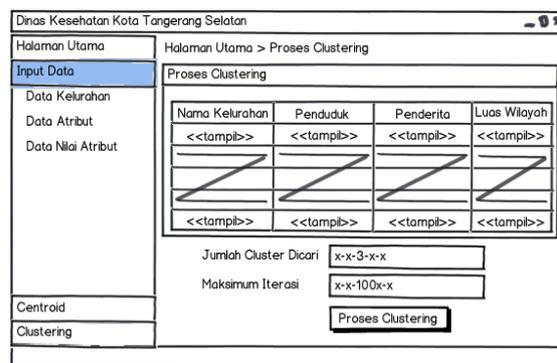
$$(1)... dij = \sqrt{\sum_{k=1}^m Xij - Cij^2}$$

3.3. Rancangan Layar

Pada rancangan layar proses clustering ini admin dapat melakukan proses clustering dari semua

data yang ada sehingga menghasilkan informasi yang diperlukan, sesuai dengan topik pembahasan

pada penelitian kali ini yaitu implementasi algoritma K-Means Clustering untuk mengetahui daerah potensi penyebaran penyakit demam berdarah (DBD) pada Dinas Kesehatan Kota Tangerang Selatan (Gambar.2).



Gambar 2 : Rancangan Layar Proses Clustering

4. HASIL DAN PEMBAHASAN

4.1. Tentukan Jumlah Cluster dan Tentukan Titik Pusat Awal

Dalam penelitian kali ini titik pusat awal ditentukan secara random dan didapat titik pusat dari setiap cluster, dapat dilihat pada (tabel 1) berikut ini.

Tabel 1 : Titik Pusat Centroid Awal

| Titik Pusat Awal | Kelurahan | Penderita | Penduduk |
|------------------|-------------|-----------|----------|
| C1 (SPORADIS) | Rawa Buntu | 25 | 33768 |
| C2 (POTENSI) | Pd.Ranji | 18 | 39956 |
| C3 (ENDEMIS) | Serua Indah | 5 | 19796 |

4.2. Hitung Jarak dengan Rumus Euclidean Distance

Rumus formula Eulidean Distance seperti yang terlihat pada rumus persamaan (1), berikut :

$$(1)... dij = \sqrt{\sum_{k=1}^m Xij - Cij^2}$$

$$d(C1, data1) = \sqrt{(data1 \text{ dim } ensi1 - C1 \text{ dim } ensi1)^2 + (data1 \text{ dim } ensi2 - C1 \text{ dim } ensi2)^2}$$

$$d(C1, data1) = \sqrt{(12 - 25)^2 + (38318 - 33768)^2}$$

$$d(C1, data1) = \sqrt{169 + 20702500}$$

$$d(C1, data1) = \sqrt{20702669}$$

$$d(C1, data1) = 4550.0185713907$$

$$d(C2, data1) = \sqrt{(data1 \text{ dim ensi1} - C2 \text{ dim ensi1})^2 + (data1 \text{ dim ensi2} - C2 \text{ dim ensi2})^2}$$

$$d(C2, data1) = \sqrt{(12-18)^2 + (38318-39956)^2}$$

$$d(C2, data1) = \sqrt{36 + 2683004}$$

$$d(C2, data1) = \sqrt{2683080}$$

$$d(C2, data1) = 1638.010988741$$

$$d(C3, data1) = \sqrt{(data1 \text{ dim ensi1} - C3 \text{ dim ensi1})^2 + (data1 \text{ dim ensi2} - C3 \text{ dim ensi2})^2}$$

$$d(C3, data1) = \sqrt{(12-5)^2 + (38318-19796)^2}$$

$$d(C3, data1) = \sqrt{49 + 343064484}$$

$$d(C3, data1) = \sqrt{343064533}$$

$$d(C3, data1) = 18522.001322751$$

4.3. Menghitung Centroid Baru

Berikut ini membangkitkan *centroid* baru dengan menggunakan rumus persamaan (2), yaitu :

$$(2)... C = \frac{\sum m}{n}$$

Menghitung C1

$$C1, d1 = \frac{31271+31987+36293+27230+34451+33768+31907+36463+27230+36073+27192+28031+31300}{13}$$

$$= 31784.307692307$$

$$C1, d2 = \frac{5+12+14+5+13+25+14+7+5+18+10+19+17}{13}$$

$$= 12.6153846154$$

Menghitung C2

$$C2, d1 = \frac{38318+38601+57269 + 39539 + 37136+41733+47736+39956+ 40020+ 49719+44567+44454+57269+52773+42496+46807}{16}$$

$$= 44899.5625$$

$$C2, d2 = \frac{12+20+31+25+5+12+9+18+9+6+9+9+31+23+18+20}{16}$$

$$= 16.0625$$

Menghitung C3

$$C3, d1 = \frac{17611+25057+11700+22310+16579+22633+19796+9179+ 25549+11046+15096+5247+16102+13851+8854+24488+ 18258+19775+13305+15683+21930+11327+7870}{23}$$

$$= 16228.086956521$$

$$C3, d2 = \frac{7+11+7+8+13+11+5+1+15+3+11+12+2+5+11+17+ 19+21+7+2+19+16+9}{23}$$

$$= 10.0869565217$$

4.4. Menghitung BCV (Between Cluster Variation)

$$d(C1, C2) = \sqrt{(25-18)^2 + (33768-39956)^2 + (3.28-2.06)^2}$$

$$= 6188.0040795397.$$

$$d(C1, C3) = \sqrt{(25-5)^2 + (33768-19796)^2 + (3.28-1.93)^2}$$

$$= 13972.014379555.$$

$$d(C2, C3) = \sqrt{(18-5)^2 + (39956-19796)^2 + (2.06-1.93)^2}$$

$$= 20160.004191887.$$

$$BCV = d(C1, C2) + d(C1, C3) + d(C2, C3).$$

$$BCV = 6188.0040795397 + 13972.014379555 + 20160.004191887.$$

$$BCV = 40320.022465964.$$

4.5. Menghitung WCV (Within Cluster Variation)

WCV = Jumlahkan Seluruh Hasil Kuadrat Minimal Jarak

$$WCV = 2683083.4225 + 6235409.04 + 4774229.7744 + 3172131.5876 + 1836031.6244 + 299740142.41 + 173938.2809 + 7952572.61 + 27678165.41 + 6375747.21 + 65545220.16 + 6320208.6864 + 10349155.9241 + 8048605.0441 + 3157765.0169 + 60528482.9321 + 0 + 0 + 4177.2116 + 42745845.4161 + 112720705.0289 + 95316313.2116 + 21261402.2704 + 466633.2916 + 0 + 33097112.3489 + 76562504.1089 + 22090036.0361 + 211673450.0144 + 3463443.1664 + 13645645.7396 + 35343026.3225 + 119727402.8224 + 22015008.1936 + 20232085.4225 + 299740142.41 + 7263353.0804 + 42745845.4161 + 5313074.0961 + 2365641.9881 + 697.1764 + 42133085.4761 + 16916780.8224 + 4554152.2916 + 71724082.25 + 164275517.24 + 142229492.0529 + 43244001.0144 + 32913207.6569 + 6451600.36 + 46936205.25 + 6091088.6724$$

$$WCV = 2279833646.9927$$

Nilai WCV adalah : 2279833646.9927

Selanjutnya adalah menghitung nilai Rasio, menggunakan ru,us persamaan (3), yaitu:

$$(3)... Rasio = \frac{BCV}{WCV}$$

yang didapat nilai BCV dibagi nilai WCV, Nilai BCV = 40320.022650982 dibagi dengan nilai WCV = 2279833646.9927, maka nilai Rasio nya adalah 1.7685510828462.

Setelah di hitung sampai iterasi ke-5 tidak ada data yang berpindah cluster, maka iterasi berhenti pada iterasi ke-5, berikut ini tabel perbandingan BCW, WCV, Rasio dan Iterasi, dapat dilihat pada (Tabel 2) sebagai berikut :

Tabel 2 : Perbandingan BCV, WCV, Rasio Iterasi 1, 2, 3, 4, 5

| Proses | WCV | BCV | Rasio |
|-----------|-----------------|-----------------|------------------|
| Iterasi 1 | 2279833646.9927 | 40320.022650982 | 1.7685510828462 |
| Iterasi 2 | 1416920765.049 | 57342.952378623 | 4.0470119284787 |
| Iterasi 3 | 1362849409.4504 | 62033.23056328 | 4.55177303770411 |
| Iterasi 4 | 1355933244.9515 | 63159.878722362 | 4.6580375464007 |
| Iterasi 5 | 1355933244.9515 | 63159.878722362 | 4.6580375464007 |

Dari hasil perhitungan iterasi 1 hingga iterasi 5, maka hasil Clustering yang dihasilkan adalah sebagai berikut (Gambar 3).

| Kelurahan | Cluster |
|-------------------|---------|
| Pt. Jangung Timur | C1 |
| Pt. Picing | C1 |
| Buana | C1 |
| Panas Makur Jaya | C1 |
| Jk. Gudang Timur | C1 |
| Jk. Gudang Barat | C1 |
| Pongong | C1 |
| Pengarang | C1 |
| Pt. Denta | C2 |
| Dakikan | C2 |
| Kadangan | C2 |
| Pangas | C2 |
| Benda Baru | C2 |
| Pakarang | C2 |
| Pempas | C1 |

Gambar 3 : Tampilan Layar Hasil Proses Cluster

Hasil klasterisasi C1 (Sporadis) terdapat 19 Kelurahan, C2 (Potensi) terdapat 13 Kelurahan, dan C3 (Endemis) terdapat 20 Kelurahan dari total 52 Kelurahan, seperti yang dapat dilihat pada grafik penyebaran DBD (gambar 4) dibawah ini.



Gambar 4 : Grafik Penyebaran DBD

5. KESIMPULAN

Melalui proses pengerjaan dan pengujian dalam penelitian ini, maka dapat ditarik kesimpulan bahwa metode klasterisasi k-means dapat digunakan untuk mengelompokkan daerah potensi. Dimana C1 (Sporadis) berpusat di Kelurahan Rawa Buntu, C2 (Potensi) berpusat di Kelurahan Pondok Ranji dan C3 (Endemis) berpusat di Kelurahan Serua Indah yang ditentukan secara random dari 52 Kelurahan.

Hasil klasterisasi C1 (Sporadis) terdapat 19 kelurahan, diantaranya adalah kelurahan pisangang, cirende, paku jaya, serua, cabe ilir, cabe udik, pondok aren, pondok jagung, sawah baru, rawa buntu, ciater, bambu apus, cipayung, jelupang, pondok jagung timur, pondok pucung, kademangan, rengas, rempoa.

Hasil klasterisasi C2 (Potensi) terdapat 13 kelurahan, diantaranya adalah kelurahan pamulang barat, pamulang timur, pondok betung, jombang, pondok ranji, pondok kacang timur, jurang mangu barat, jurang mangu timur, pondok karya, pondok kacang barat, pondok benda, benda baru, kedaung.

Hasil klasterisasi C3 (Endemis) terdapat 20 kelurahan, diantaranya adalah kelurahan paku alam, pondok jaya, parigi lama, setu, ciputat, serua indah, lengkong karya, lengkong wetan, bakti jaya, kranggan, pakulon, parigi baru, muncul, buaran, rawa mekar, lengkong gudang timur, lengkong gudang barat, serpong, cilenggang, babakan.

Beberapa saran yang dapat diberikan untuk pengembangan program lebih lanjut dengan harapan dapat menghasilkan penelitian yang lebih baik lagi di waktu yang akan datang dapat berupa, Program aplikasi clustering daerah potensi penyebaran penyakit Demam Berdarah ini dapat mencari nilai centroid pertama secara otomatis dibantu dengan algoritma tertentu, agar hasil cluster lebih optimal.

Diharapkan adanya penelitian dan kajian lebih lanjut mengenai pengembangan program aplikasi clustering yang menggunakan metode K-Means Clustering dimasa yang akan datang. Serta melakukan evaluasi secara berkala terhadap kinerja program sebagai bahan masukan untuk keperluan pengembangan program selanjutnya.

6. DAFTAR PUSTAKA

- [1] Setiawan, R., 2016, Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi, *Jurnal Lentera ICT ISSN: 2338-3143, Volume 3, Nomor 1*, hal.76-92.
- [2] Jajuli, M., Defiyanti, S., Rohmawati, N.W., 2015, Implementasi Algoritma K-Means Dalam Pengklasteran Mahasiswa Pelamar Beasiswa, *Jurnal Ilmiah Teknologi Informasi Terapan ISSN: 2407-3911, Volume 1, Nomor 2*, hal.62-68.
- [3] Abdurrahman, G., 2016, *Clustering Data Ujian Tengah Semester (UTS) Data Mining Menggunakan*

- Algoritma K-Means, *Skripsi*, FT, Universitas Muhammadiyah Jember.
- [4] Dhuhita, W.M.P., 2015, *Clustering Menggunakan Metode K-Means Untuk Menentukan Status Gizi Balita*, *Jurnal Informatika*, Volume 15, Nomor 2, hal.160-174.
- [5] Hastuti, N.F., 2013, Pemanfaatan Metode K-Means *Clustering* Dalam Penentuan Penerima Beasiswa, *Skripsi*, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Sebelas Maret Surakarta.
- [6] Nengsih, Warnia, S.Kom, M.Kom., Riveranda, Okta., 2016, *K-Means Analysis* Klasterisasi Kasus HIV AIDS di Indonesia, *Skripsi*, Program Studi Sistem Informasi, Politeknik Caltex Riau. Pekanbaru.
- [7] Ong, J.O., 2013, Implementasi Algoritma K-Means *Clustering* Untuk Menentukan Strategi Marketing President University, *Jurnal Ilmiah Teknik Industri ISSN: 1412-6869*, Volume 12, Nomor 1, hal.10-20.
- [8] Wardhani, A.K., 2016, Implementasi Algoritma K-Means Untuk Pengelompokan Penyakit Pasien Pada Puskesmas Kajen Pekalongan, *Jurnal Transformatika*, Volume 14, Nomor 1, hal.30-37.
- [9] Abeiza, E., Sumadikarta, I., 2014, Penerapan Algoritma K-Means pada Data Mining untuk memilih Produk dan Pelanggan Potensial, *Jurnal Satya Informatika*, Volume 1, Nomor 1, hal.12-22.