

PENGELOMPOKAN DATA KESEHATAN KOTA BANDUNG MENGGUNAKAN K-MEANS CLUSTERING
GROUPING OF HEALTH DATA IN CITY OF BANDUNG USING K-MEANS CLUSTERING

Silvia Latifah Putri¹, Dr, Purba Daru Kusuma, S.T., M.T.², Roswan Latuconsina, S.T., M.T.³

^{1,2,3}Prodi S1 Teknik Komputer, Fakultas Teknik Elektro, Universitas Telkom

¹silvialatifahp@student.telkomuniversity.ac.id, ²purbadaru@telkomuniversity.co.id,

³roswan@telkomuniversity.ac.id

Abstrak

Pada abad 21 ini era modern semakin maju, salah satunya yang banyak dibutuhkan sekarang adalah *Data Science* yang sangat berguna bagi berbagai perusahaan dalam mengumpulkan, mengkaji, dan menganalisa terhadap suatu data dan permasalahan yang ada. Salah satunya adalah *Dana Mining* dengan memanfaatkan *Clustering* (pengelompokan data). Pada penelitian ini yaitu mengelompokkan data kesehatan pada penyakit demam berdarah, diare, dan TB BTA+ yang sering terjadi di Kota Bandung berdasarkan jumlah penduduk dan jumlah pengidap penyakit demam berdarah, diare, dan TB BTA+ sesuai dengan jenis kelaminnya. Data yang digunakan berasal dari Dinas Kesehatan Kota Bandung dan Dinas Kependudukan dan Catatan Sipil Kota Bandung.

Pengelompokan data ini menggunakan metode *K-Means Clustering*. *K-Means Clustering* sendiri adalah pengelompokan data yang ada kedalam beberapa kelompok, dimana setiap satu *cluster* memiliki karakteristik yang sama. Perhitungan clustering memanfaatkan persamaan *Euclidean Distance* dimana jarak antar data dengan *centroid*. Penelitian ini bertujuan untuk melakukan analisa multiaspek atas data penyakit demam berdarah, diare, dan TB BTA+ dan membangun sebuah sistem berbasis *website* yang memiliki kemampuan untuk melakukan klasterisasi.

Kata kunci: Penyakit DBD, Penyakit Diare, Penyakit TB BTA+, *K-Means Clustering*, *Euclidean Distance*

Abstract

In the 21st century modern era is develop increasingly, one of which is much needed now is Data Science which is very useful for various companies collecting in collecting, reviewing, and analyzing existing data and problems. One of them is Dana Mining which is utilizing Clustering (grouping data). In this study, grouping health data on dengue fever, diarrhea, and tuberculosis which often occurs in City of Bandung based on the population and the number of people with dengue fever, diarrhea, and tuberculosis according their sex. The data used for this study came from Department Health Office City of Bandung and Department Population and Civil Registration Office City of Bandung.

This grouping data uses K-Means Clustering method. K-Means Clustering itself is grouping of data in several groups, where each cluster has the same characteristics. Clustering calculation uses the Euclidean Distance Space equation where the distance between data and centroid. This final project aims to analyze multifaceted diseases of dengue fever, diarrhea, and tuberculosis and build an interactive website that has ability to do classifying.

Keywords: *Dengue Fever, Diarrhea, Tuberculosis, K-Means Clustering, Euclidean Distance Space*

1. Pendahuluan

Kesehatan adalah keadaan seseorang sehat atau tidak ada gangguan mulai dari badan, dan sosial, dan seseorang dapat hidup produktif secara sosial dan ekonomis. Masih banyak masalah kesehatan yang terjadi di Indonesia yang memang masih menyandang predikat “negara berkembang”. Seperti kasus beberapa penyakit yang terjadi di Kota Bandung, salah satunya adalah penyakit demam berdarah, diare, dan TB BTA+. Di Kota Bandung sendiri memiliki rasio penderita penyakit yang cukup signifikan dari tahun ke tahun, dimana pada tahun 2014 sesuai dengan profile kesehatan yang dikeluarkan oleh Dinas Kesehatan Kota Bandung mempunyai rincian sebagai berikut[1]: Penyakit Demam Berdarah terdapat 3.132 penderita, penyakit Diare terdapat 54.7722 penderita, dan penyakit TB BTA+ dengan penderita baru terdapat 940 penderita. Pada tahun 2015 mempunyai rincian sebagai berikut[2]: Penyakit Demam Berdarah terdapat 3.648 penderita, penyakit Diare terdapat 55.361 penderita, dan TB BTA+ dengan penderita terdapat 1.023 penderita. Pada tahun 2016 mempunyai rincian sebagai berikut[3]: Penyakit Demam Berdarah terdapat 3.860 penderita, penyakit Diare terdapat 55.054 penderita, penyakit TB BTA+ dengan penderita baru terdapat 1.107 penderita. Pada tahun 2017 mempunyai rincian sebagai berikut[4]: Penyakit Demam Berdarah terdapat 1.816 penderita, penyakit Diare terdapat 55.004 penderita, penyakit TB BTA+ dengan penderita baru terdapat 1.003 penderita. Pada tahun 2018 memiliki rincian sebagai berikut[5]: Penyakit Demam Berdarah terdapat 2.826 penderita, penyakit Diare terdapat 58.019 penderita, dan penyakit TB BTA+ dengan penderita baru terdapat 2.056 penderita.

Data dikeluarkan oleh Dinas Kesehatan Kota Bandung. Permasalahan data kesehatan ini dapat dibantu dengan teknik data mining yaitu dengan memanfaatkan teknik *clustering*. Salah satu metode yang sering digunakan adalah *K-Means Clustering*. *K-Means Clustering* adalah metode yang sering digunakan oleh ilmuwan dan metode paling

sederhana[6]. Pada *K-Means* sendiri melakukan proses pengelompokannya sesuai dengan k yang dimasukkan, namun pada pencarian *centroid* mengalami sedikit kelemahan, yaitu harus ditentukan terlebih dahulu *centroid* mana yang akan digunakan, karena hasil dari *clustering* itu sendiri bergantung pada penentuan pusat *centroid*-nya. Penggunaan metode *K-Means Clustering* sendiri dapat diterapkan untuk hal pengelompokan data kesehatan yang ada di Dinas Kesehatan yang diharapkan berguna untuk mengetahui kecamatan mana yang mengalami penderita penyakit tertentu tertinggi.

2. Dasar Teori

2.1 Demam Berdarah

Demam Berdarah Dengue (DBD) adalah virus Dengue ditularkan kepada manusia melalui nyamuk *Aedes Aegypti* dan *Aedes Albopictus* yang biasa hidup dan disekitar rumah[7].

2.2 Diare

Diare adalah salah satu masalah kesehatan yang sering terjadi di Indonesia, salah satunya adalah penyakit Diare. Ciri-ciri seseorang terkena diare sendiri adalah penyakit yang membuat penderitanya sering buang air besar dengan kondisi tinja seperti air kencing, dimana volume menggunakan kamar mandi sering, selain itu tanda-tanda yang lain adalah mengalami dehidrasi, kulit menjadi keriput[8].

2.3 TB BTA+

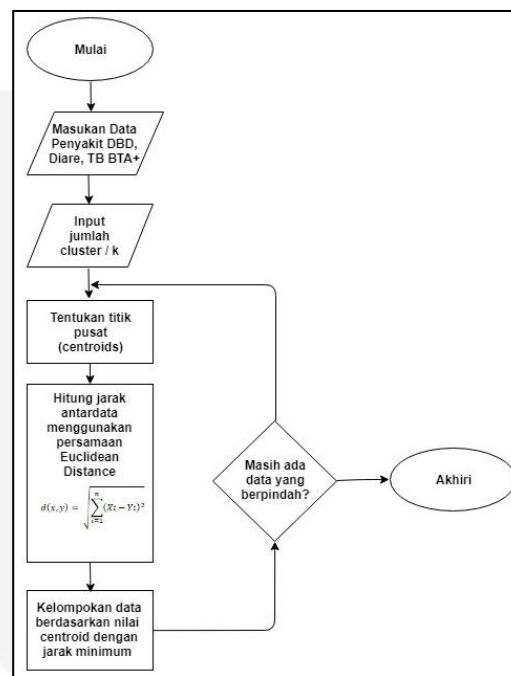
Micobacterium Tuberculosis Basil Tahan Asam (TB BTA+) adalah penyakit menular yang sebagian besar disebabkan oleh kuman. Biasanya kuman tersebut menyebar melalui udara pernapasan dan masuk kedalam paru-paru[9].

2.4 K-means Clustering

Menurut buku Rui Xu dan Wunsch II menyebutkan bahwasannya *K-Means* adalah salah satu metode yang paling populer dalam algoritma *clustering*, dimana *K-Means* bekerja mencari partisi terdekat atau optimal dari data dengan meminimalkan kriteria jumlah kesalahan kuadrat dengan prosedur iterasi yang optimal[10].

Tahapan penggunaan algoritma *K-Means Clustering* yang digunakan dalam penelitian ini seperti terlihat pada gambar 1 berikut ini.

1. Data penyakit DBD, Diare, dan TB BTA+ akan di unggah dimana data yang berupa variabel-variabel terkait berisi data numerik
2. Nilai k nanti akan ditentukan oleh pengguna secara random dimana tergantung terhadap kebutuhan yang ada
3. Sistem akan melakukan klasterisasi pada data yang telah diunggah diawal dan sudah ditentukan nilai k yang diinginkan.



Gambar 1. Flowchart K-Means Clustering

Pada proses penentuan jarak antara data dengan centroid dapat menggunakan persamaan *Euclidean Distance Space* sebagai berikut[11]:

$$d(x,y) = |x - y| = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2} ; i = 1,2,3, \dots, n \quad (1)$$

Keterangan:

X_i = objek x ke- i

Y_i = data y ke- i

n = banyaknya objek

Kemudian persamaan untuk pembaharuan suatu titik *centroid* dapat dilakukan dengan rumus berikut[12]:

$$\mu_k = \frac{1}{N_k} \sum_{q=1}^{N_k} X_q \quad (2)$$

Keterangan:

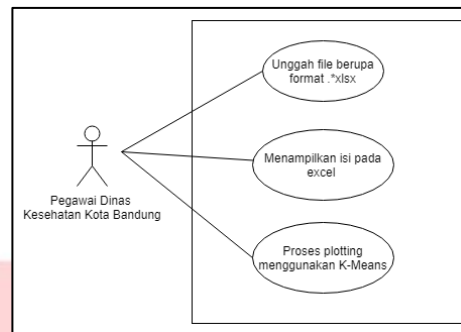
μ_k = titik centroid dari cluster ke-K

N_k = banyaknya data pada cluster ke-K

x_q = data ke-q pada cluster ke-K

3. Perancangan

Sistem yang akan dibuat pada Penelitian ini menggunakan data penyakit DBD, Diare, dan TB BTA+ di Kota Bandung dengan metode K-Means Clustering berbasis website. Website yang digunakan nantinya hanya dapat dilakukan oleh pegawai Dinas Kesehatan Kota Bandung sehingga mampu melakukan analisis terhadap data yang diunggah.



Gambar 2 Diagram use case sistem klasterisasi data kesehatan

Pada gambar 2 dijelaskan alur sistem klasterisasi sebagai berikut:

1. Pegawai Dinas Kesehatan Kota Bandung mengunggah data yang berupa file excel dengan variabel nama kecamatan, jumlah penduduk, dan jumlah pengidap penyakit pada penyakit DBD, Diare, dan TB BTA+
2. Pada sistem akan dikeluarkan isi pada file excel yang diunggah
3. Pegawai Dinas Kesehatan mendapatkan hasil klasterisasi yang memanfaatkan metode *K-Means Clustering* berupa plottingan.

4. Pengujian Sistem

Kebutuhan data yang dibutuhkan agar dapat berjalan dengan baik yang menghasilkan keluaran berupa klasterisasi kedalam beberapa *cluster*, adalah sebagai berikut: (1) data jumlah pengidap penyakit Demam Berdarah Dengue berdasarkan kecamatan dan jenis kelamin pada tahun 2014 sampai 2018 di Kota Bandung, (2) data jumlah pengidap penyakit Diare berdasarkan kecamatan dan jenis kelamin pada tahun 2014 sampai 2018 di Kota Bandung, (3) data jumlah pengidap penyakit TB BTA+ berdasarkan kecamatan dan jenis kelamin pada tahun 2014 sampai 2018 di Kota Bandung, (4) data jumlah penduduk berdasarkan kecamatan dan jenis kelamin pada tahun 2014 sampai 2018. Proses perhitungan k-means dapat dilakukan sebagai berikut:

Tabel 1 Data Penderita DBD Jenis Kelamin Laki-Laki dan Perempuan tahun 2014

Kecamatan	Penduduk Campur	Penderita Campur	Keterangan
Sukasari	71.280	111	
Sukajadi	96.218	86	C1
Cicendo	91.330	64	
Andir	97.793	72	
Cidadap	49.856	36	
Coblong	108.468	134	
Bandung Wetan	29.406	39	
Sumur Bandung	34.800	80	
Cibeunying Kaler	67.727	72	
Cibeunying Kidul	106.377	124	
Kiaracondong	118.888	172	
Batununggal	114.164	165	
Lengkong	70.759	236	
Regol	77.109	143	
Astanaanyar	71.893	96	
Bojongloa Kaler	116.140	106	C2
Bojongloa Kidul	81.286	100	
Babakan Ciparay	123.277	155	
Bandung Kulon	127.498	75	
Antapani	71.493	148	
Mandalajati	62.666	61	

Arcamanik	67.476	105	
Ujungberung	73.363	107	
Cinambo	23.186	28	
Cibiru	69.378	76	
Panyileukan	36.091	44	
Gedebage	36.091	52	
Rancasari	76.706	135	C3
Buahbatu	93.103	202	
Bandung Kidul	54.692	108	

Pertama kita harus menentukan *cluster* secara acak. Pada data diatas C1 bernilai (96.218 , 86) dan C2 bernilai (116.140 , 106) kemudian masukan rumus *Euclidean Distance* (persamaan 1) untuk menghitung jarak antar data satu dengan data lainnya sebagai berikut:

- a. Jarak C1 (96.218 , 86) dengan data pertama (71.280 , 111)
 (96.218 , 86) \leftrightarrow (71.280 , 111)
 $= \sqrt{(71.280 - 96.218)^2 + (111 - 86)^2} = 24.938,01253$
 $= \sqrt{(96.218 - 96.218)^2 + (111 - 111)^2} = 0$
- b. Jarak C2 (116.140 , 106) dengan data pertama (71.280 , 111)
 (116.140 , 106) \leftrightarrow (71.280 , 111)
 $= \sqrt{(71.280 - 116.140)^2 + (111 - 106)^2} = 44.860,00028$
- c. Jarak C3 (76.706 , 135) dengan data pertama (71.280 , 111)
 (76.706 , 135) \leftrightarrow (71.280 , 111)
 $= \sqrt{(71.280 - 76.706)^2 + (111 - 135)^2} = 5.426,053078$

Setelah melakukan perhitungan *centroid* satu, dua, dan tiga pada data kesatu hingga data ketiga puluh maka akan didapatkan hasil sebagai berikut:

Tabel 2 Jarak data pada *cluster* C1, C2, C3

Kecamatan	C1	C2	C3	Keterangan
Sukasari	24.938,01253	44.860,00028	5.426,053078	3
Sukajadi	0	19.922,01004	19.512,06153	1
Cicendo	4.888,049509	24.810,03555	14.624,17235	1
Andir	1.575,062221	18.347,0315	21.087,09411	1
Cidadap	46.362,02696	66.284,03696	26.850,18251	3
Coblong	12.250,09404	7.672,051095	31.762,00002	2
Bandung Wetan	66.812,01653	86.734,02588	47.300,09742	3
Sumur Bandung	61.418,00029	81.340,00416	41906.03609	3
Cibeunying Kaler	28.491,00344	48.413,01194	8.979,221013	3
Cibeunying Kidul	10.159,07107	9.763,016593	29.671,00204	2
Kiaracondong	22.670,16312	2.748,792462	42.182,01623	2
Batununggal	17946.17388	1.976,880624	37.458,01201	2
Lengkong	25.459,44188	45.381,1862	5.947,857597	3
Regol	19.109,08501	39.031,01754	403.0793966	3
Astanaanyar	24.325,00206	44.247,00113	4.813,158007	3
Bojongloa Kaler	19.922,01004	0	39.434,01066	2
Bojongloa Kidul	14.932,00656	34.854,00052	4.580,133732	3
Babakan Ciparay	27.059,08797	7.137,168206	46.571,00429	2
Bandung Kulon	31.280,00193	11.358,0423	50.792,03544	2
Antapani	24.725,07773	44.647,01975	5.213,016209	3
Mandalajati	33.552,00931	53.474,01893	14.040,19501	3
Arcamanik	28.742,00628	48.664,00001	9.230,048754	3
Ujungberung	22.855,00965	42.777,00001	3.343,117258	3
Cinambo	73.032,02303	92.954,03273	53.520,10696	3
Cibiru	26.840,00186	46.762,00962	7.328,23751	3
Panyileukan	60.025,01469	79.947,02404	40.513,1022	3
Gedebage	60.127,00961	80.049,01821	40.615,08481	3
Rancasari	19.512,06153	39.434,01066	0	3
Buahbatu	3.117,159123	23.037,20003	16.397,13688	1
Bandung Kidul	41.526,00583	61.448,00003	22.014,01656	3

Setelah mendapatkan nilai untuk masing-masing *cluster* satu, dua, dan tiga dan data sudah dikelompokkan masuk kedalam *cluster* satu, dua, dan tiga pada iterasi pertama kemudian akan diperbaharui *centroid* atau titik pusatnya untuk mengetahui apakah data tetap stabil atau mengalami perubahan. Setelah proses perhitungan *centroid* baru dilakukan maka akan tetap dilakukan seperti perhitungan diawal yaitu mencari nilai pada *centroid* satu, dua, dan tiga. Jika nilai dan keterangan penempatan *centroid* pada data sama atau tidak ada perubahan maka proses iterasi dihentikan, namun jika nilai dan keterangan penempatan *centroid* pada iterasi sebelum dan setelahnya berbeda maka perlu melakukan proses ulang perhitungan dan pengelompokan terhadap data.

Tabel 3 Hasil akhir cluster

Kecamatan	Cluster	Jarak dengan <i>centroid</i>
Sukasari	1	3.403,5078
Sukajadi	2	15.873,48303
Cicendo	1	16.646,58852
Andir	2	14.298,5284
Cidadap	3	12.109,72964
Coblong	2	3.623,467765
Bandung Wetan	3	8.340,301614
Sumur Bandung	3	2.946,389368
Cibeunying Kaler	1	6.956,653981
Cibeunying Kidul	2	5.714,445232
Kiaracondong	2	6.796,7469
Batununggal	2	2.073,022559
Lengkong	1	3.926,265007
Regol	1	2.425,625908
Astanaanyar	1	2.790,588989
Bojongloa Kaler	2	4.048,583343
Bojongloa Kidul	1	6.602,525321
Babakan Ciparay	2	11.185,60723
Bandung Kulon	2	15.406,62423
Antapani	1	3.190,638367
Mandalajati	1	12.017,63654
Arcamanik	1	7.207,512245
Ujungberung	1	1.320,548226
Cinambo	3	14.560,31128
Cibiru	1	5.305,668509
Panyileukan	3	1.553,326713
Gedebage	3	1.655.288975
Rancasari	1	2.022,569064
Buahbatu	1	18.419,69023
Bandung Kidul	3	16.945,79628

Pada tabel 4.0 dapat dilihat bahwasannya tidak perlu membutuhkan perhitungan iterasi selanjutnya dikarenakan nilai dan penempatan data pada *cluster* sudah stabil atau tidak ada data yang berpindah. Pada tabel diatas juga menunjukkan bahwasannya membutuhkan sekitar perhitungan 10 iterasi hingga nilai dan penempatan data stabil. Dari tabel 4.0 dapat disimpulkan bahwasannya pengelompokan data pengidap penyakit DBD pada tahun 2014 berjenis kelamin laki-laki dan perempuan dapat dikelompokkan menjadi 3 *cluster* dimana *cluster* 1 berisi kecamatan Sukasari, Cicendo, Cibeunying Kaler, Lengkong, Regol, Astaanyar, Bojongloa Kidul, Antapani, Mandalajati, Arcamanik, Ujungberung, Cibiru, Rancasari, dan Buahbatu. *Cluster* 2 berisi kecamatan Sukajadi, Andir, Coblong, Cibeunying Kidul, Kiaracondong, Batununggal, Bojongloa Kaler, Babakan Ciparay, dan Bandung Kulon. *Cluster* 3 berisi kecamatan Cidadap, Bandung Wetan, Sumur Bandung, Cinambo, Panyileukan, Gedebage, dan Bandung Kidul. Berdasarkan hasil klasterisasi diatas dapat dilihat bahwasannya titik pusat obeb pada *cluster* 1 adalah kecamatan Rancasari, *cluster* 2 adalah kecamatan Batununggal, dan *cluster* 3 adalah kecamatan Panyileukan.

5. Kesimpulan

Berdasarkan hasil klasterisasi data pengidap penyakit DBD berjenis kelamin laki-laki dan perempuan pada tahun 2014 di Kota Bandung bahwasannya data dapat diklasterisasi menggunakan 3 *cluster* berdasarkan kecamatan dan jenis kelamin. Hasil klasterisasi diatas dapat disimpulkan bahwasannya titik pusat untuk *cluster* pertama berpusat pada kecamatan Rancasari, *cluster* kedua berpusat pada kecamatan Batununggal, dan *cluster* berpusat pada kecamatan

Panyileukan. Dari hasil klasterisasi diatas dapat diketahui tingkat keakuratan kebenaran algoritma K-Means dalam pengelompokan data pengidap penyakit DBD sebesar $\frac{16}{30} = 53,3333\%$

Daftar Pustaka

- [1] Dinas Kesehatan Kota Bandung, "Profile Kesehatan Kota Bandung Tahun 2014," 2014. [Online]. Available: <https://clouddinkes.bandung.go.id/index.php/s/DbmLfdehNX6P0Pw/download>. [Diakses 29 April 2019].
- [2] Dinas Kesehatan Kota Bandung, "Profile Kesehatan Kota Bandung Tahun 2015," 2015. [Online]. Available: <https://clouddinkes.bandung.go.id/index.php/s/nTjqq5R3BT55TS0/download>. [Diakses 19 April 2019].
- [3] Dinas Kesehatan Kota Bandung, "Profile Kesehatan Kota Bandung Tahun 2016," 2016. [Online]. Available: <https://clouddinkes.bandung.go.id/index.php/s/ltyhkhCwkERFIXT/download>. [Diakses 19 April 2019].
- [4] Dinas Kesehatan Kota Bandung, "Profile Kesehatan Kota Bandung Tahun 2017," 2017. [Online]. Available: <https://clouddinkes.bandung.go.id/index.php/s/aH94RDEeFn1dLd7/download>. [Diakses 19 April 2019].
- [5] Dinas Kesehatan Kota Bandung, "Profile Kesehatan Kota Bandung Tahun 2018," 2018. [Online]. [Diakses 31 Januari 2019].
- [6] T. Alfina, B. Santosa dan A. R. Barakbah, "Analisa Perbandingan Metode Hierarchical Clustering, K-means dan Gabungan Keduanya dalam Cluster Data (Studi kasus: Problem Kerja Praktek Jurusan Teknik Industri ITS)," *JURNAL TEKNIK ITS*, vol. 1, no. ISSN: 2301-9271, 2012.
- [7] Kementerian Kesehatan Republik Indonesia, "Demam Berdarah Dengue (DBD)," 25 April 2017. [Online]. Available: <http://www.depkes.go.id/development/site/depkes/index.php?cid=1-17042500004&id=demam-berdarah-dengue-dbd-.html>. [Diakses 28]
- [8] Kemeterian Kesehatan Republik Indonesia, "Situasi DIARE di Indonesia," Juni 2011. [Online]. Available: <http://www.depkes.go.id/download.php?file=download/pusdatin/buletin/buletin-diare.pdf>. [Diakses 23 Juni 2019]
- [9] Kementerian Kesehatan Republik Indonesia, "Info Datin:Tuberkulosis," 2018. [Online]. Available: <http://www.depkes.go.id/resources/download/pusdatin/infodatin/infodatin%20tuberkulosis%202018.pdf>. [Diakses 22 Juni 2019].
- [10] R. Xu dan D. C. Wunsch II, *Clustering*, Hoboken, New Jersey: A John Wiley & Sons, INC., Publication, 2009.
- [11] N. Bouhmala, "How Good Is The Euclidean Distance Metric For The Clustering Problem," dalam *2016 5th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)*, Kumamoto, Japan, 2016.
- [12] K. Rajeswari, O. Acharya, M. Sharma, M. Kopnar dan K. Karandikar, "Improvement in k-Means Clustering Algorithm Using Data Clustering," dalam *2015 International Conference on Computing Communication Control and Automation*, Pune, India, 2015.