

KLASTERISASI NEGARA PENDAFTAR PATEN DI INDONESIA MENGGUNAKAN K-MEANS**¹Deny Novianti, ² Nicodias Palasara**¹Ilmu Komputer, Ilmu Komputer, STMIK Nusa Mandiri
Jl. Damai No. 08, Warung Jati Barat (Margasatwa)²Sistem Informasi, Sistem Informasi, STMIK Nusa Mandiri
Jl. Damai No. 08, Warung Jati Barat (Margasatwa)Email: ddenynovianti@gmail.com, nico.ncp@nusamandiri.ac.id

(Diterima: 19 Juli 2019, direvisi: 8 Agustus 2019, disetujui: 18 Agustus 2019)

ABSTRAK

Dengan semakin banyaknya paten yang mendaftar dari berbagai negara di Asia Tenggara, dapat diketahui negara manakah yang paling banyak mendaftarkan paten dari tahun 2016-2018. Berdasarkan data yang ada, tidak disebutkan negara mana sajakah yang termasuk dalam pendaftaran patennya yang tertinggi ataupun terendah. Sehingga tidak dapat diketahui Negara mana saja yang berpotensi paling banyak mendaftarkan paten untuk ke depannya. Data yang digunakan untuk dikelompokkan menggunakan metode *K-Means Cluster Analysis* adalah *database KI* dengan parameternya yaitu jenis Paten, yaitu paten dan paten sederhana yang ada dari tahun 2016 sampai dengan tahun 2018. Tahapan dalam menggunakan metode tersebut dimulai dari penentuan titik pusat awal, perhitungan jarak antar cluster dan menentukan besarnya rasio. Berdasarkan hasil penelitian ini, diperoleh hasil India dan Singapore adalah Negara tertinggi selama tiga tahun terakhir dalam pendaftaran paten di Indonesia. Kemudian ada China, Malaysia, dan Thailand adalah Negara dengan intensitas sedang dalam pendaftaran paten di Indonesia selama tiga tahun terakhir, dan yang terakhir terdapat 18 negara yang masuk dalam kategori terendah dalam pendaftaran paten di Indonesia dalam tiga tahun terakhir.

Keywords: Clustering, K-Means, HKI, Patents, Data Mining

1 PENDAHULUAN

Menurut Direktorat Jenderal Kekayaan Intelektual Kementrian Hukum dan HAM, Paten diartikan sebagai hak *eksklusif inventor* atas invensi di bidang teknologi untuk selama waktu tertentu melaksanakan sendiri atau memberikan persetujuan kepada pihak lain untuk melaksanakan invensinya. Invensi adalah ide inventor yang dituangkan ke dalam suatu kegiatan pemecahan masalah yang spesifik di bidang teknologi, dapat berupa produk atau proses atau penyempurnaan dan pengembangan produk atau proses [1].

Dikenal dua macam bentuk paten, yaitu paten biasa dan paten sederhana. Paten merupakan hak eksklusif yang diberikan oleh negara kepada inventor atas hasil invensinya di bidang teknologi yang untuk selama waktu tertentu melaksanakan sendiri invensinya atau memberikan persetujuannya kepada pihak lain untuk melaksanakannya. Paten sederhana adalah hak eksklusif yang diberikan oleh negara kepada inventor atas hasil-hasil invensinya berupa produk atau alat yang baru dan memiliki nilai kegunaan praktis disebabkan oleh bentuk, konfigurasi, konstruksi atau komponennya [2].

Secara umum, hal-hal ataupun subjek yang dapat dipatenkan adalah proses, mesin, barang (yang di produksi serta dapat digunakan), metode bisnis, perangkat lunak, teknik medis, teknik olahraga, dan lain sebagainya. Syarat hasil penemuan di Indonesia yang dapat dipatenkan yaitu Invensi memiliki aspek kebaruan, atau dengan kata lain invensi itu belum pernah diungkapkan sebelumnya, Invensi harus mengandung langkah inventif, yaitu mengandung langkah yang tidak diduga sebelumnya bagi seseorang dengan keahlian tertentu di bidang teknik, Invensi dapat diterapkan dalam industri[2]. Semakin banyaknya paten yang masuk dari berbagai negara di Asia Tenggara, maka akan dapat diketahui kelompok negara mana sajakah yang memiliki pendaftaran paten terbanyak selama 3 tahun terakhir. Pada penelitian ini K-Means digunakan untuk mengelompokkan negara mana sajakah yang pendaftaran patennya tertinggi dan terendah .

2 TINJAUAN PUSTAKA

Data Mining memiliki beberapa pandangan, seperti *knowledge discover* ataupun *pattern recognition*. Kedua istilah tersebut sebenarnya memiliki ketepatan masing-masing, istilah *knowledge discovery* atau penemuan pengetahuan tepat karena digunakan karena tujuan utama dari data mining memang untuk mendapat pengetahuan yang masih tersembunyi di dalam bongkahan data [5]. Menurut Widodo dalam [6], Data mining adalah analisa terhadap data untuk menemukan hubungan yang jelas serta menyimpulkannya yang belum diketahui sebelumnya dengan cara terkini dipahami dan berguna bagi pemilik data tersebut. Data mining juga merupakan metode yang digunakan dalam pengolahan data berskala besar oleh karena itu data mining memiliki peranan yang sangat penting dalam beberapa bidang kehidupan diantaranya yaitu bidang industri, bidang keuangan, cuaca, ilmu dan teknologi. Dalam data mining juga terdapat metode metode yang dapat digunakan seperti klasifikasi, clustering, regresi, seleksi variabel, dan market basket analisis [7].

Cluster Analysis merupakan salah satu metode objek mining yang bersifat tanpa latihan (*unsupervised analysis*), sedangkan *K-Means Cluster Analysis* merupakan salah satu metode cluster analysis non hirarki yang berusaha untuk mempartisi objek yang ada ke dalam satu atau lebih *cluster* atau kelompok objek berdasarkan karakteristiknya, sehingga objek yang mempunyai karakteristik yang sama dikelompokkan dalam satu *cluster* yang sama dan objek yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam *cluster* yang lain [8]. Analisis Pengelompokan/ *Clustering* merupakan proses membagi data dalam suatu himpunan ke dalam beberapa kelompok yang kesamaan datanya dalam suatu kelompok lebih besar dari pada kesamaan data tersebut dengan data dalam kelompok lain [9]. Potensi clustering adalah dapat digunakan untuk mengetahui struktur dalam data yang dapat dipakai lebih lanjut dalam berbagai aplikasi secara luas seperti klasifikasi, pengolahan gambar, dan pengenalan pola [10].

Metode ini telah banyak digunakan dalam berbagai penelitian, seperti yang pernah dilakukan oleh [9], Untuk melakukan penilaian terhadap hasil ekspor buah-buahan berdasarkan negara tujuan dapat menerapkan metode clustering K-Means. Data diolah untuk memperoleh nilai dari produksi ekspor buah-buahan berdasarkan negara tujuan. Data tersebut diolah menggunakan *Rapidminer* untuk ditentukan nilai centroid dalam 3 cluster yaitu cluster tingkat ekspor tinggi, cluster tingkat ekspor sedang dan cluster tingkat ekspor rendah. Sehingga diperoleh penilaian berdasarkan indeks ekspor buah-buahan dengan 2 negara cluster tingkat ekspor tinggi yakni India dan Pakistan, 3 negara cluster tingkat ekspor sedang yakni Singapura, Bangladesh dan Negara lainnya dan 6 negara cluster tingkat ekspor rendah yakni Hongkong, Tiongkok, Malaysia, Nepal, Vietnam dan Iran. Hasil yang dari penelitian dapat digunakan untuk mengetahui jumlah ekspor buah-buahan menurut negara tujuan. Penelitian selanjutnya dilakukan [11], salah satu proses data mining adalah Clustering. Atribut yang digunakan dalam pengelompokan prestasi siswa adalah Nama, Ekstrakurikuler, Nilai yang meliputi Nilai Tugas, Nilai Uts, Nilai Uas, jumlah ketidakhadiran siswa (absensi), dan Nilai sikap. Studi kasus pada 20 siswa dengan perhitungan jarak menggunakan manhattan distance, chbychep distance dan euclidian distance menghasilkan akurasi sebesar 67%. Penelitian selanjutnya dilakukan oleh [12], algoritma K-Means yang merupakan metode data clustering non hirarki yang mempartisi data ke dalam cluster sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok lain. Tujuan dari penelitian ini adalah untuk mengelompokkan data obat-obatan di Rumah Sakit Umum Daerah Pekanbaru yang dapat digunakan sebagai referensi dalam pengambilan keputusan dalam perencanaan dan pengendalian pasokan medis di rumah sakit tersebut. Berikut hasil penelitian yang dilakukan oleh [8], Algoritma K-Means tidak terpengaruh terhadap urutan objek yang digunakan, hal ini dibuktikan ketika penulis mencoba menentukan secara acak titik awal pusat cluster dari salah satu objek pada permulaan perhitungan. Jumlah keanggotaan cluster yang dihasilkan berjumlah sama ketika menggunakan objek yang lain sebagai titik awal pusat cluster tersebut. Penelitian yang dilakukan [13], untuk melakukan penilaian terhadap pengelompokan jumlah wisatawan ke objek wisata unggulan di Prov. DKI Jakarta dapat diterapkan dengan metode clustering K-Means. Data diolah untuk memperoleh jumlah wisatawan yang berkunjung ke objek wisata unggulan di Prov. DKI Jakarta. Data tersebut diolah menggunakan Aplikasi XLSTAT. Variable yang digunakan jumlah pengunjung. Data diolah dengan melakukan K-Means yang dikelompokkan menjadi 3 cluster yaitu cluster tinggi (C1) yakni jumlah kunjungan wisatawan tinggi, cluster sedang (C2) yakni jumlah kunjungan wisatawan sedang dan cluster

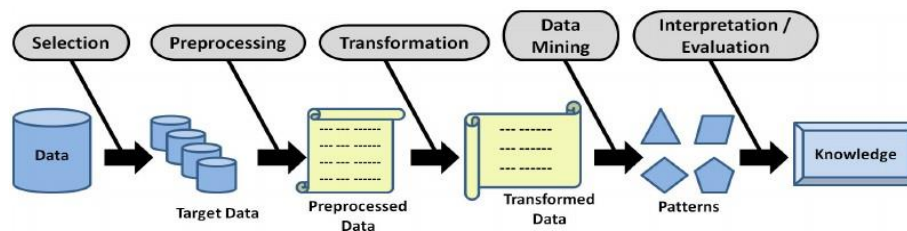
rendah (C3) yakni jumlah kunjungan wisatawan rendah. Centroid data C1= 15.438.488, Centroid data C2= 4.464.577 dan Centroid data C3= 342.332. Penelitian selanjutnya oleh [6], dalam penerapan ini, digunakan penerapan clustering dengan menggunakan algoritma k-means. Dari data yang diolah dengan sampel data yang diambil di Swalayan Fadhilla Bengkulu, maka menghasilkan dua jenis kelompok data. Yaitu data penjualan rendah dan data penjualan tinggi. Sehingga dengan adanya pengelompokan data ini pihak swalayan Fadhilla dapat mengetahui jenis barang yang laris terjual dan tidak. Sehingga barang yang ada di gudang tidak menumpuk. Berikutnya penelitian yang dilakukan[14], Pengolahan data dengan menggunakan metode clustering dengan algoritma K-Means yakni usia, jenis kendaraan dan jenis pelanggaran. Data kemudian ditransformasi untuk menentukan jarak pengelompokan kemudian diproses dengan tools Matlab dan menghasilkan pusat/cluster antar variable. Dengan pengelompokan objek diperoleh hasil yakni usia diantara 17 sampai dengan 37 tahun, yang melakukan pelanggaran dengan tidak menggunakan sefty belt dan melanggar rambu lalu lintas lebih banyak menggunakan sepeda motor dan mobil.

3 METODE PENELITIAN

Data penelitian ini menggunakan data dari data base KI tahun 2016-2018 yang terdiri dari 26 negara. Variabel yang digunakan adalah paten dan paten sederhana. Data akan diolah dengan melakukan clustering pendaftaran paten berdasarkan 3 cluster yakni pendaftaran paten tertinggi, pendaftaran paten sedang dan pendaftaran paten terendah.

3.1. Knowledge Discovery In Database (KDD)

KDD merupakan salah satu metode yang dapat digunakan untuk mendapatkan pengetahuan yang berasal dari database yang tersedia. Untuk lebih jelasnya, berikut gambaran dari metode KDD [15][12].



Gambar 1 Metode knowledge discovery in database

1. Selection

Penentuan variable untuk menghindari terjadinya data kembar dan pengulangan

2. Preprocessing

Terbagi menjadi 2 bagian, yaitu:

a. Cleaning

Menghapus data tidak terpakai serta mengatasi data yang tidak berpendirian

b. Integration

atribut yang memncari entitas yang unik

3. Transformation

Melakukan perubahan data dalam data mining

4. Data mining

Tahap untuk mencari nilai seperti informasi yang tidak diketahui secara manual dari suatu clustering

5. Evaluation/ Interpretation

Hasil dari tahap ini adalah melakukan pemeriksaan terhadap hasil yang sudah dikelola dari tahapan-tahapan diatas.

6. Knowledge

Bentuk yang muncul akan dijelaskan kepada pengguna. Dan hasil dalam proses pengolahan ini dapat dipahami dan dimengerti oleh orang lain

2.2. Proses Algoritma K-Means[12]

1. Ditentukan K untuk jumlah cluster.

2. K sebagai titik *cluster* yang disebut *centroid*, untuk menghitung *centroidcluster* ke-*i* selanjutnya, menggunakan cara seperti dibawah ini:

$$v = \frac{\sum_{i=1}^n X_i}{n} : i = 1, 2, 3, \dots, n \quad (1)$$

3. Jarak tiap obyek ke tiap-tiap *centroid* dari tiap-tiap *cluster* menggunakan perhitungan[9], dengan cara dibawah ini:

$$d(x_i, c_j) = ||x_i - c_j|| = \sqrt{\sum_{k=1}^n (x_{ik} - c_{jk})^2} : i = 1, 2, 3, \dots, n \quad (2)$$

4. Pemberian lokasi dari obyek ke dalam *centroid* yang terdekat, dalam tiap-tiap *cluster* pada iterasi dapat melakukan dengan cara *hardk-means* yaitu tiap obyek ditunjuk sebagai anggota *cluster* dengan cara menghitung jarak ke terdekat dari titik *cluster*.

5. Melakukan iterasi yang selanjutnya menentukan titik *centroid* baru dengan cara kesamaan.

6. lalu lakukan langkah tiga jika letak *centroid* terbaru tidak sama.

4 HASIL DAN PEMBAHASAN

Data dari database KI tahun 2016-2018. Parameter yang digunakan adalah tahun dan jenis paten. Proses *cluster* dilakukan untuk mengetahui Negara mana saja yang mendaftarkan patennya paling banyak berdasarkan jenis patennya.

4.1 Selection

Dilakukan proses pemilihan data yang akan digali (*field* yang dibutuhkan dalam proses *data mining*). Berikut ini adalah Tabel 1 meruapakan data awal yang akan digunakan dalam penelitian ini.

Tabel 1 Data Selection

No	Negara	Jenis	2016	2017	2018
1	Armenia	Paten	0	0	2
2	Azerbaijan	Paten	0	0	7
3	Brunei Darussalam	Paten Sederhana	0	0	0
4	China	Paten Sederhana	0	0	0
6	Georgia	Paten	1	1	0
7	Georgia	Paten Sederhana	0	0	0
8	Hong Kong	Paten	40	42	11
9	Hong Kong	Paten Sederhana	0	0	0
11	India	Paten	1808	1508	1038
12	India	Paten Sederhana	0	0	0
14	Indonesia	Paten	10366	10876	10187
15	Indonesia	Paten Sederhana	0	0	0
17	Iran (Islamic Republic of)	Paten	0	2	0
18	Iraq	Paten	0	0	2
19	Israel	Paten	202	178	174
20	Israel	Paten Sederhana	0	0	0
21	Japan	Paten	23610	23548	22229
22	Japan	Paten Sederhana	1	0	0
24	Kazakhstan	Paten Sederhana	0	0	0
25	Malaysia	Paten	530	273	550
26	Malaysia	Paten Sederhana	0	0	0
28	Oman	Paten	0	2	0
29	Philippines	Paten	76	11	0
30	Philippines	Paten Sederhana	0	0	0
32	Qatar	Paten	0	0	1

33	Qatar	Paten Sederhana	0	0	0
34	Republic of Korea	Paten	3617	3822	5655
35	Republic of Korea	Paten Sederhana	0	0	0
37	Russian Federation	Paten	463	225	109
38	Russian Federation	Paten Sederhana	0	0	0
40	Saudi Arabia	Paten	26	0	0
41	Saudi Arabia	Paten Sederhana	0	0	0
42	Singapore	Paten	343	674	1066
43	Singapore	Paten Sederhana	0	0	0
45	Sri Lanka	Paten	0	1	0
46	Thailand	Paten	65	33	82
47	Thailand	Paten Sederhana	0	0	0
49	Turkey	Paten	34	38	13
50	Turkey	Paten Sederhana	0	0	0
52	United Arab Emirates	Paten	96	3	17
53	United Arab Emirates	Paten Sederhana	0	0	0
54	Viet Nam	Paten	0	0	1
55	Viet Nam	Paten Sederhana	0	0	0

Data Negara yang mendaftarkan paten berdasarkan jenis dan juga tahun pendaftarannya dapat dilihat pada Tabel 1.

4.2 Preprocessing

4.2.1 Data Cleaning

Pada tahap ini yang dilakukan adalah menghilangkan atribut yang tidak relevan atau tidak konsisten. Maka atribut yang tidak relevan akan dihilang seperti Tabel 2 dibawah ini:

Tabel 2 Data Cleaning

Negara	2016		2017		2018	
	Paten	Paten Sederhana	Paten	Paten Sederhana	Paten	Paten Sederhana
Armenia	0	0	0	0	2	0
Azerbaijan	0	0	0	0	7	0
Brunei	0	0	0	0	0	0
Darussalam	0	0	0	0	0	0
China	0	15	0	70	0	29
Georgia	1	0	1	0	0	0
Hong Kong	40	1	42	1	11	1
India	1808	5	1508	27	1038	0
Iran						
(Islamic Republic of)	0	0	2	0	0	0
Iraq	0	0	0	0	2	0
Israel	202	0	178	0	174	0
Kazakhstan	0	0	0	0	0	0
Malaysia	530	89	273	8	550	4
Oman	0	0	2	0	0	0
Philippines	76	0	11	0	0	0
Qatar	0	0	0	0	1	0
Russian Federation	463	0	225	0	109	3
Saudi Arabia	26	0	0	0	0	0
Singapore	343	3	674	0	1066	6
Sri Lanka	0	0	1	0	0	0

Turkey	34	2	38	0	13	0
United Arab Emirates	96	0	3	0	17	0
Viet Nam	0	0	0	0	1	0

Dalam penelitian ini, data yang dihilangkan ada tiga Negara, yaitu Indonesia, Jepang, dan Korea Selatan. Negara-negara tersebut dihilangkan karena datanya *outlier* (jauh lebih tinggi dari yang lain).

4.2.2 Data Integration

Dilakukan terhadap atribut-atribut yang mengidentifikasi entitas yang unik. Pada tabel ini variabel tahun dan jenis paten yang akan di ubah. Untuk perubahan data tersebut dapat dilihat pada Tabel 3 berikut ini:

Tabel 3 Data Integration

Negara	x1	x2	x3	x4	x5	x6
Armenia	0	0	0	0	2	0
Azerbaijan	0	0	0	0	7	0
Brunei Darussalam	0	0	0	0	0	0
China	0	15	0	70	0	29
Georgia	1	0	1	0	0	0
Hong Kong	40	1	42	1	11	1
India	1808	5	1508	27	1038	0
Iran (Islamic Republic of)	0	0	2	0	0	0
Iraq	0	0	0	0	2	0
Israel	202	0	178	0	174	0
Kazakhstan	0	0	0	0	0	0
Malaysia	530	89	273	8	550	4
Oman	0	0	2	0	0	0
Philippines	76	0	11	0	0	0
Qatar	0	0	0	0	1	0
Russian Federation	463	0	225	0	109	3
Saudi Arabia	26	0	0	0	0	0
Singapore	343	3	674	0	1066	6
Sri Lanka	0	0	1	0	0	0
Thailand	65	28	33	14	82	8
Turkey	34	2	38	0	13	0
United Arab Emirates	96	0	3	0	17	0
Viet Nam	0	0	0	0	1	0

Dari yang awalnya variable jenis da tahun, di ubah menjadi data x1, ..., x6. Pertama x1 dan x2 di ubah untuk jenis paten tahun 2016, kemudian x3 dan x4 untuk jenis paten tahun 2017, dan yang terakhir x5 dan x6 untuk jenis paten tahun 2018.

4.3 Transformation

Pada tahap *transformation* ini dilakukan transformasi data-data yang terdapat pada tabel. Data yang telah di *transformasi* dapat dilihat pada Tabel 4 berikut:

Tabel 4 Transformation

Negara	x1	x2	x3	x4	x5	x6
Armenia	0.00	0.00	0.00	0.00	0.00	0.00
Azerbaijan	0.00	0.00	0.00	0.00	0.02	0.00
Brunei Darussalam	0.00	0.00	0.00	0.00	0.00	0.00
China	0.00	0.62	0.00	1.52	0.00	0.37
Georgia	0.00	0.00	0.00	0.00	0.00	0.00
Hong Kong	0.10	0.04	0.10	0.02	0.03	0.01
India	4.38	0.21	3.66	0.59	2.52	0.00
Indonesia	25.11	92.07	26.37	96.18	24.76	99.04
Iran (Islamic Republic of)	0.00	0.00	0.00	0.00	0.00	0.00
Iraq	0.00	0.00	0.00	0.00	0.00	0.00
Israel	0.49	0.00	0.43	0.00	0.42	0.00
Japan	57.20	1.81	57.10	0.93	54.03	0.21
Kazakhstan	0.00	0.00	0.00	0.00	0.00	0.00
Malaysia	1.28	3.66	0.66	0.17	1.34	0.05
Oman	0.00	0.00	0.00	0.00	0.00	0.00
Philippines	0.18	0.00	0.03	0.00	0.00	0.00
Qatar	0.00	0.00	0.00	0.00	0.00	0.00

Republic of Korea	8.76	0.25	9.27	0.28	13.74	0.10
Russian Federation	1.12	0.00	0.55	0.00	0.26	0.04
Saudi Arabia	0.06	0.00	0.00	0.00	0.00	0.00
Singapore	0.83	0.12	1.63	0.00	2.59	0.08
Sri Lanka	0.00	0.00	0.00	0.00	0.00	0.00
Thailand	0.16	1.15	0.08	0.30	0.20	0.10
Turkey	0.08	0.08	0.09	0.00	0.03	0.00
United Arab Emirates	0.23	0.00	0.01	0.00	0.04	0.00
Viet Nam	0.00	0.00	0.00	0.00	0.00	0.00

Transformasi yang dilakukan pada tabel sebelumnya adalah mengubah datanya menjadi bentuk %, ini bertujuan untuk memudahkan dalam melakukan perhitungan nantinya.

4.4 Data Mining

4.4.1 Proses Training

1. Iterasi 1

Tentukan K jumlah pusat cluster secara acak. Pada percobaan pertama ini ditentukan 3 data secara acak sebagai titik pusat awal untuk perhitungan jarak dari seluruh kelompok *cluster* yang akan dibentuk.

Jumlah *cluster* = 3 (diambil secara acak)

Jumlah data = 23

Jumlah atribut = 6

Data acak tersebut dapat dilihat pada Tabel 5 berikut:

Tabel 5 Pusat Cluster Iterasi 1

Negara	x1	x2	x3	x4	x5	x6	Cluster
Russian Federation	1.12	0.00	0.55	0.00	0.26	0.04	C1
Thailand	0.16	1.15	0.08	0.30	0.20	0.10	C2
Hong Kong	0.10	0.04	0.10	0.02	0.03	0.01	C3

Dari Tabel 5 dapat dilihat bahwa masing-masing data memiliki nilai pusat cluster yang terlihat pada Tabel 6.

Tabel 6 Jarak Antar Pusat Cluster Iterasi 1

C1	C2	1,60
C1	C3	1,14
C2	C3	1,16

2. Hitung jarak tiap data dengan masing-masing *cluster* pusat dengan menggunakan persamaan (2) yaitu persamaan *Euclidean Distance*. Hasil perhitungan tersebut dapat dilihat pada Tabel 7 berikut ini.

Tabel 7 Perhitungan Euclidean Distance Iterasi 1

Negara	C1	C2	C3
Armenia	1.27	1.22	0.15
Azerbaijan	1.27	1.22	0.15
Brunei Darussalam	1.28	1.22	0.15
China	2.10	1.38	1.65
Georgia	1.27	1.22	0.15
Hong Kong	1.14	1.16	0.00
India	5.08	6.08	6.13
Iran (Islamic Republic of)	1.27	1.22	0.15
Iraq	1.27	1.22	0.15
Israel	0.66	1.31	0.65
Kazakhstan	1.28	1.22	0.15
Malaysia	3.82	3.03	4.07
Oman	1.27	1.22	0.15
Philippines	1.10	1.21	0.13
Qatar	1.28	1.22	0.15
Russian Federation	0.00	1.60	1.14
Saudi Arabia	1.22	1.22	0.12
Singapore	2.59	3.12	3.08
Sri Lanka	1.27	1.22	0.15
Thailand	1.60	0.00	1.16
Turkey	1.16	1.13	0.05
United Arab Emirates	1.06	1.21	0.17
Viet Nam	1.28	1.22	0.15

Anggota dipilih dari yang terkecil diantara 3 *cluster* jika terkecil pada bagian C1 maka termasuk sebagai anggota C1 yaitu sebanyak 3 data, jika terkecil pada bagian C2 maka termasuk sebagai anggota C2 yaitu sebanyak 3 data, dan jika terkecil pada bagian C3 maka termasuk sebagai anggota C3 yaitu sebanyak 17 data. Berikut adalah rasio pada iterasi ke 1 yang ditampilkan pada Tabel 8.

Tabel 8 Perhitungan Ratio Iterasi 1

Iterasi	Ratio
1	0.09
0	0

Karena rasio sebelumnya belum ada maka iterasi dilanjutkan.

3. Lakukan iterasi ke 2

Tentukan posisi *centroid* baru dengan cara menghitung rata-rata dari data yang ada pada centroid yang sama atau anggota yang sama. Hasil perhitungan tersebut dapat dilihat pada Tabel 9 sebagai berikut:

Tabel 9 Jarak Antar Pusat Cluster Iterasi 2

C1	C2	3,21
C1	C3	3,31
C2	C3	2,04

Kemudian hitung jarak tiap data dengan masing-masing *cluster* pusat. Perhitungannya sama dengan tahap perhitungan pada iterasi 1. Hasil dari perhitungan tersebut ditunjukkan pada Tabel 10.

Tabel 10 Perhitungan Euclidean Distance Iterasi 2

Negara	C1	C2	C3
Armenia	3.39	2.07	0.08
Azerbaijan	3.38	2.07	0.08
Brunei Darussalam	3.39	2.07	0.09
China	3.68	1.65	1.68
Georgia	3.39	2.07	0.08
Hong Kong	3.26	1.99	0.08
India	2.96	5.79	6.19
Iran (Islamic Republic of)	3.39	2.07	0.08
Iraq	3.39	2.07	0.08
Israel	2.62	1.94	0.70
Kazakhstan	3.39	2.07	0.09
Malaysia	3.89	2.27	4.11
Oman	3.39	2.07	0.08
Philippines	3.27	2.03	0.12
Qatar	3.39	2.07	0.08
Russian Federation	2.31	2.07	1.19
Saudi Arabia	3.35	2.06	0.05
Singapore	1.55	3.11	3.11
Sri Lanka	3.39	2.07	0.08
Thailand	3.31	0.89	1.20
Turkey	3.27	1.96	0.09
United Arab Emirates	3.23	2.02	0.17
Viet Nam	3.39	2.07	0.08

Anggota dipilih dari yang terkecil diantara 3 *cluster* jika terkecil pada bagian C1 maka termasuk sebagai anggota C1 yaitu sebanyak 2 data, jika terkecil pada bagian C2 maka termasuk sebagai anggota C2 yaitu sebanyak 3 data, dan jika terkecil pada bagian C3 maka termasuk sebagai anggota C3 yaitu sebanyak 18 data. Rasio pada iterasi ke 2 yang ditampilkan pada Tabel 11 berikut ini:

Tabel 11 Perhitungan Ratio Iterasi 2

Iterasi	Ratio
2	0.39
1	0.09

Karena hasil iterasi ke-2 tidak sama dengan iterasi ke-1, maka perlu dilakukan kembali perhitungan ke iterasi ke-3 dan seterusnya sampai mendapatkan hasil yang sama.

Lakukan iterasi ke 3

Tentukan posisi *centroid* baru dengan cara menghitung rata-rata dari data-data yang ada pada *centroid* yang sama atau anggota yang sama. Hasil perhitungan tersebut dapat dilihat pada Tabel 12 sebagai berikut:

Tabel 12 Jarak Antar Pusat Cluster Iterasi 3

C1	C2	4,16
C1	C3	4,34
C2	C3	1,99

Kemudian hitung jarak tiap data dengan masing-masing *cluster* pusat. Perhitungannya sama dengan tahap perhitungan pada iterasi 1 dan iterasi ke 2. Hasil dari perhitungan tersebut ditunjukkan pada Tabel 13.

5 Tabel 13 Perhitungan Euclidean Distance Iterasi 3

Negara	C1	C2	C3
Armenia	4.52	2.07	0.19
Azerbaijan	4.51	2.07	0.18
Brunei Darussalam	4.52	2.07	0.19
China	4.71	1.65	1.67
Georgia	4.52	2.07	0.19
Hong Kong	4.39	1.99	0.07
India	2.06	5.79	6.08
Iran (Islamic Republic of)	4.52	2.07	0.19
Iraq	4.52	2.07	0.19
Israel	3.75	1.94	0.60
Kazakhstan	4.52	2.07	0.19
Malaysia	4.40	2.27	4.06
Oman	4.52	2.07	0.19
Philippines	4.40	2.03	0.10
Qatar	4.52	2.07	0.19
Russian Federation	3.46	2.07	1.09
Saudi Arabia	4.49	2.06	0.15
Singapore	2.06	3.11	3.04
Sri Lanka	4.52	2.07	0.19
Thailand	4.37	0.89	1.18
Turkey	4.40	1.96	0.10
United Arab Emirates	4.36	2.02	0.12
Viet Nam	4.52	2.07	0.19

Anggota dipilih dari yang terkecil diantara 3 *cluster* jika terkecil pada bagian C1 maka termasuk sebagai anggota C1 yaitu sebanyak 2 data, jika terkecil pada bagian C2 maka termasuk sebagai anggota C2 yaitu sebanyak 3 data, dan jika terkecil pada bagian C3 maka termasuk sebagai anggota C3 yaitu sebanyak 18 data. Rasio pada iterasi ke 3 yang ditampilkan pada Tabel 14 berikut ini:

Tabel 14 Perhitungan Ratio Iterasi 3

Iterasi	Ratio
3	0.55
2	0.39

Karena hasil iterasi ke-3 tidak sama dengan iterasi ke-2, maka perlu dilakukan kembali perhitungan ke iterasi ke-4 dan seterusnya sampai mendapatkan hasil yang sama.

Lakukan iterasi ke 4

Tentukan posisi *centroid* baru dengan cara menghitung rata-rata dari data-data yang ada pada *centroid* yang sama atau anggota yang sama. Hasil perhitungan tersebut dapat dilihat pada Tabel 15 sebagai berikut:

6 Tabel 15 Jarak Antar Pusat Cluster Iterasi 4

C1	C2	4,16
C1	C3	4,34
C2	C3	1,99

Kemudian hitung jarak tiap data dengan masing-masing *cluster* pusat. Perhitungannya sama dengan tahap perhitungan pada iterasi 1, iterasi 2, dan iterasi 3. Hasil dari perhitungan tersebut ditunjukkan pada Tabel 16.

Tabel 16 Perhitungan Euclidean Distance Iterasi 4

Negara	C1	C2	C3
Armenia	4.52	2.07	0.19
Azerbaijan	4.51	2.07	0.19
Brunei Darussalam	4.52	2.07	0.20
China	4.71	1.65	1.67
Georgia	4.52	2.07	0.19
Hong Kong	4.39	1.99	0.08
India	2.06	5.79	6.08
Iran (Islamic Republic of)	4.52	2.07	0.19
Iraq	4.52	2.07	0.19
Israel	3.75	1.94	0.59
Kazakhstan	4.52	2.07	0.20
Malaysia	4.40	2.27	4.05
Oman	4.52	2.07	0.19
Philippines	4.40	2.03	0.10
Qatar	4.52	2.07	0.19
Russian Federation	3.46	2.07	1.09
Saudi Arabia	4.49	2.06	0.15
Singapore	2.06	3.11	3.03
Sri Lanka	4.52	2.07	0.19
Thailand	4.37	0.89	1.17
Turkey	4.40	1.96	0.10
United Arab Emirates	4.36	2.02	0.12
Viet Nam	4.52	2.07	0.19

Anggota dipilih dari yang terkecil diantara 3 *cluster* jika terkecil pada bagian C1 maka termasuk sebagai anggota C1 yaitu sebanyak 2 data, jika terkecil pada bagian C2 maka termasuk sebagai anggota C2 yaitu sebanyak 3 data, dan jika terkecil pada bagian C3 maka termasuk sebagai anggota C3 yaitu sebanyak 18 data. Rasio pada iterasi ke 4 yang ditampilkan pada Tabel 17 berikut ini:

Tabel 17 Perhitungan Ratio Iterasi 4

Iterasi	Ratio
4	0.55
3	0.55

Iterasi ke 4 dengan iterasi ke-3 hasilnya sudah sama maka tidak perlu lagi melanjutkan ke iterasi ke-5 atau cukup berhenti di iterasi ke-4.

Klasterisasi data pendaftaran paten yang dihitung dengan algoritma *k-means* di dapatkan hasil *cluster* nya setelah melakukan iterasi ke-4. Tabel 18 menjelaskan bahwa *Cluster 1* terdapat kelompok Negara dengan pendaftaran paten tertinggi yaitu Negara India dan Singapore. Tabel 19 menjelaskan bahwa *Cluster 2* adalah kelompok Negara dengan pendaftaran paten sedang yaitu Negara China, Malaysia, dan Thailand. Terakhir pada Tabel 20 menjelaskan bahwa pada *Cluster 3* merupakan pendaftaran paten terendah yang memiliki 18 negara yaitu Armenia, Azerbaijan, Brunei Darussalam, Georgia, Hongkong, Iran, Iraq, Israel, Kazakhstan, Oman, Philippines, Qatar, Russian Federation, Saudi Arabia, Sri Lanka, Turkey, United Arab Emirates, Vietnam.

Tabel 18 Anggota Cluster 1 (Negara Dengan Pendaftaran Paten Tertinggi)

Negara	x1	x2	x3	x4	x5	x6	Cluster
India	4.38	0.21	3.66	0.59	2.52	0.00	C1
Singapore	0.83	0.12	1.63	0.00	2.59	0.08	C1

Tabel 19 Anggota Cluster 2 (Negara Dengan Pendaftaran Paten Sedang)

Negara	x1	x2	x3	x4	x5	x6	Cluster
China	0.00	0.62	0.00	1.52	0.00	0.37	C2
Malaysia	1.28	3.66	0.66	0.17	1.34	0.05	C2
Thailand	0.16	1.15	0.08	0.30	0.20	0.10	C2

Tabel 20 Anggota Cluster 3 (Negara Dengan Pendaftaran Paten Terendah)

Armenia	0.00	0.00	0.00	0.00	0.00	0.00	C3
Azerbaijan	0.00	0.00	0.00	0.00	0.02	0.00	C3
Brunei Darussalam	0.00	0.00	0.00	0.00	0.00	0.00	C3
Georgia	0.00	0.00	0.00	0.00	0.00	0.00	C3

Hong Kong	0.10	0.04	0.10	0.02	0.03	0.01	C3
Iran (Islamic Republic of)	0.00	0.00	0.00	0.00	0.00	0.00	C3
Iraq	0.00	0.00	0.00	0.00	0.00	0.00	C3
Israel	0.49	0.00	0.43	0.00	0.42	0.00	C3
Kazakhstan	0.00	0.00	0.00	0.00	0.00	0.00	C3
Oman	0.00	0.00	0.00	0.00	0.00	0.00	C3
Philippines	0.18	0.00	0.03	0.00	0.00	0.00	C3
Qatar	0.00	0.00	0.00	0.00	0.00	0.00	C3
Russian Federation	1.12	0.00	0.55	0.00	0.26	0.04	C3
Saudi Arabia	0.06	0.00	0.00	0.00	0.00	0.00	C3
Sri Lanka	0.00	0.00	0.00	0.00	0.00	0.00	C3
Turkey	0.08	0.08	0.09	0.00	0.03	0.00	C3
United Arab Emirates	0.23	0.00	0.01	0.00	0.04	0.00	C3
Viet Nam	0.50	0.33	0.42	0.15	0.44	0.04	C3

7 KESIMPULAN

Hasil cluster data pendaftaran paten di Indonesia bahwa India dan Singapore adalah Negara tertinggi selama tiga tahun terakhir dalam pendaftaran paten di Indonesia. Kemudian ada China, Malaysia, dan Thailand adalah Negara dengan intensitas sedang dalam pendaftaran paten di Indonesia selama tiga tahun terakhir, dan yang terakhir terdapat 18 negara yang masuk dalam kategori terendah dalam pendaftaran paten di Indonesia dalam tiga tahun terakhir.

REFERENSI

- [1] Direktorat Jenderal Kekayaan Intelektual Kementrian Hukum & HAM R.I, “Pengenalan Paten - Direktorat Jenderal Kekayaan Intelektual.” [Online]. Available: <http://www.dgip.go.id/pengenalan-paten>. [Accessed: 16-Jul-2019].
- [2] Pusat Studi Energi Universitas Gadjah Mada, “PEMAHAMAN PATEN – Pusat Studi Energi,” 2012. [Online]. Available: <https://pse.ugm.ac.id/389/>. [Accessed: 16-Jul-2019].
- [3] I. Method, K. C. Based, S. Value, W. Interface, C. Study, and I. U. M. M. Magelang, “Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang,” vol. 18, no. 1, pp. 76–82, 2015.
- [4] A. K. Wardhani, “Implementasi Algoritma K-Means untuk Pengelompokan Penyakit Pasien pada Puskesmas Kajen Pekalongan,” *J. Transform.*, vol. 14, no. 1, pp. 30–37, 2016.
- [5] A. Bastian *et al.*, “No Title,” no. 1, pp. 26–32.
- [6] B. M. Metisen and H. L. Sari, “ANALISIS CLUSTERING MENGGUNAKAN METODE K-MEANS DALAM PENGELOMPOKKAN PENJUALAN PRODUK PADA SWALAYAN FADHILA,” vol. 11, no. 2, pp. 110–118, 2015.
- [7] D. Nielza Atthina, “Klasterisasi Data Kesehatan Penduduk untuk Menentukan Rentang Derajat Kesehatan Daerah dengan Metode K-Means,” *Aseminar Nas. Apl. Teknol. Infromasi*, vol. 1, no. Klustering, p. B-52-B-59, 2014.
- [8] M. N. Mara and N. Satyahadewi, “PENGKLASIFIKASIAN KARAKTERISTIK DENGAN METODE K-MEANS CLUSTER ANALYSIS,” vol. 02, no. 2, pp. 133–136, 2013.
- [9] A. P. Windarto, “Penerapan Datamining Pada Ekspor Buah-Buahan Menurut Negara Tujuan Menggunakan K-Means Clustering Method,” *Techno.Com*, vol. 16, no. 4, pp. 348–357, 2017.
- [10] J. O. Ong, “Implementasi Algotritma K-means clustering untuk menentukan strategi marketing president university,” *J. Ilm. Tek. Ind.*, vol. vol.12, no. no. juni, pp. 10–20, 2013.
- [11] F. L. Sibuea, A. Sapta, S. Informasi, and S. Royal, “PEMETAAN SISWA BERPRESTASI MENGGUNAKAN METODE K-MEANS CLUSTERING,” vol. IV, no. 1, 2017.

- [12] M. H. Adiya and Y. Desnelita, “Jurnal Nasional Teknologi dan Sistem Informasi Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru,” vol. 01, pp. 17–24, 2019.
- [13] L. Maulida, “Penerapan Datamining Dalam Mengelompokkan Kunjungan Wisatawan Ke Objek Wisata Unggulan Di Prov. Dki Jakarta Dengan K-Means,” *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 2, no. 3, pp. 167–174, 2018.
- [14] S. Kaputama, “Clustering Pelanggaran Berkendaraan Menggunakan,” pp. 1–9.
- [15] F. Gullo, “From Patterns in Data to Knowledge Discovery : What Data Mining Can Do,” *Phys. Procedia*, vol. 62, pp. 18–22, 2015.