SYSTEMIC: Information System and Informatics Journal

ISSN: 2460-8092, 2548-6551 (e)

Vol 7 No 1 - Agustus 2021

Implementasi Metode Clusterisasi K-Means Pada Pemetaan Daerah Rawan Kriminal Kota Dili Berbasis WebGIS

Elisabet Maria Pereira da Costa Leite¹, Latipah²^{1,2}) Universitas Narotama, Surabayaelisabetscosta117@gmail.com¹, latifah.rifani@narotama.ac.id²

Kata Kunci

Kriminalitas, Data-mining, clustering k-means, GIS

Abstrak

Penelitian ini bertujuan untuk memetakan sub-district yang rawan akan terjadinya tindakan kriminalitas dengan data mining untuk melakukan pengelompokan data kriminal menggunakan clustering k-means, serta melakukan visualisasi kedalam sistem informasi kriminal berbasis webgis dengan menampilkan map persebaran data kriminal dan prosentase dari masing-masing sub-district dari kota Dili. Setelah melakukan serangkaian proses maka didapatkan hasil bahwa sub-district dom aleixo dimana menunjukan daerah berstatus sangat rawan dengan jenis kriminalitas penganiayaan memiliki presentase total korban tinggi dengan nilai sebesar 49.6%, pencurian 19.6%, pelecehan 12%, prostitusi 9.6%, pembunuhan 6.8% dan terakhir narkoba dengan presentase 2.4%.

Keywords

Crime, Data-mining, clustering k-means, GIS

Abstract

This study aims to map sub-districts that are prone to criminal acts with data mining to group criminal data using k-means clustering, as well as visualize into a webgis-based criminal information system by displaying a map of the distribution of criminal data and the percentage of each sub-district. -district of the city of Dili. After carrying out a series of processes, it was found that the sub-district of dom aleixo which shows a very vulnerable status area with the type of crime of assault has a high percentage of total victims with a value of 49.6%, theft 19.6%, harassment 12%, prostitution 9.6%, murder 6.8% and finally drugs with a percentage of 2.4%.

1. Pendahuluan

Timor-Leste merupakan salah satu negara yang secara geografis terletak pada Asia Tenggara, setelah kemerdekaan Timor-Leste pada tahun 2002 yang sempat dijajah selama 24 tahun oleh Indonesia dapat dikatakan bahwa Timor-Leste sebagai Negara yang masih muda dan sedang ada pada proses perkembangan dalam berbagai bidang terutama pada bidang teknologi [1]. Pada saat ini penggunaan teknologi informasi dalam dunia pekerjaan di Timor-Leste belum dimaksimalkan sehingga hal ini bisa menjadi salah satu faktor kurangnya efektifitas dalam dunia pekerjaan, salah satu contoh adalah bidang keamanan kepolisian dimana sebagian besar pendataan yang dilakukan masih secara manual atau data masih dalam bentuk mentah, tentunya hal ini dapat mempengaruhi untuk melakukan presentase dari berbagai unit kepolisian terutama pada unit kriminal untuk melakukan presentase mengenai tindakan

kriminalitas yang terjadi pada kota Dili.

Kriminalitas atau tindakan kriminal merupakan peristiwa yang sering ditemui atau terjadi dimana saja terutama pada kota-kota besar seperti di kota Dili. Kriminalitas yang terjadi pada kota Dili sangat beragam, dan juga ditambah dengan pendataan yang masih dilakukan secara manual, tentunya hal ini membuat pihak kepolisian sulit untuk menentukan lokasi-lokasi yang rawan akan terjadinya tindakan kriminalitas. Di era globalisasi sekarang ini, teknologi semakin berkembang bahkan perkembangan teknologi saat ini banyak membantu memudahkan pekerjaan manusia supaya lebih efisien dan efektif dalam berbagai bidang. [2]

Berdasarkan uraian diatas, maka dari itu peneliti berinisiatif untuk membuat sebuah sistem informasi pemetaan daerah rawan kriminal berbasis webgis dengan memanfaatkan data yang diperoleh dari pihak kepolisian Dili untuk

Tabel 1. Dataset kriminal

JK	SD	TK	lat	lon
Pembunuhan	D. Aleixo	5	-8,5526928	125,5525064
Pelecehan	Cristo rei	2	-8,5758262	125,5667781
Pengancaman	D. Aleixo	7	-8,5526928	125,5525064
Pengancaman	Vera Cruz	3	-8,5630659	125,5746235
Pencurian	D. Aleixo	2	-8,5526928	125,5525064
Prostitusi	Nain fetu	1	-8,5553641	125,5778065
Pembunuhan	Cristo rei	1	-8,5758262	125,5667781
Pencurian	Vera cruz	2	-8,5630659	125,5746235
Pencurian	Cristo rei	2	-8,5758262	125,5667781
Pengancaman	D.Aleixo	2	-8,5526928	125,5525064
Narkoba	Nain fetu	5	-8,5553641	125,5778065
...	...	...	...	...

melakukan proses pengolahan data dengan menggunakan metode *clusterisasi k-means* kemudian implementasi ke dalam sistem informasi pemetaan daerah rawan kriminal berupa visualisasi grafik presentase total korban. K-means merupakan salah satu algoritma dari analisis cluster non hirarki yang bertujuan untuk mempartisi data kedalam beberapa kelompok cluster berdasarkan karakteristiknya. Keakuratan metode K-Means tergantung pada pemilihan datacentroid selama inisialisasi[3]

Tinjauan penelitian pertama disusun oleh Gilang Yudistira Hilman, Bandi Sasmito, dan Arwan Putra Wijaya yang berjudul "Pemetaan Daerah Rawan Kriminalitas Di Wilayah Hukum Poltabes Semarang Pada Tahun 2013 Dengan Menggunakan Metode Clustering". Hasil yang didapatkan dari penelitian ini sendiri adalah tingkat kerawanan di wilayah semarang berbeda-beda dan sering terjadi antara waktu a pukul 00.01 - 06.00 kemudian 06.01-12.00, 12.01-18.00 dan 06.00-12.00 di mana 50,03 % tindak kejahatan terjadi antara pukul 00.01-06.00. [4]

Diharapkan melalui data kriminal yang didapat dari kepolisian dapat melakukan proses clustering dengan metode k-means serta melakukan implementasi kedalam sistem informasi geografis berupa visualisasi kerawanan tiap sub-district berdasarkan total korban dari semua jenis kejahatan. Sehingga dapat mengetahui status kerawanan dari masing-masing sub-district agar lebih efisien, untuk menempatkan anggota kepolisian pada sub-district yang rawan.

Pada tahap ini akan membahas tentang proses serta metode yang digunakan dalam penelitian ini. Berikut dapat dilihat pada gambar 1 menunjukan flowchart desain penelitian.

2.1 Observasi

Observasi yang dilakukan pada penelitian ini yaitu melakukan observasi secara langsung di kepolisian pusat Dili tepatnya pada unit kriminal mengenai sistem dan juga kebutuhan data yang akan digunakan oleh peneliti.

2.2 Data Collection

Proses pengumpulan data berfungsi untuk memperoleh informasi dari perusahaan atau instansi yang bersangkutan sehingga bisa menjadikan sebagai bahan acuan penelitian untuk melakukan klasterisasi *k-means*. Pengumpulan data berarti mengumpulkan data yang dibutuhkan untuk analisis[5]. Pada penelitian ini data yang digunakan yaitu data kriminal dari tahun 2017-2019 terdiri 5 atributantara lain adalah sub-district, jenis kejahatan, total korban, latitude danjuga longitude dari masing-masing sub-district. Berikut akan ditampilkan pada tabel 1 dataset kriminal.

2.3 Preprocessing Data

Tahap berikutnya adalah *preprocessing* data, tujuan dengan melakukan *preprocessing* ini sendiri yaitu untuk melakukan pembersihan atau pengurangan atribut data yang tidak diperlukan serta kolom variabel yang tidak memiliki nilai, selain itu *preprocessing* data juga meningkatkan kualitas data dan mengurangi ukuran file log web[6].

Pada penelitian ini meliputi dua proses preprocessing antara lain yaitu data cleaning, dan data transformation.

1. Data cleaning

Tujuan utama dari data cleaning sendiri adalah untuk membersihkan data dengan beberapa teknik seperti mengisi data entry yang kosong, memperbaiki data yang tidak konsisten.

2. Data Transformation

Dalam data *transformation* ini akan dilakukan

Gambar 1. Flowchart Desain Penelitian

proses normalisasi pada dataset yang ada sehingga nilai yang didapatkan berada dalam *scope* nilai yang sama. Pada penelitian ini metode normalisasi yang digunakan adalah *minmax normalization* dengan persamaan(1).

$$Z = \frac{(x - x_{Min})}{(x_{Max} - x_{Min})} \quad 0.0 \text{ Sampai } 1.0 \quad (1)$$

Gambar 2. Flowchart clustering k-means

Keterangan :

Z = hasil data normalisasi
 x = nilai awal
 x_{min} = nilai minimal data
 x_{max} = nilaimaskimal data

2.4 Clustering K-means

Clustering K-means merupakan sebuah metode atau proses yang bertujuan untuk mengelompokkan data dengan tingkat kemiripan sama kedalam satu atau lebih kelompok. [7] Pada proses *clustering k-means* menggunakan tools dari python yaitu jupyterlab. Berikut gambar 2 adalah alur diagram dari proses *clustering k-means*.

Alur proses *clustering k-means* dapat diejelaskan sebagai berikut :

1. Menentukan total cluster yang akan digunakan dengan elbow method. Metode dan analisis ini digunakan untuk pemilihan jumlah cluster atau kelompok yang optimal.
2. Menentukan centroid awal dari setiap kelompok cluster secara random.
3. Menghitung jarak antara dataset dengan pusat cluster menggunakan rumus euclidean distance dengan persamaan (3), kemudian mengelompokkan data berdasarkan jarak terdekat dari pusat cluster.

$$Z = \sqrt{\sum (x_i - y_i)^2} \quad ni=1 \quad (2)$$

Keterangan :

z = jarak

i = setiap data
 c = data pusat centroid
 x = data atribut
 n = jumlah data

2.5 Visualisasi SIG

Visualisasi SIG pada penelitian ini adalah merepresentasikan dataset kriminal kedalam webgis berupa lokasi dari kota Dili dan juga grafik total korban dari jenis tindakan kriminal yang ada.

2. Hasil Dan Pembahasan

Pada tahapan ini, akan memaparkan hasil dari pengolahan data dengan mengimplementasi metode clustering k-means menggunakan tools dari python yaitu anaconda navigator (*jupyterlab*).

3.1 Pengolahan Data

Pada proses pengolahan data akan dilakukan analisis terhadap dataset kriminal tahun 2017-2019. Tipe dataset yang digunakan pada penelitian ini bertipe *unsupervised learning*(data tanpa label). Sebelum memulai proses pengelompokan data daerah berdasarkan status rawan atau tidaknya, hal pertama yang harus dilakukan adalah dengan mengubah atau membuat inisialisasi data awal yang bertipe kategorik menjadi data numerical seperti mana ditunjukkan pada tabel 2. Berikut adalah keterangan data inisialisasi :

1. Jenis kejahatan : penganiyaan = 1, pelecehan = 2, prostitusi = 3, narkoba = 4, pembunuhan = 5, pencurian = 6.
2. Sub-district : Dom aleixo = 1, cristorei = 2, nainfeto = 3, 21eracruz = 4, atauro = 5, metinaru = 6.

Tabel 2.Hasilinisialisasi data kategorik

JK	SD	TK	lat	lon
5	1	5	-8,5526928	125,5525064
2	2	2	-8,5758262	125,5667781
1	1	7	-8,5526928	125,5525064
1	4	3	-8,5526928	125,5525064
6	2	2	-8,5526928	125,5525064
3	3	1	-8,5553641	125,5778065
5	2	1	-8,5758262	125,5667781
6	4	2	-8,5526928	125,5525064
6	2	2	-8,5758262	125,5667781
1	1	2	-8,5526928	125,5525064
4	3	5	-8,5553641	125,5778065
...	...	...	...	...

3.2 Data Normalization (Min-max)

Pada proses normalisasi data menggunakan min-max berfungsi untuk mengubah nilai awal data dengan tipe *decimal* menjadi *float*. *Min-Max Normalization* merupakan metode normalisasi dengan melakukan transformasi linear terhadap data asli sehingga dapat menghasilkan keseimbangan nilai perbandingan antar data saat sebelum dan sesudah diproses [8]. Berikut dapat dilihat pada tabel 3 hasil dari normalisasi.

Tabel 3. Hasil normalisasi

JK	SD	TK
0.700140	0.140028	0.700140
0.577350	0.577350	0.577350
0.140028	0.140028	0.980196
0.196116	0.784465	0.588348
0.937043	0.156174	0.312348
0.688247	0.688247	0.229416
0.912871	0.365148	0.182574
0.801784	0.534522	0.267261
0.904534	0.301511	0.301511
0.408248	0.408248	0.816497
0.742781	0.557086	0.371391
...	...	...

3.3 Proses Metode Clustering K-means

Pada pemrosesan data dengan metode *clustering k-means* seperti yang telah dipaparkan pada sub bab sebelumnya tentang alur diagram dari metode ini. Terlebih dahulu yaitu menentukan jumlah cluster yang akan dibentuk, kemudian menentukan nilai awal centroid dari masing-masing *cluster*, dalam penelitian ini total dataset yang digunakan dari tahun 2017-2019 adalah 178 lalu dibagi menjadi empat (4) kelompok atau cluster. Setelah melakukan penentuan centroid proses selanjutnya yaitu menghitung jarak dari setiap data terhadap pusat *cluster* lalu mengelompokkan data berdasarkan jarak terkecil dari pusat cluster.

Berikut merupakan contoh perhitungan dari pengelompokan data terdapat 4 cluster dengan masing-masing nilai adalah sebagai berikut :

$$\begin{aligned}
 C1 &= (0.140028, 0.140028, 0.980196) \\
 C2 &= (0.688247, 0.688247, 0.229416) \\
 C3 &= (0.408248, 0.408248, 0.816497) \\
 C4 &= (0.742781, 0.557086, 0.371391)
 \end{aligned}$$

Selanjutnya menghitung jarak dari semua data dengan pusat cluster, sebagai contoh dalam eksperimen ini akan dilakukan perhitungan data ke-1 dengan pusat cluster.

$$\begin{aligned}
 d^1_{C1} &= \sqrt{(0.700140 - 0.140028)^2 + (0.140028 - 0.140028)^2} \\
 &= (0.700140 - 0.980196)^2 = 1.0980390825 \\
 d^1_{C2} &= \sqrt{(0.700140 - 0.688247)^2 + (0.140028 - 0.688247)^2} \\
 &= (0.700140 - 0.229416)^2 = 0.078821544 \\
 d^1_{C3} &= \sqrt{(0.700140 - 0.408248)^2 + (0.140028 - 0.408248)^2} \\
 &= (0.700140 - 0.816497)^2 = 0.0267979227 \\
 d^1_{C4} &= \sqrt{(0.700140 - 0.742781)^2 + (0.140028 - 0.557086)^2} \\
 &= (0.700140 - 0.371391)^2 = 0.067677921
 \end{aligned}$$

Berdasarkan perhitungan yang dilakukan pada data ke-1 dengan setiap cluster menunjukkan jarak minimum atau terkecil adalah cluster 4 dengan nilai jarak 0.067677921. Hasil yang didapatkan dari proses pembentukan cluster k-means dapat direpresentasikan kedalam bentuk array dengan status masing-masing yaitu 0 menunjukkan status wilayah aman, 1= cukup aman, 2= rawan, 3= sangat rawan.

Berikutnya menambahkan hasil perhitungan jarak pada kolom baru dalam dataset kriminal seperti yang ditampilkan pada tabel 4.

Tabel 4. Pembentukan cluster

JK	SD	TK	Clstr
0.700140	0.140028	0.700140	3 C4
0.577350	0.577350	0.577350	3 C4
0.140028	0.140028	0.980196	0 C1
0.196116	0.784465	0.588348	0 C1
0.937043	0.156174	0.312348	2 C3
0.688247	0.688247	0.229416	0 C1
0.912871	0.365148	0.182574	0 C1
0.801784	0.534522	0.267261	1 C2
0.904534	0.301511	0.301511	1 C2
0.408248	0.408248	0.816497	1 C2
0.742781	0.557086	0.371391	2 C3
...	...	...	...

3.4 Sistem Informasi Geografis

Pada tahapan ini menjelaskan bagaimana program implementasikan kedalam sistem informasi geografis dengan menampilkan visual map dari kota Dili dengan memanfaatkan longitude dan latitude yang ada. Terdapat 6 *sub-district* yang digunakan pada penelitian ini terdiri dari *sub-district* domaleixo, cristorei, nainfeto, veracruz, atauru, dan metinaru. Berikut pada gambar 4 menampilkan dashboard webgis kriminal kepolisian Dili.

Tampilan gambar 3 menunjukkan dashboard webgis kriminal yang menampilkan beberapa informasi dari masing-masing struktur website, seperti pada struktur header terdapat pada 2 menu yaitu menu login dan maps, namun kali ini akan fokus pada menu maps sebagai inti dari pembahasan.

Gambar 3. Dashboard kriminal

Berikut gambar 4 menunjukkan hasil persebaran data kriminal yang di visualisasikan kedalam GIS kemudian ditampilkan melalui menu maps.

Gambar 4. Persebaran data kriminal

Gambar 5 menampilkan map persebaran data kriminal setiap *sub-district* yang ada pada kota Dili, dimana tiap *sub-district* yang ada akan menampilkan informasi berupa presentase jenis kriminalitas, pada pembahasan ini akan diambil *sub-district* dom aleixo dan metinaru sebagai salah contoh untuk menampilkan informasi presentase tindakan kriminal yang terjadi di kota Dili, Timor-Leste seperti yang ditunjukkan pada gambar 5 dan 6.

Gambar 5. Presentase kriminal sub-district dom aleixo

Gambar 5 menampilkan presentase total korban dari *sub-district* dom aleixo berdasarkan jenis kriminalitas, informasi yang ditampilkan pada webgis kriminal yakni terdapat 6 jenis kriminalitas dengan nilai persentasenya masing-masing terdiri dari jenis kejahatan penganiayaan dengan nilai presentase total korban sebesar 49.6%, pencurian dengan nilai presentase 19.6%, pelecehan 12%, prostitusi 9.6%, pembunuhan 6.8% dan terakhir narkoba dengan presentase 2.4%.

Gambar 6. Presentase kriminal sub-district metinaru

Gambar 6 menampilkan presentase total korban *sub-district* metinaru berbeda dengan presentase sebelumnya yakni menampilkan kasus kriminalitas dengan tipe kasus yang terjadi cukup sedikit dimana hanya terdapat 3 jenis kejahatan terdiri dari kasus penganiayaan dengan nilai presentase korban sebanyak 52.4%, pencurian 38.1%, dan prostitusi dengan nilai 9.5%.

3. Kesimpulan

Berdasarkan pembahasan yang dilakukan, dengan data kriminal tahun 2017-2019 berjumlah 178 dataset yang didapatkan dari kepolisian Dili. Dengan menggunakan metode klusterisasi k-means hasil keakuratan cluster yang diperoleh cukup baik dengan nilai sebesar 0.514419. Dari semua penjelasan pada sub bab sebelumnya mengenai analisis data kriminal dan visualisasi kedalam sistem informasi geografis berdasarkan total korban dari semua tindakan kriminal maka dapat diambil kesimpulan bahwa presentase total korban pada *sub-district* dom aleixo maupun metinaru dengan jenis kejahatan yang memiliki presentase tinggi pada *sub-district* dom aleixo adalah jenis kejahatan penganiayaan. Sedangkan pada *sub-district* metinaru kasus kriminalitas yang terjadi terbilang sedikit atau berstatus aman dimana hanya terdapat 3 jenis kejahatan terdiri dari jenis kriminalitas penganiayaan yang memiliki nilai presentase tinggi, diikuti dengan pencurian, dan prostitusi.

DAFTAR PUSTAKA

- [1] R. D. Oliveira, "Por uma Geografia de Timor Leste, timorenses...", no. September, 2017, doi: 10.13140/RG.2.2.29979.57125.
- [2] R. Andrean, S. Fendy, and A. Nugroho, "Klusterisasi Pengendalian Persediaan Aki Menggunakan Metode K-Means," *JOINTECS (Journal Inf. Technol. Comput. Sci.)*, 2019, doi: 10.31328/jointecs.v4i1.998.
- [3] R. M. Esteves, T. Hacker, and C. Rong, "Competitive K-means: A new accurate and distributed K-means algorithm for large datasets," *Proc. Int. Conf. Cloud Comput. Technol. Sci. CloudCom*, vol. 1, pp. 17-24, 2013, doi: 10.1109/CloudCom.2013.89.
- [4] G. Hilman, B. Sasmito, and A. Wijaya, "Pemetaan Daerah Rawan Kriminalitas Di Wilayah Hukum Poltabes Semarang Tahun 2013 Dengan Menggunakan Metode Clustering," *J. Geod. Undip*, vol. 4, no. 1, pp. 32-42, 2015.
- [5] S. Sharma and A. Bhagat, "Data preprocessing algorithm for Web Structure Mining," *Proc. 5th Int. Conf. Eco-Friendly Comput. Commun. Syst. ICECCS 2016*, pp. 94-98, 2017, doi: 10.1109/Eco-friendly.2016.7893249.
- [6] S. K. Dwivedi and B. Rawat, "A review paper on data preprocessing: A critical phase in

- web usage mining process," *Proc. 2015 Int. Conf. Green Comput. Internet Things, ICGCIoT 2015*, pp. 506–510, 2016, doi: 10.1109/ICGCIoT.2015.7380517.
- [7] M. G. H. Omran, A. P. Engelbrecht, and A. Salman, "An overview of clustering methods," *Intell. Data Anal.*, vol. 11, no. 6, pp. 583–605, 2007, doi: 10.3233/ida-2007-11602.
- [8] D. A. Nasution, H. H. Khotimah, and N. Chamidah, "Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma K-NN," *Comput. Eng. Sci. Syst. J.*, vol. 4, no. 1, p. 78, 2019, doi: 10.24114/cess.v4i1.11458.