

Penerapan Algoritma K-Means Clustering untuk Mengelompokkan Kerawanan Rob di Daerah Pekalongan

Imam Faturohim¹, Anna Baita²

^{1,2} Fakultas Ilmu Komputer,
Universitas Amikom
Yogyakarta, Indonesia

¹imam.faturohim@students.amikom.ac.id

²anna@amikom.ac.id (Corresponding
author)

Ringkasan

Banjir rob di kawasan Kota Pekalongan telah terjadi sejak lama dan berpotensi semakin luas. Hal ini disebabkan oleh beberapa wilayah di Kota Pekalongan memiliki ketinggian tanah dibawah permukaan air laut dan mengalami penurunan muka tanah (land subsidence). Ancaman banjir dan rob semakin tinggi karena adanya pemanasan global yang berdampak pada naiknya permukaan air laut. Pemetaan kerawanan Rob ini penting agar dapat membantu pemerintah daerah dalam evaluasi dan penanganan rob. Selain itu, pemetaan ini dapat digunakan untuk mengevaluasi dan merencanakan tata ruang wilayah Kota Pekalongan. Dalam penelitian ini menggunakan metode K-Means Clustering untuk mengelompokkan data berdasarkan tinggi genangan rob dan luas area genangan rob. Hasil penelitian ini menghasilkan 3 cluster daerah. Hasil penelitian ini menunjukkan bahwa daerah dengan kerawanan tinggi bencana rob di kota Pekalongan ada 7 Kelurahan yakni Bandengan, Degayu, Kandang Panjang, Krapyak, Padukuhan Kraton, Panjang Wetan dan Panjang Baru. Hasil Clustering disajikan dalam bentuk peta menggunakan ArcGis.

Kata kunci: Rob, K-Means, Clustering, Pekalongan, ArcGis

1. Pendahuluan

Kota Pekalongan yang terletak di dataran rendah pantai utara Pulau Jawa mengalami banjir rob dengan frekuensi dan luas genangan yang semakin meningkat[1]. Banjir rob di kawasan Kota Pekalongan telah terjadi sejak lama dan berpotensi semakin luas. Hal ini disebabkan oleh beberapa wilayah di Kota Pekalongan memiliki ketinggian tanah dibawah permukaan air laut dan mengalami penurunan muka tanah (land subsidence). Ancaman banjir dan rob semakin tinggi karena adanya pemanasan global yang berdampak pada naiknya permukaan air laut[2].

Badan Penanggulangan Bencana Daerah (BPBD) Kota Pekalongan untuk tahun 2021 belum memiliki peta wilayah daerah rawan rob. Data yang tersedia di BPBD merupakan data statistik per kejadian. Penanggulangan bencana rob, akan lebih tepat sasaran jika sudah ada pengelompokan tingkat kerawanan rob. Sehingga memudahkan untuk penentuan strategi jangka panjang maupun menengah dalam penanggulangan banjir rob. Peta wilayah daerah rawan rob juga dapat digunakan untuk melakukan evaluasi terhadap rencana Tata Ruang wilayah Kota Pekalongan.

Pada penelitian Iskandar telah membahas pemodelan geospasial genangan banjir rob kota pekalongan tahun 2020 dan memprediksi genangan banjir rob 2025 berdasarkan data pasang surut, tipe pasang surut, laju kenaikan permukaan air laut dan laju penurunan muka tanah[3]. Pada penelitian usulan akan melakukan pengelompokan kerawanan rob di kota pekalongan dengan data 2021 dan melakukan pemetaan hasilnya.

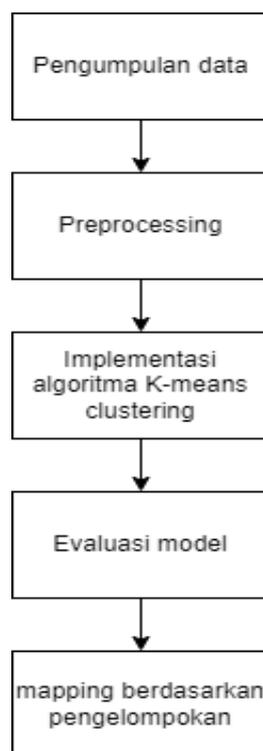
Metode yang dapat digunakan untuk pengelompokan data rawan bencana rob adalah metode clustering. Ada beberapa algoritma untuk melakukan clustering data, antara lain Hierarchical clustering, K-means clustering, k-medoids clustering[4]. K-means clustering, algoritma ini mengelompokkan data yang ada kedalam

beberapa kelompok, dimana data dalam satu kelompok mempunyai karakteristik yang sama satu sama lainnya dan mempunyai karakteristik yang berbeda dengan data yang ada pada kelompok lainnya. Algoritma ini memiliki proses komputasi yang cepat[5].

Dengan adanya pemaparan di atas, maka dalam penelitian ini akan membahas penerapan algoritma k-means clustering untuk mengelompokkan kerawanan rob di daerah Pekalongan

2. Metode Penelitian

Metode penelitian yang digunakan dalam penelitian ini digambarkan oleh bagan berikut:



Gambar 1. Alur Penelitian

Berdasarkan Gambar 1, tersebut di atas, ada beberapa tahap yang dijalankan dalam penelitian ini:

1. Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang didapatkan dari BPBD Kota Pekalongan dan Bappeda Kota Pekalongan. Data yang diperoleh adalah data kejadian rob dari tahun 2021 yang berupa data shape file. Data tersebut terdiri dari: data kelurahan, luas kelurahan, tinggi genangan air, kecamatan, luas area rob, dan tinggi tanah. Data shape file merupakan format data geospasial yang berbentuk vektor dan direpresentasikan dalam bentuk garis (lines), titik (points) dan area(polygon)[6].

2. *Preprocessing*

Ada beberapa tahap yang dilakukan pada tahap *preprocessing* ini, antara lain: penanganan data kosong, *encoder*, normalisasi data, dan menghitung nilai korelasi antar parameter.

Penanganan data kosong ini bertujuan untuk membersihkan data dari *field* yang bernilai kosong. Untuk dapat diproses menggunakan algoritma *k-means* maka semua data harus berupa data *numeric*. Sehingga data-data *non numeric* dikonversi terlebih dahulu menjadi data *numeric*. Langkah berikutnya adalah normalisasi data. Proses ini bertujuan agar menyamakan rentang data sehingga tidak terjadi kesenjangan antar parameter. Tahap terakhir dari *preprocessing* adalah menghitung nilai korelasi antar parameter. Proses ini dilakukan untuk mengetahui menyeleksi fitur mana yang akan dijadikan parameter dalam proses *clustering*.

3. Implementasi Algoritma *K-Means Clustering*

Clustering adalah jenis metode statistik yang dipakai untuk mengelompokkan banyak data atau objek ke dalam beberapa kelompok (cluster), sehingga objek-objek yang berada dalam satu kelompok akan mempunyai

kemiripan yang tinggi [7].

Langkah-langkah algoritma K-Means adalah sebagai berikut[8]:

a. Menentukan jumlah *cluster* (*k*) dalam dataset

b. Menentukan nilai pusat (*centroid*)

Penentuan *centroid* pada awal proses dilakukan secara acak. Untuk menghitung jarak setiap data *input* terhadap masing-masing *centroid* dengan menggunakan rumus jarak Euclidean Distance yaitu:

$$d(x_i, \mu_j) = \sqrt{\sum (x_i, \mu_j)^2} \quad (1)$$

Keterangan:

x_i = data kriteria

μ_j = centroid pada cluster ke-*js*

c. Mengelompokkan setiap data berdasarkan kedekatannya dengan centroid atau mencari jarak terkecil.

d. Melakukan update nilai centroid. Nilai centroid baru dihitung dari rata-rata cluster yang bersangkutan dengan menggunakan rumus yaitu:

$$\mu_j(t+1) = \frac{1}{N_{sj}} \sum_{j \in s_j} x_j \quad (2)$$

Keterangan:

$\mu_j(t+1)$ = centroid baru pada iterasi (*t+1*)

N_{sj} = Data pada cluster *Sj*

e. Apabila data setiap cluster belum berhenti, lakukan perulangan dari langkah b hingga e, sampai anggota tiap cluster tidak ada yang berubah

4. Evaluasi Model

Evaluasi model, dilakukan untuk mengetahui apakah jumlah cluster yang terbaik. Adapun metode yang digunakan adalah metode elbow [9] dan metode silhouette [10][11]

5. Mapping berdasarkan hasil *cluster*

Tahap selanjutnya adalah memindahkan data hasil clustering kedalam software ArcGis untuk dilakukan Mapping data. ArcGIS biasa digunakan dalam proses pembuatan peta, pemodelan, pengelolaan data spasial secara efektif dan efisien serta membantu SIG dalam melakukan analisis[12]. Proses mapping dilakukan dengan menggabungkan hasil dari clustering ke dalam attribute tabel yang sebelumnya sudah dilakukan proses overlay untuk pembagian cluster di setiap kelurahan yang ada pada Kota Pekalongan.

3. Hasil dan Pembahasan

Implementasi dari setiap tahapan adalah sebagai berikut:

3.1. Pengumpulan Data

Data yang diperoleh dari BPBD merupakan data shape file. Data tersebut berisi tentang gambaran wilayah Kota Pekalongan dan data bencana yang ada di Kota Pekalongan, data tersebut merupakan rekap data pada tahun 2021. data tersebut diubah ke dalam format .csv agar dapat diakses melalui python. Data tersebut ditunjukkan oleh Gambar 2 berikut ini.

Dari Gambar ?? tersebut diketahui ada beberapa atribut yang akan digunakan antara lain: kelurahan, luas kelurahan, tinggi genangan, luas area genangan, dan tinggi tanah. Data tersebut akan diproses menggunakan Python.

```
import numpy as np
import pandas as pd
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn import preprocessing

url="https://drive.google.com/file/d/1iYRhXUxNAazJ88uyrdp8pKxfG7bv710U/view?usp=sharing"
file_id=url.split('/')[-2]
link="https://drive.google.com/uc?id="+file_id
df=pd.read_csv(link)
df
```

	KELURAHAN	LuasKel	Tinggi	Kec	LuasArea	Tinggi tanah
0	Kelurahan Bandengan	221.56	70	Pekalongan Utara	123.90	1.44
1	Kelurahan Bandengan	221.56	50	Pekalongan Utara	95.00	1.44
2	Kelurahan Banyurip	173.78	-10	Pekalongan Barat	0.42	4.87

0s completed at 8:33 PM

Gambar 2. Data Rob

```
df.isna().sum()
```

KELURAHAN	0
LuasKel	0
Tinggi	0
Kec	0
LuasArea	0
Tinggi tanah	0
dtype:	int64

Gambar 3. Pengecekan Data Kosong

3.2. Preprocessing

Langkah pertama dari preprocessing adalah penanganan data kosong, yang ditunjukkan oleh Gambar 3 berikut ini.

Dari Gambar 3 tersebut diketahui bahwa tidak ada data yang kosong, sehingga tidak perlu adanya penghapusan data.

Langkah selanjutnya adalah proses encoder, yakni mengubah data *non numeric* menjadi *numeric*[13]. Untuk dapat melihat tipe data dari setiap attribute dapat dilihat dalam Gambar 4 berikut ini.

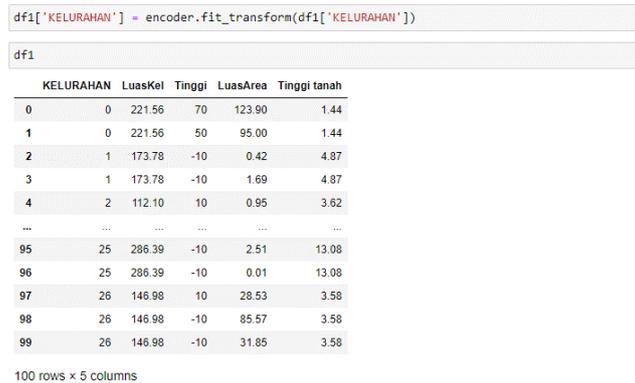
```
df.dtypes
```

KELURAHAN	object
LuasKel	float64
Tinggi	int64
LuasArea	float64
Tinggi tanah	float64
dtype:	object

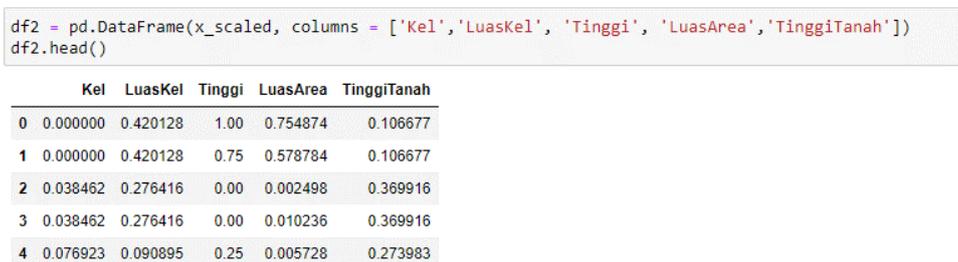
Gambar 4. Tipe Data

Dari Gambar 4 tersebut terlihat bahwa hanya data kelurahan yang belum bertipe *numeric*. Sehingga data kelurahan perlu diencoder terlebih dahulu. Hasil encoder diperlihatkan oleh Gambar 5.

Dari Gambar 5 terlihat bahwa data kelurahan sudah berubah menjadi data *numeric*. Selanjutnya adalah proses normalisasi data. Normalisasi adalah proses penskalaan nilai atribut dari data sehingga bisa terletak pada rentang tertentu[14]. Proses normalisasi yang digunakan dalam penelitian ini menggunakan algoritma min-max scaling. Min-Max scaling merupakan metode normalisasi yang sering digunakan untuk mengatasi permasalahan nilai antar fitur yang memiliki jarak terlampaui jauh[15]. Hasil normalisasi ditunjukkan oleh Gambar 6.

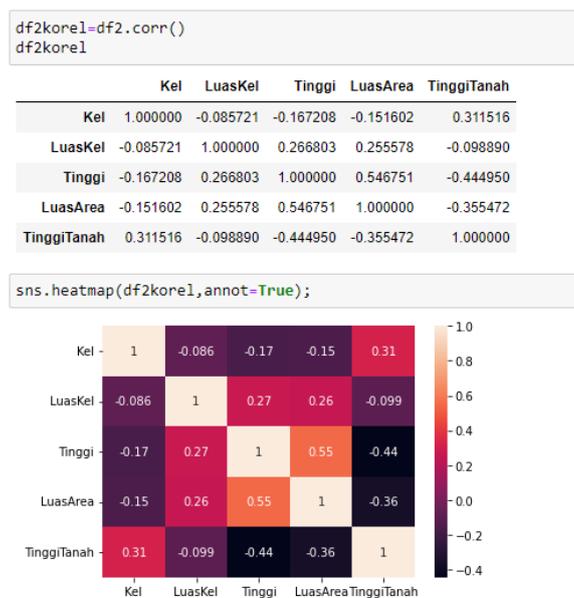


Gambar 5. Proses Encoder



Gambar 6. Hasil Normalisasi

Dari Gambar 6 terlihat bahwa rentang data keseluruhan atribut sudah sama, yakni antara 0 dan 1. setelah proses normalisasi maka langkah selanjutnya adalah proses pencarian korelasi antar atribut. Pencarian korelasi antribut ditunjukkan oleh Gambar 7 berikut ini.



Gambar 7. Korelasi Antar Atribut

Dari Gambar 7, diketahui bahwa nilai korelasi tertinggi antar atribut adalah korelasi antara atribut tinggi ge-nangan air dan luar area, yakni sebesar 0.55. maka parameter yang dijadikan inputan untuk proses clustering adalah atribut dengan nilai korelasi tertinggi.

3.3. Implementasi Algoritma K-Means Clustering

Proses clustering dilakukan dengan menggunakan parameter tinggi genangan air dan luas area rob. Dalam penelitian ini, data dikelompokkan menjadi 3 cluster. Proses dan hasil clustering ditunjukkan oleh Gambar 8 berikut ini.



Gambar 8. Proses Clustering

Dari Gambar 8 diatas dapat dibaca bahwa untuk cluster 0 merupakan data yang mempunyai nilai tinggi genangan yang tinggi dan mempunyai nilai luas area yang tinggi dan sedang. Cluster 0 diwakili oleh sebaran data yang berwarna merah. Cluster 1 diwakili oleh warna kuning merupakan data yang mempunyai nilai tinggi genangan yang sedang dan nilai luas area yang rendah dan sedang. Sebaran data yang berwarna hijau merupakan data cluster 2. Data cluster 2 merepresentasikan data dengan nilai tinggi genangan yang rendah dan nilai luas area yang rendah dan sedang. Nilai Centroid dari tiap cluster diwakili oleh titik yang berwarna hitam.

Sehingga dapat disimpulkan bahwa cluster 0 merupakan daerah yang memiliki tingkat kerawanan rob paling tinggi. Cluster 1 merupakan daerah yang memiliki tingkat kerawanan rob sedang. Cluster 2 merupakan daerah yang memiliki tingkat kerawanan rob rendah. Setelah data cluster didapatkan, maka tahap selanjutnya adalah meletakkan data hasil cluster ke dalam tabel untuk digabungkan dengan attribute yang lain. Proses tersebut ditunjukkan oleh Gambar 9 berikut.

	KELURAHAN	LuasKel	Tinggi	Kec	LuasArea	Tinggi tanah	cluster
0	Kelurahan Bandengan	221.56	70	Pekalongan Utara	123.90	1.44	0
1	Kelurahan Bandengan	221.56	50	Pekalongan Utara	95.00	1.44	0
2	Kelurahan Banyurip	173.78	-10	Pekalongan Barat	0.42	4.87	2
3	Kelurahan Banyurip	173.78	-10	Pekalongan Selatan	1.69	4.87	2
4	Kelurahan Bendan Kergon	112.10	10	Pekalongan Barat	0.95	3.62	1
...
95	Kelurahan Sokoduwet	286.39	-10	Pekalongan Selatan	2.51	13.08	2
96	Kelurahan Sokoduwet	286.39	-10	Pekalongan Timur	0.01	13.08	2
97	Kelurahan Tirta	146.98	10	Pekalongan Barat	28.53	3.58	1
98	Kelurahan Tirta	146.98	-10	Pekalongan Barat	85.57	3.58	2
99	Kelurahan Tirta	146.98	-10	Pekalongan Barat	31.85	3.58	2

100 rows x 7 columns

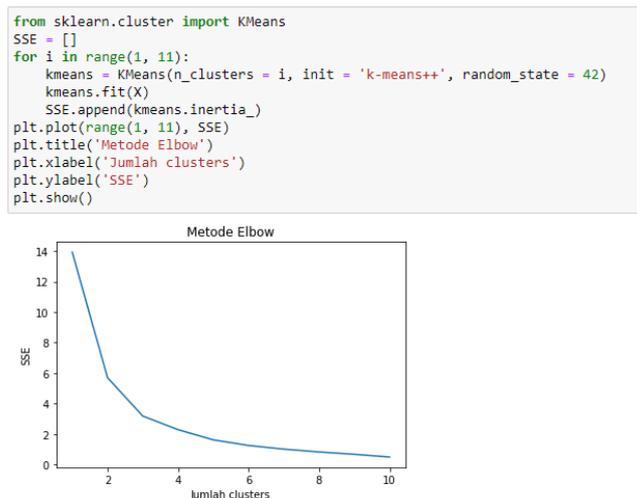
Gambar 9. Proses Clustering

3.4. Evaluasi

Tahap evaluasi ini bertujuan untuk menguji pemodelan k-means clustering apakah sudah baik atau belum dalam menentukan jumlah cluster. Proses evaluasi hasil cluster dilakukan dengan metode elbow dan silhouette. Hasil evaluasi menggunakan elbow method ditunjukkan oleh Gambar 10 berikut ini.

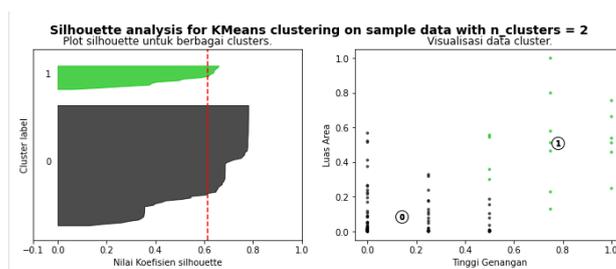
Pada Gambar 10, tampak syntax untuk mencari SSE (Sum of Square Error) dari masing-masing cluster dan didapatkan titik siku pada cluster nomor 3. Titik siku menunjukkan jumlah cluster yang optimal. Sehingga dari metode Elbow dapat disimpulkan bahwa jumlah cluster yang optimal untuk proses clustering adalah 3 cluster.

Pengujian selanjutnya adalah menggunakan metode Silhouette yang dilakukan dengan membandingkan data apabila dibagi menjadi beberapa cluster, dalam penelitian ini pembagian cluster akan dibagi menjadi 5 bagian dalam metode Silhouette yaitu dari cluster 2 sampai dengan 6. Pada proses ini nantinya data akan



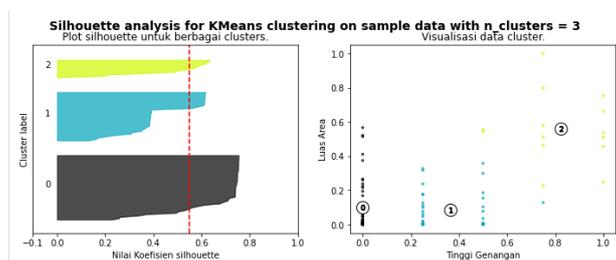
Gambar 10. Elbow Method

dilihat dari persebaran data apakah sudah optimal dalam pembagian cluster dilihat dari jarak terdekat setiap cluster dan apakah ada nilai negatif yang ada dalam cluster tersebut.



Gambar 11. Silhouette 2 Cluster

Gambar 11 diatas merupakan pembagian metode *Silhouette* dengan 2 cluster. Gambar tersebut menunjukkan bahwa persebaran data pada cluster ini masih berjauhan setiap clusternya. Jadi perlu adanya evaluasi untuk jumlah cluster.

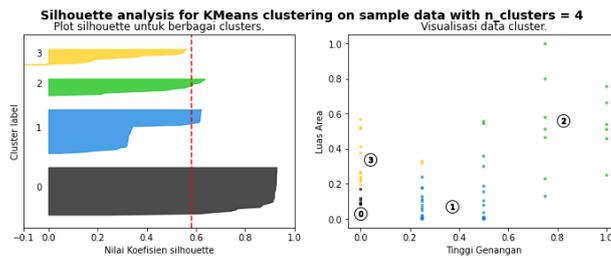


Gambar 12. Silhouette 3 Cluster

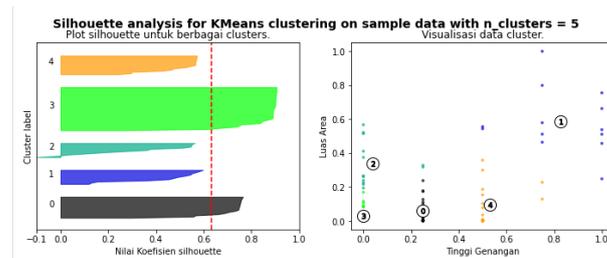
Pada saat jumlah cluster diubah menjadi 3 cluster, maka persebaran data yang sudah dekat di setiap cluster dan tidak ada nilai negative dalam setiap cluster. Hal tersebut dapat dilihat di Gambar 12.

Hasil yang ditunjukkan oleh cluster 4 pada Gambar 13, nilai persebaran data sudah dekat dari setiap cluster akan tetapi masih memiliki nilai negatif pada cluster ke-3.

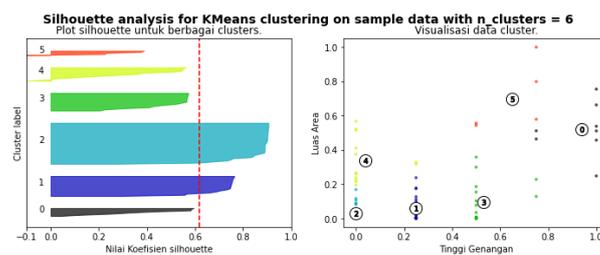
Pembagian data dengan jumlah cluster 5, seperti yang ditunjukkan oleh Gambar 14, menghasilkan persebaran data yang lebih dekat karena jumlah cluster yang banyak akan tetapi memiliki nilai negatif pada cluster ke-2.



Gambar 13. Silhouette 4 Cluster



Gambar 14. Silhouette 5 Cluster



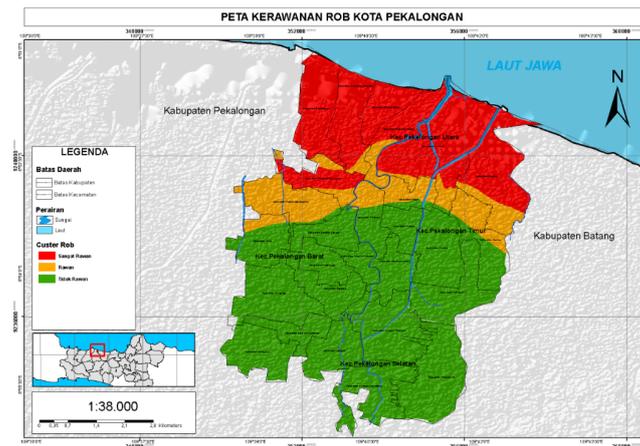
Gambar 15. Silhouette 6 Cluster

Pada data dengan jumlah cluster sebanyak 6 cluster seperti pada Gambar 15, persebaran data menjadi lebih dekat dibandingkan cluster lainnya tetapi dalam cluster ini memiliki nilai negatif yang terdapat dalam cluster ke-4 dan ke-5.

Setelah melakukan proses evaluasi dengan metode silhouette didapatkan hasil bahwa cluster yang tidak memiliki nilai negatif adalah jumlah cluster 2 dan 3, dengan melihat dan mempertimbangkan hasil dari metode Elbow, maka cluster yang paling optimal dalam proses clustering pada penelitian ini berjumlah 3 cluster.

3.5. Mapping

Setelah mendapatkan data dari hasil clustering, tahap selanjutnya adalah memindahkan data hasil clustering kedalam software ArcGis untuk dilakukan Mapping data. Pada tahap ini proses yang dilakukan adalah dengan menggabungkan hasil dari clustering ke dalam attribute tabel yang sebelumnya sudah dilakukan proses overlay untuk pembagian cluster di setiap kelurahan yang ada pada Kota Pekalongan. Hasil Mapping ditunjukkan oleh Gambar 16 berikut ini.



Gambar 16. Hasil Mapping

Pada gambar 16 tersebut di atas, area yang berwarna merah adalah cluster 0, yakni daerah yang memiliki tingkat kerawanan rob paling tinggi. Cluster 0 ini meliputi 7 kelurahan yaitu kelurahan Bandengan, Degayu, Kandang panjang, Krapyak, Padukuhan kraton, Panjang wetan, Panjang baru.

Area yang berwarna kuning merupakan cluster 1, yakni daerah yang memiliki tingkat kerawanan rob sedang. Cluster 1 ini meliputi 11 kelurahan yaitu kelurahan Bendan kergon, Degayu, Gamer, Kandang panjang, Kauman, Klego, Krapyak, Padukuhan kraton, Panjang wetan, Pasir kraton kramat, Tirto.

Area yang berwarna hijau merupakan cluster 2, yakni daerah yang memiliki tingkat kerawanan rob rendah. Cluster 2 ini meliputi 22 kelurahan yaitu Banyurip, Bendan Kergon, Buaran kradenan, Gamer, Jenggot, Kali baros, Kauman, Klego, Krapyak, Kuripan kertoharjo, Kuripan yosorejo, Medono, Noyontaansari, Padukuhan kraton, Pasir kraton kramat, Podosugih, Poncol, Pringrejo, Sapuro Kebulen, Setono, Soko duwet, Tirto.

Ada beberapa kelurahan yang terbagi menjadi 2 cluster, hal ini disebabkan karena ketinggian tanah yang berbeda didalam satu kelurahan dan juga tinggi genangan di daerah tersebut. Dalam penelitian ini semakin tinggi genangan rob dan Luas area genangan rob menjadi faktor utama kerawanan rob di Kota Pekalongan.

3.6. Subjudul 1

1. First level item
2. First level item
 - (a) Second level item
 - (b) Second level item
 - i. Third level item
 - ii. Third level item
 - A. Fourth level item
 - B. Fourth level item

3.6.1. Subjudul 2

Untuk daftar bernomor, gunakan a. b. c. dan seterusnya

4. Simpulan

Dari penelitian ini dapat disimpulkan bahwa:

- a. Tingkat kerawanan rob di kota pekalongan dapat dikelompokkan menggunakan algoritma k-means clustering menjadi 3 cluster dengan menggunakan parameter tinggi genangan dan luas area genangan

- b. Hasil dari penelitian ini menunjukkan bahwa daerah dengan kerawanan tinggi bencana rob di Kota Pekalongan ada 7 kelurahan, daerah dengan kerawanan rob sedang ada 11 kelurahan dan daerah dengan kerawanan rob rendah ada 22 kelurahan.

Semua referensi harus dikutip dalam makalah dengan referensi minimal 15, disarankan untuk menggunakan alat referensi seperti Mendeley atau enote, Minimal 80% referensi berasal dari jurnal dan diterbitkan dalam 5 tahun terakhir. Gunakan mendeley atau *reference tool* lainnya untuk mengisi daftar isi pada *file template.bib*

Pustaka

- [1] S. Miftakhudin, “Strategi penanganan banjir rob kota pekalongan,” *JURNAL LITBANG KOTA PEKALONGAN*, vol. 19, no. 1, 2021.
- [2] M. A. Salim and A. B. Siswanto, “Kajian penanganan dampak banjir kabupaten pekalongan,” *Rang Teknik Journal*, vol. 4, no. 2, p. 295–303, 2021.
- [3] S. W. d. B. R. S. Adi Iskandar, M. Helmi. Analisis geospasial area genangan banjir rob dan dampaknya pada penggunaan lahan tahun 2020-2025 di kota pekalongan provinsi jawa tengah. [Online]. Available: <http://ejournal2.undip.ac.id/index.php/ijoice/Diterima>
- [4] C. Nisa1 and W. Yustanti, “Studi perbandingan algoritma klastering dalam pengelompokan persediaan produk (studi kasus subdirektorat perencanaan sarana prasarana dan logistik ptn x),” *JEISBI*, vol. 02, 2021.
- [5] U. K. I. Kamila, “Perbandingan algoritma k-means dan k-medoids untuk pengelompokan,” *Jurnal Ilmiah Rekayasa dan Manajemen Sistem Informasi*, vol. 5, no. 1, pp. 119–125, 2019.
- [6] C. S. Santoso and A. I. Nurhidayat. Sistem informasi real time gis untuk monitoring sistem cors (continously operating reference station) di kantor wilayah badan pertahanan provinsi jawa timur. [Online]. Available: <https://id.m.wikipedia.org/wiki/JSON>
- [7] H. M. L. Petra Refialy and M. S. Pesulima, “Perbaikan kinerja clustering k-means pada data ekonomi nelayan dengan perhitungan sum of square error (sse) dan optimasi nilai k cluster,” *Techno.Com*, vol. 20, no. 2, pp. 321–329, 2021.
- [8] N. H. R. Kesuma Dinata and N. Azizah, “Analisis k-means clustering pada data sepeda motor,” *INFORMAL*, vol. 5, no. 1, pp. 10–17, 2020.
- [9] E. M. S. R. M. A. Syakur, B. K. Khotimah and B. D. Satoto, “Integration k-means clustering method and elbow method for identification of the best customer profile cluster,” *IOP Conference Series: Materials Science and Engineering*, vol. 336, no. 1, pp. 891–921, 1905.
- [10] K. R. Shahapure and C. Nicholas, “Cluster quality analysis using silhouette score,” *IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*, 2020.
- [11] J. D. K. J. P. F. Wang, H.-H. Franco-Penya and R. Ross. An analysis of the application of simplified silhouette to the evaluation of k-means clustering validity. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-62416-7_21
- [12] M. J. R. D. Indraswari, N. Hanifah and Y. Priyana, “Analisis aplikasi arcgis 10.3 untuk pembuatan daerah aliran sungai dan penggunaan lahan di das samajid kabupaten sampang, madura,” in *Prosiding Seminar Nasional Geografi UMS IX 2018*, vol. IX, Surakarta, Indonesia, Juni 2018, pp. 478–489.
- [13] A. S. W. D McGinnis, C. Siu and H. Huang, “Category encoders: a scikit-learn-contrib package of transformers for encoding categorical data,” *The Journal of Open Source Software*, vol. 3, no. 21, p. 501, 2018.
- [14] H. H. K. D. Azzahra Nasution and N. Chamidah, “Perbandingan normalisasi data untuk klasifikasi wine menggunakan algoritma k-nn,” *Journal of Computer Engineering, System, and Science*, vol. 4, no. 1, pp. 78–82, 2019.
- [15] N. G. Ramadhan and A. Khoirunnisa, “Klasifikasi data malaria menggunakan metode support vector machine,” *Jurnal Media Informatika Budidarma*, vol. 5, no. 4, p. 1580, 2021.