

# *Clustering* Daerah Rawan Angin Puting Beliung pada Kabupaten di Sumatera Utara dengan Algoritma *K-Means*

Rizky Syahrul Ramadhan, Rakhmat Kurniawan R dan Muhammad Siddik Hasibuan

Program Studi Ilmu Komputer, Fakultas Sains dan Teknologi,  
Universitas Islam Negeri Sumatera Utara

E-mail: riskisyahrul99@gmail.com, rakhmat.kr@uinsu.ac.id, muhammadsiddik@uinsu.ac.id.

## Abstrak

Provinsi Sumatera Utara merupakan daerah dengan topografi yang berbeda-beda di tiap kabupatennya dimana terdiri dari dataran tinggi, dataran rendah, pegunungan dan pantai. Dengan keadaan seperti itu sering terjadi bencana alam salah satunya angin puting beliung. Masalah yang terjadi adalah bagaimana mengetahui daerah yang rawan angin puting beliung agar mengurangi kerugian dan korban jiwa. Salah satu cara yang dapat dilakukan adalah untuk mengetahui daerah yang rawan angin puting beliung adalah dengan menggunakan teknik *data mining*. Adapun metode yang digunakan dalam penelitian ini adalah algoritma *k-means clustering*. Algoritma *k-means clustering* dapat mengelompokkan data yang memiliki karakteristik yang sama menjadi satu kelompok dan data yang memiliki karakteristik yang berbeda dikelompokkan menjadi kelompok lainnya. Data yang digunakan yaitu parameter angin puting beliung dari 33 kabupaten/kota dan 181 column data. Hasilnya diperoleh cluster 0 atau daerah dengan potensial rawan bencana angin puting beliung rendah sebanyak 19 daerah dan cluster 1 atau daerah dengan potensial rawan bencana angin puting beliung tinggi sebanyak 14 daerah.

**Kata kunci:** Angin puting beliung, *K-means*, *Cluster*

## Pendahuluan

Provinsi Sumatera Utara memiliki 33 Kabupaten/Kota dengan topografi yang berbeda-beda di tiap kabupatennya dimana terdiri dari dataran tinggi, dataran rendah, pegunungan dan daerah pantai dengan iklim yang bervariasi pada tiap dataran dengan sebagian iklim yang cukup panas dan sebagian lagi iklim sedang. Dengan topografi dan iklim yang bervariasi tersebut, Sumatera Utara memiliki potensi untuk terkena bencana alam, salah satunya angin puting beliung. Angin puting beliung adalah angin kencang yang muncul secara tiba-tiba, mempunyai pusat dan bergerak melingkar seperti spiral hingga menyentuh permukaan bumi [1]. Faktor pemanasan global merupakan pendukung terjadinya perubahan pola arah angin yang tidak beraturan dan sulit untuk diprediksi serta waktu terjadinya cenderung singkat. Dibutuhkan cara untuk mengidentifikasi daerah rawan terjadinya angin puting beliung untuk melakukan tindakan yang mengurangi kerugian, salah satunya dapat menggunakan *data mining*.

*Data mining* merupakan teknik yang meng-

gabungkan teknik analisis data dan menemukan pola-pola yang penting pada data [2]. *Data mining* dapat melakukan kegiatan seleksi, eksplorasi dan pemodelan dari jumlah data yang besar untuk menemukan kecenderungan yang biasanya tidak disadari keberadaannya oleh pengguna. Hasilnya, *data mining* mampu menganalisa data yang besar menjadi informasi yang mempunyai arti bagi pendukung keputusan [3]. Adapun metode dalam *data mining* yang digunakan untuk mengidentifikasi potensi daerah rawan angin puting beliung yaitu *clustering* menggunakan algoritma *k-means*.

Algoritma *k-means clustering* merupakan algoritma *unsupervised learning* yang mengelompokkan data kedalam beberapa kelompok berdasarkan karakteristik atau ciri-ciri data. Tujuan utama *k-means clustering* adalah mengelompokkan data berdasarkan kemiripan pada objek data dan sebaliknya meminimalkan kemiripan terhadap kelas yang lain [4]. Metode *clustering* dengan algoritma *k-means* dipilih dalam penelitian ini karena kemampuannya yang dapat mengelompokkan berdasarkan kemiripan objek. Tujuan dari penelitian ini diperoleh *clustering* potensi daerah rawan angin put-

ing beliung yang dapat digunakan untuk mengembangkan strategi mitigasi dan penanganan yang lebih presisi serta evaluasi kondisi alam.

## Metode Penelitian

### Algoritma *K-Means Clustering*

Algoritma *k-means clustering* ditemukan oleh beberapa orang yaitu Lloyd pada tahun 1957 dan 1982, Forgey tahun 1965, Friedman dan Rubin tahun 1967 serta McQueen tahun 1967. Konsep awal ditemukan oleh Lloyd tahun 1957, tetapi baru dipublikasikan pada tahun 1982. Pada tahun 1965, Forgey mempublikasikan teknik yang sama dengan nama yang dikenal Lloyd-Forgey pada beberapa sumber [4].

Algoritma *k-means clustering* adalah algoritma *unsupervised learning*. *Unsupervised learning* yaitu metode dimana tidak satupun atribut yang digunakan untuk memandu proses pembelajaran [5]. Metode *unsupervised learning* ini seperti metode yang bekerja sendiri dan memahami pola. Metode ini tidak membutuhkan data training dan hasil pada metode ini tidak diketahui oleh siapapun. Pada *unsupervised learning* hanya ada data masukan (*input*) dan parameter tanpa adanya kelas atau label (*actual output*) [6].

Algoritma *k-means* hanya bisa menggunakan data bersifat numerik atau angka dalam mengolah atau pengaplikasiannya pada data. Ketidakmiripan suatu data dengan kelompok dapat direpresentasikan dengan jarak antara data dengan *centroid* (titik pusat) *cluster*. Jumlah terbaik dari kelompok tergantung pada jarak yang disebut apriori [7]. Apabila jarak suatu data dengan *centroid* atau titik pusat *cluster* jauh, maka ketidakmiripannya tinggi. Namun apabila jika jarak data dengan *centroid* atau titik pusat *cluster* dekat, maka tingkat ketidakkemiripan dengan datanya rendah. Berikut ini adalah *pseudocode* metode algoritma *k-means*.

```

Algoritma k-means clustering
Input:  $x, k, n, s$ 
Output:  $c$ 
Procedure
 $C \leftarrow 0;$ 
For  $i \leftarrow$  to  $n$ 
 $S_i \leftarrow \text{RandomSample}(x, s)$ 
 $(a_i, c_i) \leftarrow \text{kmeans}(S_i, k)$ 
 $C \leftarrow \text{AppendRows}(c, c_i)$ 
End for
Return  $c_{\text{final}} \leftarrow \text{kmeans}(c, k)$ 
End Procedure
    
```

Gambar 1: Pseudocode Algoritma K-Means [8]

Berdasarkan Gambar 1, dapat diperoleh

langkah-langkah algoritma *k-means* adalah sebagai berikut:

1. Mempersiapkan data yang akan di *cluster*.
2. Menentukan jumlah  $k$  atau *cluster* sesuai yang diinginkan.
3. Memilih titik pusat awal *cluster* (*centroid*) secara acak.
4. Menghitung jarak data dengan *centroid* dengan metode *Euclidean Distance*.

$$d_{ij} = \sqrt{(x_{1j} - Y_{1k})^2 + (x_{2j} - Y_{2k})^2 + \dots + (x_{ij} - Y_{jk})^2} \quad (1)$$

Bahwa  $d_{ij}$  merupakan jarak antara  $x_i$  dan  $x_j$ ,  $X_{ij}$  adalah nilai data titik  $i$  pada atribut data ke  $j$ , dan  $Y_{jk}$  adalah nilai data titik pusat *cluster* pada atribut data ke  $j$ .

5. Tempatkan data ke dalam *cluster* terdekat dengan cara menghitung jarak terpendek data dengan *centroid*.
6. Tentukan kembali *centroid* baru dengan menghitung rata-rata dari tiap data yang berada di *cluster* yang sama.

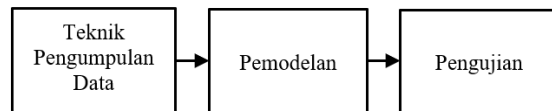
$$\mu_j(t+1) = \frac{1}{N_{S_j}} \sum_{j \in S_j} X_j \quad (2)$$

Dimana  $\mu_j(t+1)$  adalah *centroid* baru pada iterasi ke  $(t+1)$  dan  $N_{S_j}$  adalah banyaknya data dalam *cluster*  $S_j$ .

7. Lakukan kembali perhitungan jarak tiap data dengan *centroid* baru dengan metode *Euclidean Distance*. Lakukan evaluasi sampai data tidak berpindah *cluster* lagi.

### Tahapan Penelitian

Tahap-tahap dalam metode penelitian ini terdiri dari 3 bagian, dapat dilihat pada Gambar 2.

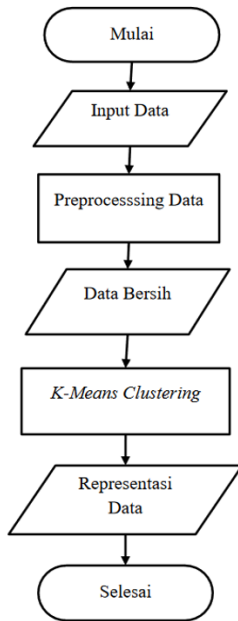


Gambar 2: Tahapan Penelitian

1. Metode pengumpulan data dilakukan dengan studi literatur dan pengambilan sumber data sebagai dataset yang akan digunakan untuk klasterisasi. Pada penelitian ini studi literatur digunakan untuk menemukan referensi yang terkait atau sumber yang diidentifikasi dengan masalah yang diteliti sebagai acuan dan pembanding penelitian. Referensi didapat dengan cara mencari jurnal, ebook, tugas akhir dan situs internet. *Output* dari studi literatur ini adalah terkoleksinya

referensi yang relevan dengan perumusan masalah [9]. Sumber data pada penelitian ini didapat dengan cara pengambilan data dari website <https://power.larc.nasa.gov/> dan <https://earth.google.com/> dimana data parameter penyebab angin puting beliung yang digunakan yaitu suhu udara (C), kecepatan angin (m/s), curah hujan (mm/day), kelembapan spesifik (g/kg), arah angin (derajat) dan ketinggian daratan (mdpl) pada tiap bulan dari tahun 2019 sampai tahun 2021 di tiap kabupaten/kota dengan ekstension file .csv.

2. Pemodelan adalah pembuatan model matematika yang mewakili dari sebuah sistem yang nyata baik secara kualitatif dan dimana mampu menggambarkan secara jelas hubungan serta interaksi antara faktor penggerak sistem. Sistem nyata adalah sistem yang sedang berlangsung di dunia nyata dan menjadi permasalahan yang sedang diteliti [10]. Tahap pemodelan ini dimulai dengan menginput data berupa parameter bencana angin puting beliung ke dalam file ekstension .csv. Kemudian melakukan *preprocessing* data yaitu melakukan penyetaraan data yang tidak konsisten agar datanya tidak mendominasi hasil. Setelah *preprocessing* data, diperoleh data bersih dan selanjutnya penerapan *clustering* menggunakan algoritma *k-means*. Tahap terakhir yaitu merepresentasikan data kedalam bentuk visualisasi. Tahapan pemodelan pada penelitian ini dapat dilihat pada Gambar 3.



Gambar 3: Flowchart Pemodelan Sistem

3. Pengujian pada penelitian ini dilakukan menggunakan bahasa pemrograman python.

Pengujian bertujuan mengetahui apakah sistem berjalan dengan baik dan sesuai. Dalam penelitian ini juga peneliti menggunakan *library* sebagai pembantu dalam mengolah dan eksekusi data.

## Hasil dan Pembahasan

Dalam penelitian ini pengujian dilakukan dengan menggunakan metode k-means clustering menggunakan bahasa pemrograman python dengan google colabatory sebagai editor dan eksekusi program yang sifatnya berbasis cloud. Berikut ini adalah penerapan algoritma k-means menggunakan bahasa pemrograman python pada data parameter penyebab angin puting beliung:

### 1. Import Data

Pada Gambar 4 adalah *library* yang digunakan untuk *clustering* data menggunakan algoritma *k-means* dimana terdapat beberapa *library* yaitu *library* *pandas* berfungsi sebagai membaca dan menulis data, membersihkan dan mengolah data serta manipulasi data, *numpy* berfungsi untuk melakukan operasi matematika numerik yang bersifat *array* serta matriks, *matplotlib* berfungsi sebagai memvisualisasikan data ke dalam bentuk grafik, animasi ataupun interaktif dan *scikit learn* (*sklearn*) berfungsi sebagai pelatihan model data dengan *machine learning* dan menyederhanakan proses membangun, melatih dan mengevaluasi model *machine learning* dengan memanfaatkan *library numpy* dan *pandas* untuk penanganan data yang efisien dan perhitungan numerik dalam alur kerja kerja machine learning.

```

[ ] #import library
import pandas as pd
import numpy as np
from matplotlib import pyplot as plt
from sklearn.preprocessing import MinMaxScaler
from sklearn.cluster import KMeans
  
```

Gambar 4: Library yang digunakan

### 2. Menampilkan Data

Gambar 5 adalah perintah yang digunakan untuk menampilkan dataset yang telah *input* pada google collab. Pada Gambar 6 adalah tampilan dataset yang telah di *import* dalam bentuk *table* melalui perintah yang digunakan pada Gambar 5.

```

#Menampilkan data file csv
df = pd.read_csv(r'data terakhir.csv', delimiter = ';')
df
  
```

Gambar 5: Membaca Data

Kabupaten/Kota	Suhu Udara (C) Januari 2019	Suhu Udara (C) Februari 2019	Suhu Udara (C) Maret 2019	Suhu Udara (C) April 2019	Suhu Udara (C) Mei 2019	Suhu Udara (C) Juni 2019	Suhu Udara (C) Juli 2019	Suhu Udara (C) Agustus 2019	Suhu Udara (C) September 2019	Suhu Udara (C) Oktober 2019	Arah Angin (derajat) April 2021	Arah Angin (derajat) Mei 2021	Arah Angin (derajat) Juni 2021	Arah Angin (derajat) Juli 2021	Arah Angin (derajat) Agustus 2021	Arah Angin (derajat) September 2021	Arah Angin (derajat) Oktober 2021	Arah Angin (derajat) November 2021	Arah Angin (derajat) Desember 2021	Ketinggian Daratan (mop)
0 Nias	27.62	28.15	28.15	28.26	28.78	28.31	28.26	27.87	27.63	—	281.31	272.31	276.19	283.00	293.62	273.81	263.31	279.25	278.25	53
1 Mandailing Natal	23.96	24.52	24.34	24.62	24.62	24.23	23.95	23.74	24.18	—	255.88	245.81	246.94	250.75	256.25	253.31	249.38	254.06	282.69	387
2 Tapanuli Selatan	23.32	23.90	23.83	24.21	24.12	23.85	23.40	23.22	23.58	—	252.12	240.62	249.81	251.88	250.38	250.00	251.44	254.00	267.50	543
3 Tapanuli Tengah	23.04	23.62	23.42	23.77	23.66	23.03	23.00	22.59	23.08	—	250.25	240.94	248.38	259.50	260.50	250.12	251.81	255.44	263.50	17
4 Tapanuli Utara	23.04	23.62	23.42	23.77	23.66	23.03	23.00	22.59	23.08	—	250.25	240.94	248.38	259.50	260.50	250.12	251.81	255.44	263.50	1302
5 Tobo	22.76	23.23	23.58	24.15	23.88	23.63	23.42	23.33	23.40	—	292.62	79.62	240.81	245.50	267.75	241.38	240.88	253.88	323.56	922
6 Labuhanbatu	25.72	26.18	26.46	26.80	26.66	26.19	26.06	26.12	26.23	—	244.88	159.62	210.56	204.62	197.62	199.44	223.25	249.00	341.50	34
7 Asahan	25.49	26.13	26.69	27.12	26.69	26.55	26.39	26.55	26.33	—	346.19	59.19	302.12	272.50	346.81	7.56	328.75	318.94	325.12	22
8 Simalungun	21.90	22.39	22.58	23.30	23.01	22.39	22.50	22.31	22.38	—	281.75	30.88	231.12	232.19	255.62	241.88	248.06	255.12	312.81	642
9 Dairi	21.90	22.39	22.58	23.30	23.01	22.39	22.50	22.31	22.38	—	281.75	30.88	231.12	232.19	255.62	241.88	248.06	255.12	312.81	1543
10 Karo	21.90	22.39	22.58	23.30	23.01	22.39	22.50	22.31	22.38	—	281.75	30.88	231.12	232.19	255.62	241.88	248.06	255.12	312.81	1228
11 Deli Serdang	25.41	25.98	26.42	27.08	26.78	26.83	26.58	26.66	26.32	—	301.81	78.06	214.31	207.62	264.75	182.94	264.62	268.06	323.06	15
12 Langkat	22.72	23.19	23.48	24.21	23.73	23.35	23.23	23.14	23.08	—	177.50	132.69	217.38	220.31	206.88	198.94	218.75	214.81	6.44	10
13 Nias Selatan	27.62	28.15	28.15	28.26	28.78	28.31	28.26	27.87	27.63	—	281.31	272.31	276.19	283.00	293.62	273.81	263.31	279.25	278.25	204
14 Humbang Hasundutan	20.33	20.80	20.76	21.46	21.21	20.22	20.45	20.03	20.39	—	261.62	240.38	237.12	244.06	252.06	243.44	240.81	247.88	282.81	1405
15 Paspak Bharat	24.02	24.65	24.55	24.87	24.83	24.23	24.29	23.98	24.17	—	259.88	250.31	245.94	260.06	269.94	256.56	255.12	255.69	272.06	981
16 Sumsel	20.33	20.80	20.76	21.46	21.21	20.22	20.45	20.03	20.39	—	332.44	240.38	237.12	244.06	252.06	243.44	240.81	247.88	282.81	994
17 Serdang Bedagai	27.76	28.33	29.13	29.50	29.05	28.78	28.65	28.88	28.47	—	332.44	67.50	150.56	134.75	311.75	65.94	329.19	315.06	326.62	4
18 Batu Bara	25.49	26.13	26.69	27.12	26.69	26.55	26.39	26.55	26.33	—	346.19	59.19	302.12	272.50	346.81	7.56	328.75	318.94	325.12	26
19 Padang Lawas Utara	25.58	26.11	26.17	26.81	26.58	26.07	25.92	25.83	26.08	—	245.25	220.81	229.44	224.62	227.06	227.38	239.56	244.94	301.69	105
20 Padang Lawas	23.96	24.52	24.34	24.62	24.62	24.23	23.95	23.74	24.18	—	255.88	245.81	246.94	250.75	256.25	253.31	249.38	254.06	282.69	285
21 Labuhanbatu Selatan	25.72	26.18	26.46	26.80	26.66	26.19	26.06	26.12	26.23	—	244.88	159.62	210.56	204.62	197.62	199.44	223.25	249.00	341.50	44
22 Labuhanbatu Utara	22.76	23.23	23.58	24.15	23.88	23.63	23.42	23.33	23.40	—	292.62	79.62	240.81	245.50	267.75	241.38	240.88	253.88	323.56	34

Gambar 6: Tampilan Dataset yang telah Diimport

### 3. Preprocessing Data

Tahap *preprocessing* data dimulai dengan melakukan normalisasi data menggunakan metode *minmax scaler*. *Minmax scaler* adalah teknik mengubah atau mentransformasikan skala nilai data dengan menjadi lebih kecil tanpa melakukan modifikasi atau perubahan isi data dengan rentang nilai tepi batas atas data 1 dan nilai tepi bawah data 0. Tujuan normalisasi adalah data yang digunakan setara tanpa mendistorsi perbedaan dalam rentang nilai, tidak ada data yang mendominasi hasil serta menghilangkan perbedaan diantara data dimana dapat meningkatkan kinerja *machine learning*.

```
#Melakukan preprocessing data
X = df.iloc[:, 1:182]
normalisasi = MinMaxScaler()
Y = normalisasi.fit_transform(X)
Y
```

Gambar 7: Perintah Normalisasi Data

Pada Gambar 7 adalah perintah untuk melakukan normalisasi pada dataset yang telah *diimport* menggunakan metode *minmax scaler* dimana dataset yang digunakan dimulai dari baris ke-2 sampai baris ke-183 dan hasil normalisasi data disimpan didalam variabel Y.

Pada Gambar 8 adalah hasil normalisasi data menggunakan metode *minmax scaler* dimana hasilnya dalam bentuk *array* multidimensi. Data ini yang akan digunakan pada penerapan algoritma *k-means*.

```
array([[0.95795007, 0.96330275, 0.88291517, ..., 0.61884183, 0.81122784,
0.03183886],
[0.47700394, 0.48754915, 0.42771804, ..., 0.37693268, 0.8244792 ,
0.2488629 ],
[0.39290407, 0.40629096, 0.36678614, ..., 0.37635648, 0.77914403,
0.35022742],
...,
[0.66754271, 0.67889908, 0.67622461, ..., 0.70344761, 0.94496508,
0.01949318],
[0.39290407, 0.40629096, 0.36678614, ..., 0.37635648, 0.77914403,
0.20012995],
[1. , 1. , 0.91995221, ..., 0.59483338, 0.80507969,
0.11500975]])
```

Gambar 8: Hasil Normalisasi Data

```
[ ] #Inisiasi algoritma k-means
kmeans = KMeans(n_clusters = 2, random_state = 100, max_iter = 10 )
kmeans.fit(Y)
labels = kmeans.labels_
kmeans.labels_
```

Gambar 9: Menjalankan Algoritma K-Means

```
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870:
warnings.warn(
array([1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 1, 0, 0, 0, 1, 1, 0, 0, 0,
0, 1, 1, 0, 1, 1, 1, 1, 1, 0, 1], dtype=int32)
```

Gambar 10: Label Cluster

### 4. Menjalankan Algoritma K-Means

Pada Gambar 9 merupakan perintah untuk memulai atau inisiasi algoritma *k-means* pada dataset yang disimpan dalam variabel Y dimana dengan jumlah *cluster* adalah 2.

Pada Gambar 10 terlihat label *cluster* yang didapat dengan melalui inisiasi algoritma *k-means* pada dataset penyebab angin puting beliung yang

berjumlah 2 *cluster*, dimana label *cluster* 0 adalah 1 dan label *cluster* 1 adalah 2.

```
[ ] kmeans.cluster_centers_
```

Gambar 11: Perintah Mendapatkan Centroid Cluster

Gambar 11 adalah perintah yang digunakan untuk mendapatkan titik pusat (*centroid*) pada sebuah dataset yang telah di *cluster*. Hasilnya dapat dilihat pada Gambar 12.

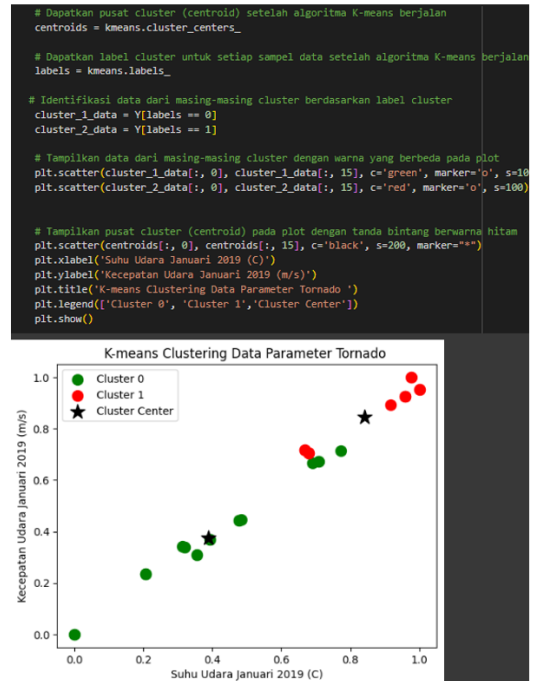
```
array([[0.38882357, 0.39394357, 0.37074766, 0.35755433, 0.378558,
0.39996311, 0.38087291, 0.37294913, 0.3974075, 0.38382712,
0.39836415, 0.40641291, 0.3767927, 0.36010736, 0.38049496,
0.37568058, 0.40520297, 0.39341945, 0.39298246, 0.38495881,
0.38641475, 0.39847886, 0.4136633, 0.39976046, 0.37244996,
0.36842105, 0.37866041, 0.38149763, 0.3918533, 0.37739835,
0.37887648, 0.39486336, 0.39660974, 0.39694799, 0.41097289,
0.38702018, 0.25988416, 0.24499629, 0.21137521, 0.16721229,
0.39550332, 0.29227324, 0.28902765, 0.31904199, 0.1600726,
0.16283566, 0.24099723, 0.3258832, 0.22923977, 0.32068543,
0.17105263, 0.28333993, 0.18058734, 0.24458204, 0.20353331,
0.31829574, 0.31336565, 0.31926925, 0.25621747, 0.20741509,
0.23931624, 0.2002448, 0.18049933, 0.26926011, 0.21185876,
0.23379878, 0.25116099, 0.22883295, 0.27825077, 0.36104593,
0.3106383, 0.1375774, 0.64119171, 0.6445418, 0.65118179,
0.65421206, 0.91813847, 0.69778772, 0.78169336, 0.61935549,
0.57329274, 0.70139078, 0.67094252, 0.69931019, 0.6177597,
0.62335264, 0.65121548, 0.64914974, 0.61984004, 0.81934985,
0.68753463, 0.66737449, 0.73534839, 0.6600756, 0.52516663,
0.69359162, 0.30910338, 0.76325795, 0.61615545, 0.70600552,
0.56849266, 0.65813567, 0.57245054, 0.87850615, 0.74158023,
0.43773165, 0.91382456, 0.67016715, 0.50665457, 0.53675856,
0.49298675, 0.48112838, 0.47983595, 0.46665154, 0.47203551,
0.47273802, 0.47272509, 0.50981267, 0.51004527, 0.52308403,
0.51436782, 0.49736842, 0.48582996, 0.47876393, 0.50149862,
0.45968246, 0.48577884, 0.46565966, 0.46493373, 0.47368421,
```

Gambar 12: Centroid Cluster

5. Visualisasi Cluster Data

Gambar 13 adalah perintah yang digunakan untuk menampilkan hasil *cluster* serta visualisasi hasil

*cluster* data dalam bentuk plot *scatter*. Terdapat *cluster* 0 yang direpresentasikan dengan titik berwarna hijau dan *cluster* 1 yang direpresentasikan dengan titik berwarna merah. Sedangkan simbol bintang berwarna hitam merepresentasikan titik pusat (*centroid*) *cluster* data. Sumbu *x* pada grafik diberi label suhu udara pada bulan Januari di tahun 2019 dalam skala celcius dan sumbu *y* pada grafik diberi label kecepatan udara pada bulan Januari di tahun 2019 dalam skala meter/second.



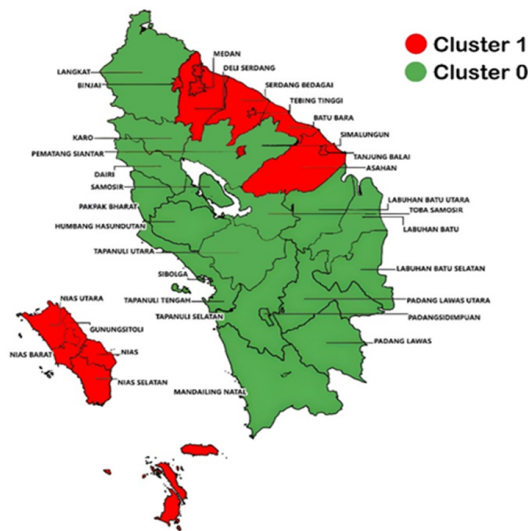
Gambar 13: Visualisasi Algoritma K-Means Pada Data

Kabupaten/Kota	Suhu Udara (C) Januari 2019	Suhu Udara (C) Februari 2019	Suhu Udara (C) Maret 2019	Suhu Udara (C) April 2019	Suhu Udara (C) Mei 2019	Suhu Udara (C) Juni 2019	Suhu Udara (C) Juli 2019	Suhu Udara (C) Agustus 2019	Suhu Udara (C) September 2019	...	Arah Angin (derajat) Mei 2021	Arah Angin (derajat) Juni 2021	Arah Angin (derajat) Juli 2021	Arah Angin (derajat) Agustus 2021	Arah Angin (derajat) September 2021	Arah Angin (derajat) Oktober 2021	Arah Angin (derajat) November 2021	Arah Angin (derajat) Desember 2021	Ketinggian (mtp)	Cluster	
0	Nias	27.62	28.15	28.25	28.78	28.31	28.28	27.87	27.63	...	272.31	278.19	283.00	293.62	273.81	263.31	279.25	278.25	53	1	
1	Mandailing Natal	23.96	24.52	24.34	24.62	24.62	24.23	23.95	23.74	24.18	...	245.81	246.94	250.75	256.25	253.31	249.38	254.06	282.89	387	0
2	Tapaneli Selatan	23.32	23.90	23.83	24.21	24.12	23.65	23.40	23.22	23.58	...	240.62	248.81	251.88	250.38	250.00	251.44	254.00	267.50	543	0
3	Tapaneli Tengah	23.04	23.62	23.42	23.77	23.66	23.03	23.00	22.59	23.08	...	240.94	245.38	250.50	260.50	250.12	251.81	255.44	263.50	17	0
4	Tapaneli Utara	23.04	23.62	23.42	23.77	23.66	23.03	23.00	22.59	23.08	...	240.94	245.38	250.50	260.50	250.12	251.81	255.44	263.50	1302	0
5	Toba	22.78	23.23	23.58	24.15	23.88	23.63	23.42	23.33	23.40	...	79.62	240.81	245.50	267.75	241.38	240.88	253.88	323.56	922	0
6	Labuhatbaru	25.72	26.18	26.46	26.80	26.66	26.19	26.06	26.12	26.23	...	199.62	210.56	204.62	197.62	199.44	223.25	240.00	341.50	34	0
7	Acaban	25.49	26.13	26.69	27.12	26.69	26.55	26.39	26.55	26.33	...	59.19	302.12	272.50	348.81	7.56	328.00	318.94	325.12	22	1
8	Simakrong	21.90	22.38	22.58	23.30	23.01	22.30	22.50	22.31	22.38	...	30.88	231.12	232.19	255.62	241.88	240.06	255.12	312.81	642	0
9	Dani	21.90	22.38	22.58	23.30	23.01	22.30	22.50	22.31	22.38	...	30.88	231.12	232.19	255.62	241.88	240.06	255.12	312.81	1543	0
10	Karo	21.90	22.38	22.58	23.30	23.01	22.30	22.50	22.31	22.38	...	30.88	231.12	232.19	255.62	241.88	240.06	255.12	312.81	1228	0
11	Deli Serdang	25.41	25.98	26.42	27.08	26.78	26.83	26.58	26.66	26.32	...	78.06	214.31	207.62	264.75	182.94	264.62	288.06	323.06	15	1
12	Langkat	22.72	23.19	23.48	24.21	23.73	23.35	23.23	23.14	23.08	...	132.69	217.38	220.31	208.88	198.94	218.75	214.81	6.44	10	0
13	Nias Selatan	27.62	28.15	28.15	28.26	28.78	28.31	28.28	27.87	27.63	...	272.31	278.19	283.00	293.62	273.81	263.31	279.25	278.25	204	1
14	Humbang Hasondutan	20.33	20.80	20.76	21.46	21.21	20.22	20.45	20.03	20.39	...	240.38	237.12	244.06	252.06	243.44	240.81	247.88	282.81	1405	0
15	Pangkal Bharat	24.02	24.65	24.55	24.87	24.83	24.23	24.29	23.98	24.17	...	258.31	245.94	269.06	269.94	256.56	255.12	255.69	272.06	981	0
16	Samosir	20.33	20.80	20.76	21.46	21.21	20.22	20.45	20.03	20.39	...	240.38	237.12	244.06	252.06	243.44	240.81	247.88	282.81	954	0
17	Serdang Bedagai	27.78	28.33	28.13	28.50	28.95	28.78	28.65	28.98	28.47	...	67.50	158.56	134.75	311.75	65.94	329.19	315.06	326.62	4	1
18	Batu Bara	25.49	26.13	26.69	27.12	26.69	26.55	26.39	26.55	26.33	...	59.19	302.12	272.50	348.81	7.56	328.00	318.94	325.12	26	1
19	Padang Lenses Utara	25.58	26.11	26.17	26.61	26.58	26.07	25.92	25.83	26.08	...	220.81	229.44	234.62	227.06	227.38	239.56	244.94	301.69	105	0
20	Padang Lenses Selatan	23.96	24.52	24.34	24.62	24.62	24.23	23.95	23.74	24.18	...	245.81	246.94	250.75	256.25	253.31	249.38	254.06	282.89	285	0
21	Labuhatbaru	25.72	26.18	26.46	26.80	26.66	26.19	26.06	26.12	26.23	...	199.62	210.56	204.62	197.62	199.44	223.25	240.00	341.50	44	0
22	Labuhatbaru	22.78	23.23	23.58	24.15	23.88	23.63	23.42	23.33	23.40	...	79.62	240.81	245.50	267.75	241.38	240.88	253.88	323.56	34	0

Gambar 14: Hasil Cluster Tiap Data

## 6. Hasil *Clustering* Algoritma *K-Means*

Pada Gambar 14 adalah hasil penerapan algoritma *k-means clustering* pada data penyebab bencana alam angin puting beliung yang berada di kolom *Cluster*. Dimana *cluster* 0 merepresentasikan daerah dengan rendah rawan bencana alam angin puting beliung dan *cluster* 1 merepresentasikan daerah dengan tinggi rawan bencana alam angin puting beliung. Hasil *cluster* menggunakan algoritma *k-means* didapati 19 data kategori *cluster* 0 dan 14 data kategori *cluster* 1. Hasil representasi *cluster* kerawanan angin puting beliung di tiap kabupaten/kota di Provinsi Sumatera Utara dapat dilihat pada Gambar 15.



Gambar 15: Peta Kerawanan Angin Puting Beliung Tahun 2019-2021

## Penutup

*Clustering* daerah rawan angin puting beliung pada kabupaten/kota di Sumatera Utara menghasilkan dua *cluster* yaitu *cluster* 0 yang mewakili daerah potensial rawan bencana angin puting beliung rendah terdiri dari 19 kabupaten/kota yaitu Mandailing Natal, Tapanuli Selatan, Tapanuli Tengah, Tapanuli Utara, Toba, Labuhanbatu, Simalungun, Dairi, Karo, Langkat, Humbang Hasundutan, Pakpak Bharat, Samosir, Padang Lawas Utara, Padang Lawas, Labuhanbatu Selatan, Labuhanbatu Utara, Sibolga dan Padang Sidempuan. *Cluster* 1 yang mewakili daerah dengan potensial rawan bencana angin puting beliung tinggi terdiri dari 14 kabupaten/kota yaitu Nias, Asahan, Deli Serdang, Nias Selatan, Serdang Bedagai, Batubara, Nias Utara, Nias Barat, Tanjung Balai, Pematang Siantar, Tebing Tinggi, Medan, Binjai, Gunungsitoli.

## Daftar Pustaka

- [1] Rohmat, "Penanggulangan Bencana Alam Klimatologis", Penerbit Duta, 2019.
- [2] N. Hasibuan, N. Silalahi, S. Darma Nasution, D. U. Sutiksno, H. Nurdiyanto dan E. Buulolo, "Implementasi Data Mining untuk Pengaturan Layout Minimarket dengan Menerapkan Association Rule", *J. Ris. Komput.*, vol. 4, no. 4, hal. 6–11, 2017.
- [3] L. Suriani, "Pengelompokan Data Kriminal pada Poldasu Menentukan Pola Daerah Rawan Tindak Kriminal menggunakan Data Mining Algoritma K-Means Clustering", *J. Sist. Komput. dan Inform.*, vol. 1, no. 2, hal. 151, doi: 10.30865/json.v1i2.1955, 2020.
- [4] Mustika Mustika, Yunita Ardilla, Abraham Manuhutu, Nazaruddin Ahmad, Imanuddin Hasbi, Guntoro Guntoro, Melda Agnes Manuhutu, Mohamad Ridwan, Hozairi Hozairi, Anindya Khrisna Wardhani, Syariful Alim, Ikhsan Romli, Yoga Religia, D. Tri Octafian, Unggul Utan Sufandi dan Iin Ernawati, "Data Mining dan Aplikasinya", Penerbit Widina, 2021.
- [5] A. M. Siregar dan A. Puspabhuana, *Data Mining: Pengolahan Data Menjadi Informasi dengan RapidMiner*, Surakarta: CV Kekata Group, 2017.
- [6] M. Z. Nasution dan M. S. Hasibuan, "Pendekatan Initial Centroid Search untuk Meningkatkan Efisiensi Iterasi Klustering K-Means", *Techno.Com*, vol. 19, no. 4, hal. 341–352, doi: 10.33633/tc.v19i4.3875, 2020.
- [7] R. Kurniawan, M. S. Hasibuan dan R. Hasibuan, "KLIK: Kajian Ilmiah Informatika dan Komputer Klasterisasi Wilayah Prioritas Vaksin Menggunakan Algoritma K-Means Clustering", *Media Online*, vol. 4, no. 3, hal. 1585–1592, doi: 10.30865/klik.v4i3.1334, 2023.
- [8] Y. Aswan, S. Defit, dan G. W. Nurcahyo, "Algoritma K-Means Clustering dalam Mengklasifikasi Data Daerah Rawan Tindak Kriminalitas (Polres Kepulauan Mentawai)", *J. Sistim Inf. dan Teknol.*, vol. 3, hal. 245–250, doi: 10.37034/jsisfotek.v3i4.73, 2021.
- [9] D. Pilendia, "Pemanfaatan Adobe Flash Sebagai Dasar Pengembangan Bahan Ajar Fisika: Studi Literatur", *J. Tunas Pendidik.*, vol. 2, no. 2, hal. 1–10, doi: 10.52060/pgsd.v2i2.255, 2020.
- [10] E. Syahrin, J. Santony dan J. Na'am, "Pemodelan Penjualan Produk Herbal menggunakan Metode Monte Carlo", *J. KomtekInfo*, vol. 5, no. 3, hal. 33–41, doi: 10.35134/komtekinfo.v5i3.29, 2019.