
Analisis *K-Means* dan *Naive Bayes* Untuk Pengelompokan Rawan Bencana di Daerah Kabupaten Labuhanbatu

Nadira Jannah Adeni Lubis¹, Syaiful Zuhri Harahap², Irmayanti Ritonga³

Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Labuhanbatu^{1,2,3}

Email : dlubis845@gmail.com¹, syaifulzuhriharahap@gmail.com²,
irmayantiritonga2@gmail.com³

Corresponding Author: dlubis845@gmail.com

Abstract

A natural disaster is an event that arises from the state of nature and has a significant impact on humans. Natural disasters can include events such as floods, volcanic eruptions, earthquakes, tsunamis, landslides, blizzards, droughts, hail, heat waves, hurricanes, tropical storms, typhoons, tornadoes, wildfires, and the spread of disease. Natural disasters that hit Labuhanbatu Regency include various types, such as floods, fires, tornadoes, and landslides. Each region within the district has specific characteristics associated with a particular type of natural disaster. In order to understand the level of vulnerability to disasters in Labuhanbatu District, K-Means and Naive Bayes methods are implemented to classify the level of vulnerability to frequent disasters. The results of this analysis will improve understanding of the level of vulnerability to disasters in Labuhan Batu Regency, enabling interested parties to identify areas that require increased attention in disaster mitigation and management efforts. In addition, the use of a combination of K-Means and Naive Bayes methods can serve as a solid basis for the development of more effective early warning systems in the future.

Keywords: *Analysis, K-Means, Naive Bayes, Natural Disasters.*

I. Pendahuluan

Bencana alam adalah suatu kejadian yang timbul dari keadaan alam dan memiliki dampak yang signifikan pada manusia. Bencana alam dapat mencakup peristiwa seperti banjir, letusan gunung berapi, gempa bumi, tsunami, tanah longsor, badai salju, kekeringan, hujan es, gelombang panas, hurikan, badai tropis, topan, tornado, kebakaran hutan, dan penyebaran penyakit. Bencana adalah suatu peristiwa atau rangkaian peristiwa yang menyebabkan atau meningkatkan

jumlah korban dan/atau merusak harta benda, infrastruktur, layanan penting, atau sarana kehidupan pada tingkat yang melebihi kapasitas norma. Bencana alam yang melanda Kabupaten Labuhanbatu mencakup berbagai jenis, seperti banjir, kebakaran, puting beliung, dan tanah longsor. Tiap daerah dalam kabupaten ini memiliki karakteristik khusus terkait dengan jenis bencana alam tertentu. Dalam rangka memahami tingkat kerentanan terhadap bencana di Kabupaten Labuhanbatu, metode *K-Means* dan *Naive Bayes*

diimplementasikan untuk mengelompokkan tingkat kerentanan terhadap bencana yang sering terjadi. *K-Means* merupakan metode pengelompokan non-hirarki yang bertujuan untuk membagi data yang ada menjadi satu atau lebih kelompok. Metode *K-Means* digunakan untuk mengklasifikasikan daerah geografis berdasarkan karakteristik tertentu, seperti topografi, curah hujan, penggunaan lahan, dan faktor-faktor lain yang terkait dengan tingkat kerentanan terhadap bencana. Penggunaan kombinasi metode *K-Means* dan *Naive Bayes* dapat menjadi dasar yang solid untuk pengembangan sistem peringatan dini yang lebih efektif di masa yang akan datang. Penelitian ini memberikan kontribusi yang signifikan dalam upaya meningkatkan kesiapsiagaan dan *respons* terhadap bencana alam, sekaligus mendukung strategi pembangunan berkelanjutan di Kabupaten Labuhan Batu.

II. Landasan Teori Bencana

Bencana merupakan kejadian atau serangkaian kejadian yang menimbulkan ancaman dan gangguan terhadap kehidupan serta penghidupan masyarakat. Bencana dapat disebabkan oleh faktor alam, faktor non-alam, atau faktor manusia, yang mengakibatkan timbulnya kerugian jiwa, kerusakan lingkungan, kehilangan harta benda, dan dampak psikologis.

Terdapat tiga jenis bencana berdasarkan penyebabnya, yaitu sebagai berikut:

1. Bencana Geologis
Proses-proses geologi, baik yang bersifat endogenik maupun eksogenik, memiliki potensi

untuk menimbulkan risiko dan bahaya bagi kehidupan manusia. Bencana yang diakibatkan oleh proses-proses geologi ini dikenal sebagai bencana geologi. Bencana geologis terdiri dari:

- a. Gempa Bumi (*Earthquake*)
Gempa bumi adalah getaran yang berasal dari dalam bumi yang disebabkan oleh peristiwa yang terjadi di dalam perut bumi.
- b. Tsunami,
Tsunami adalah sebuah ombak yang terjadi setelah sebuah gempa bumi, gempa laut, gunung api meletus, atau hantaman meteor dilaut.
- c. Gunung berapi (*Volcano*),
Gunung berapi adalah elevasi tanah atau bukit yang terbentuk dari akumulasi semua material hasil erupsi yang melewati satu atau beberapa saluran (*Vulkanic Vent*) di seluruh permukaan bumi.

2. Bencana Meteorologis

Semua peristiwa bencana meteorologis saat ini termasuk fenomena alam yang dapat diprediksi secara cukup akurat setelah adanya sistem pemantauan yang terpadu melibatkan stasiun pemantau dan satelit cuaca. Bencana Meteorologis terdiri dari:

- a. Banjir (*Flood*)
- b. Gelombang Laut (*Wave*),
- c. Kebakaran Liar (*Wildfire*)
- d. Kekeringan (*Drought*)
- e. Topan (*Storm*)

3. Bencana

Antropogenik (*Anthropogenic*)
Antropogenik adalah istilah

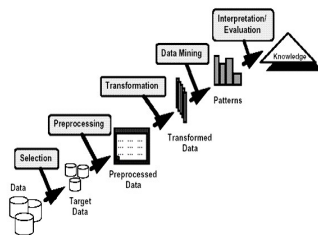
yang dapat diartikan secara simpel sebagai kegiatan manusia, baik yang dilakukan dengan sengaja maupun tidak sengaja, yang berkelanjutan dan berdampak negatif bagi masyarakat karena dapat memicu atau mempercepat terjadinya bencana.

Knowledge Discovery In Database

Knowledge Discovery In Database (KDD) merupakan metode untuk memperoleh pengetahuan dari database yang ada. Dalam struktur database, terdapat tabel-tabel yang memiliki keterkaitan atau relasi satu sama lain. Informasi yang ditemukan dalam proses ini dapat dijadikan dasar pengetahuan (*knowledge base*) untuk keperluan pengambilan keputusan. Proses KDD secara garis besar dapat dijelaskan sebagai berikut :

1. *Data Selection*
2. *Pre-processing / Cleaning*
3. *Transformation*
4. *Data Mining*
5. *Interpretation / Evaluation*

Proses *Knowledge Discovery in Database* (KDD) dapat digambarkan sebagai berikut :



Gambar 1. Proses Knowledge Discovery In Database

Sumber :

<https://dosbing.id/2023/04/21/proses-tahapan-data-mining/>

Data Mining

Data mining merupakan analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan cara berbeda dengan cara yang berbeda dengan sebelumnya, yang dapat dipahami dan bermanfaat bagi pemilik data.

Clustering

Clustering atau klasifikasi adalah metode yang digunakan untuk membagi rangkaian data menjadi beberapa *group* berdasarkan kesamaan-kesamaan yang telah ditentukan sebelumnya.

Algoritma K-Means

Salah satu algoritma *Clustering* adalah *K-Means* yang digunakan untuk mengelompokkan data menjadi beberapa kelompok dengan beberapa *Cluster*. Data-data dipilih menjadi beberapa kelompok dengan kriteria yang telah ditentukan lalu dikumpulkan menjadi satu dalam sebuah *Cluster*. Dimana setiap *Cluster* memiliki titik pusat yang disebut *Centroid*.

Berikut adalah tahapan-tahapan untuk melakukan optimasi menggunakan algoritma *K-Means*:

- a. Pilihlah jumlah *Cluster* (*k*) yang diinginkan pada dataset
- b. Tentukan titik pusat (*Centroid*) secara acak/random pada tahap awal
- c. Hitunglah jarak terdekat setiap data dengan *Centroid*. Untuk menghitung jarak terdekat dengan *Centroid* adalah *Euclidean distance* (*d*). Dapat menggunakan rumus dibawah ini:

$$d_e = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2}$$

Keterangan:

(x,y)=Koordinat objek

(s,t)=Koordinat *Centroid*

i=Banyaknya objek

- d. Hitung kembali pusat *Cluster* dengan keanggotaan *Cluster* yang sekarang. Pusat *Cluster* adalah rata-rata dari semua data dalam sebuah *Cluster*. Dapat dihitung menggunakan rumus:

$$V_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj}$$

V_{ij} = *Centroid* rata-rata

pada *Cluster* ke-i untuk

variabel ke-j

N_i = Jumlah

anggota *Cluster* ke-i

i, k= Indeks dari

Cluster

j = Indeks variabel

X_{kj} = Nilai data ke-k

variabel ke-j untuk *Cluster*

tersebut

Algoritma *Naïve Bayes*

Naïve Bayes merupakan metode probabilistic pengklasifikasian sederhana

berdasarkan Teorema Bayes dimana pengklasifikasian dilakukan melalui trainingset sejumlah data secara efisien. Teorema Bayes, yang diusulkan oleh ilmuwan Inggris Thomas Bayes, menggambarkan prediksi peluang di masa depan berdasarkan pengalaman di masa lalu, sehingga dikenal sebagai Teorema Bayes. Persamaan Teorema Bayes.

$$P(H | X) = \frac{P(X|H)P(H)}{P(X)}$$

Dimana :

X = data dengan kelas yang belum diketahui

H = hipotesis data X merupakan suatu kelas spesifik

$P(H|X)$ = probabilitas hipotesis H berdasar kondisi X (Probabilitas posterior)

$P(H)$ = probabilitas hipotesis H (Probabilitas prior)

$P(X|H)$ = Probabilitas X berdasarkan kondisi pada hipotesis H

$P(X)$ = Probabilitas X

RapidMiner

Rapidminer adalah koleksi dari algoritma *learning machine* yang digunakan untuk tugas-tugas data mining. Rapidminer berisi *tool* untuk data pre-processing, klasifikasi, regresi, clustering, rule association, dan memvisualisasikan data tersebut menjadi mudah untuk dapat dipahami.

III. Metode Penelitian

Metode penelitian yang digunakan adalah sebagai berikut :

1. Observasi (*Observation*)

Observasi dilakukan pengamatan pada BPBD Kabupaten Labuhanbatu melalui pengumpulan data dengan cara mengamati dan menganalisa data-data riwayat bencana yang terjadi.

2. Wawancara (*Interview*)

Wawancara yang dilakukan pada penelitian ini adalah dengan tanya jawab secara langsung kepada pegawai yang kompeten di BPBD Kabupaten untuk mendapatkan data-data umum mengenai jenis bencana yang tercatat.

3. Studi Pustaka (*Literature Review*)

Studi pustaka data dilakukan dengan pengambilan data-data

dari catatan kuliah, referensi buku, dan artikel jurnal ilmiah, yang mendukung penelitian.

IV. Hasil dan Pembahasan Analisis

Analisis adalah proses penguraian suatu hal menjadi bagian-bagian yang lebih kecil untuk memahami, mengevaluasi dan mendapatkan wawasan yang lebih mendalam mengenai aspek tertentu.

Pengumpulan Data

Dalam pengumpulan data dilakukan observasi yaitu pengamatan secara langsung di tempat penelitian. Dataset yang digunakan adalah dataset daerah bencana alam periode tahun 2022-2023 yang di dapat melalui pengumpulan data dari Badan Penanggulangan Bencana Daerah. Jumlah *record* dataset ini yaitu berjumlah 9 data.

Tabel 1. Data Bencana Kabupaten Labuhanbatu

id	Kecamatan	Kode Wilayah	JenisBencana	Jumlah	Satuan	Tahun
1	Bilah Barat	12.10.07	Banjir	1	Desa	2022
2	Bilah Barat	12.10.07	Puting Beliuung	1	Desa	2023
3	Bilah Barat	12.10.07	Kebakaran	3	Desa	2023
4	Bilah Barat	12.10.07	Tanah Longsor	1	Desa	2023
5	Bilah Hilir	12.10.08	Kebakaran	2	Desa	2023
6	Bilah Hilir	12.10.08	Korban Hanyut	1	Desa	2023
7	Bilah Hilir	12.10.08	Banjir	1	Desa	2023
8	Bilah Huhu	12.10.09	Banjir	1	Desa	2022
9	Bilah Huhu	12.10.09	Korban Hanyut	1	Desa	2023
10	Bilah Huhu	12.10.09	Kebakaran	4	Desa	2023
11	Bilah Huhu	12.10.09	Puting Beliuung	1	Desa	2023
12	Pangkatan	12.10.14	Banjir	1	Desa	2023
13	Panai Hilir	12.10.19	Korban Hanyut	1	Desa	2023
14	Panai Hilir	12.10.19	Kebakaran	2	Desa	2023
15	Panai Hulu	12.10.20	-	-	Desa	2023
16	Panai Tengah	12.10.18	Korban Hanyut	1	Desa	2023
17	Panai Tengah	12.10.18	Kebakaran	1	Desa	2023
18	Rantau Utara	12.10.01	Kebakaran	4	Desa	2023
19	Rantau Utara	12.10.01	Puting Beliuung	2	Desa	2023
20	Rantau Utara	12.10.01	Korban Hanyut	1	Desa	2023
21	Rantau Selatan	12.10.02	Puting Beliuung	2	Desa	2023
22	Rantau Selatan	12.10.02	Korban Hanyut	1	Desa	2023
23	Rantau Selatan	12.10.02	Kebakaran	1	Desa	2023
24	Rantau Selatan	12.10.02	Banjir	1	Desa	2023

Tabel Diatas merupakan data bencana daerah di Kabupaten Labuhanbatu yang memiliki 6atribut yaitu atribut kecamatan, Kode wilayah, jenis bencana, Jumlah, Satuan dan tahun.

Data Preprocessing

Pada tahap ini akan dilakukan preprocessing data dengan mengurangi atribut yang tidak digunakan dalam proses data mining. Terdapat 4 atribut yang akan digunakan yaitu Kode Kecamatan, Jenis Bencana, Jumlah dan Tahun.

Tabel 2. Data Preprocessing

Kecamatan	Jenis bencana	Jumlah	Tahun
Bilah Barat	Banjir	1	2022
Bilah Barat	Puting Beliuung	1	2023
Bilah Barat	Kebakaran	3	2023
Bilah Barat	Tanah Longsor	1	2023
Bilah Hilir	Kebakaran	2	2023
Bilah Hilir	Korban Hanyut	1	2023
Bilah Hilir	Banjir	1	2023
Bilah Hulu	Banjir	1	2022
Bilah Hulu	Korban Hanyut	1	2023
Bilah Hulu	Kebakaran	4	2023
Bilah Hulu	Puting Beliuung	1	2023
Pangkatan	Banjir	1	2023
Panai Hilir	Korban Hanyut	1	2023
Panai Hilir	Kebakaran	2	2023
Panai Hulu	-	-	2023
Panai Tengah	Korban Hanyut	1	2023
Panai Tengah	Kebakaran	1	2023

Tabel diatas menunjukkan data yang akan dipakai ke proses selanjutnya dengan mengurangi/mengeliminasi atribut atribut yang tidak sesuai hingga menghasilkan data yang sesuai.

Data Transformation

Pada tahap ini dilakukan transformasi data dengan bentuk yang sesuai pada proses data mining. Dari 24 *record* data ditransformasikan menjadi 9 data yang merupakan 9 Kecamatan dengan 1 kolom kecamatan dan 5 kolom kejadian bencana yaitu Banjir, puting beliuung, kebakaran, korban hanyut, tanah longsordan 1 kolom total bencana yang terjadi di setiap kecamatan yang di kalkulasikan pada periode 2022-2023.

Tabel 3. Data Transformation

No	Kecamatan	Banjir	Puting belung	Kebakaran	Korban Hanyut	Tanah Longsor
1	Bilah Barat	1	1	3	-	1
2	Bilah Hilir	1	-	2	1	-
3	Bilah Hulu	1	1	4	1	-
4	Pangkatan	1	-	-	-	-
5	Panai Hilir	-	-	2	1	-
6	Panai Hulu	-	-	-	-	-
7	Panai Tengah	-	-	1	1	-
8	Rantau Utara	-	2	4	1	-
9	Rantau Selatan	1	2	1	1	-

Algoritma K-Means

Pada tahap ini dilakukan penghitungan menggunakan algoritma *K-Means*. Setelah melakukan tahapan diatas, maka dapat diketahui bahwa :

- Jumlah Cluster : 3
- Jumlah Data : 9
- Jumlah Atribut : 6

1. Menentukan centroid awal secara acak.

$$C0 \text{ (Pangkatan)} = (1,0,0,0,0)$$

$$C1 \text{ (Bilah Hulu)} = (1,1,4,1,0)$$

$$C2 \text{ (Bilah Hilir)} = (1,0,2,1,0)$$

Pada *Cluster 0* bencana di Kabupaten Labuhanbatu yang menjadi jumlah bencana paling sedikit di ambil pada data ke 4 yaitu Pangkatan ,lalu untuk *Cluster1*dengan bencana jumlah Paling banyak adalah data ke 3 yaitu Bilah Hulu,dan untuk *Cluster* dengan jumlah bencana menengah adalah data ke 2 yaitu Bilah Hilir.

2. Menghitung centroid terdekat
Langkah selanjutnya ialah menghitung jarak setiap data dengan *Centroid* dengan persamaan *Euclidean Distance*. Pada tahap ini jarak terdekat antara data dengan *Cluster* akan menentukan suatu data masuk ke dalam *Cluster* mana. Berikut adalah hasil perhitungan jarak pada langkah 1

$$\text{Bilah Barat (C0)} = 3,31662$$

$$\sqrt{(1-1)^2 + (1-0)^2 + (3-0)^2 + (0-0)^2 + (1-0)^2}$$

$$\text{Bilah Barat (C1)} = 1,73205$$

$$\sqrt{(1-1)^2 + (1-1)^2 + (3-4)^2 + (0-1)^2 + (1-0)^2}$$

$$\text{Bilah Barat (C2)} = 2$$

$$\sqrt{(1-1)^2 + (1-0)^2 + (3-2)^2 + (0-1)^2 + (1-0)^2}$$

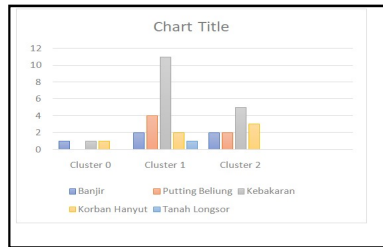
Tabel 4. Jarak Ke Centroid

Kecamatan	Jarak ke Centroid		
	Cluster 0	Cluster 1	Cluster 2
Bilah Barat	3,31662	1,73205	2,00000
Bilah Hilir	2,23606	2,23606	0
Bilah Hulu	4,24264	0	2,23606
Pangkatan	0	4,24264	2,23606
Panai Hilir	2,44948	2,44948	1,00000
Panai Hulu	1,00000	4,35889	2,44948
Panai Tengah	1,73205	3,31662	2,00000
Rantau Utara	4,69041	1,41421	3,00000
Rantau Selatan	2,44948	3,16227	2,23606

3. Melakukan Pengelompokan Berdasarkan Cluster
Melakukan pengelompokan berdasarkan *cluster*. Berikut adalah hasil pengelompokan *cluster* pada Iterasi 1:

Tabel 5. Hasil Pengelompokan Cluster

Kecamatan	Jarak terdekat
Bilah Barat	C1
Bilah Hilir	C2
Bilah Hulu	C1
Pangkatan	C0
Panai Hilir	C2
Panai Hulu	C0
Panai Tengah	C0
Rantau Utara	C1
Rantau Selatan	C2



Gambar 2. Grafik Hasil Pengelompokan Cluster

Hasil yang didapat dari grafik tersebut adalah *Cluster 0* sebagai bencana dalam jumlah paling sedikit yaitu 1 bencana banjir, 1 bencana kebakaran, 2 bencana korban hanyut. Untuk *Cluster 1* sebagai bencana dalam jumlah paling banyak yaitu dengan 2 bencana banjir, 4 bencana puting beliung, 11 bencana kebakaran, 2 bencana korban hanyut dan 1 bencana tanah longsor. Pada *Cluster 2* sebagai bencana menengah yaitu dengan 2 bencana banjir, 2 bencana puting beliung, 5 bencana kebakaran, 3 bencana korban hanyut dan 0 bencana tanah longsor.

Algoritma Naive Bayes

Data training dan *testing* digunakan untuk klasifikasi berjumlah 35 data uji, dimana data training selanjutnya akan diproses menggunakan naive bayes. Adapun kelas label data faktual ditunjukkan pada tabel 6.

Tabel 6. Data Latih Bencana Kabupaten Labuhanbatu

No	Jenis bencana	Tahun	Bencana	Total kerugian	Main
1	Bencana alam	2022	Banjir	--100Juta	Ringan
2	Bencana alam	2023	Banjir	--500Juta	Berat
3	Bencana alam	2023	Banjir	--100Juta	Ringan
4	Bencana alam	2023	Banjir	--100Juta	Ringan
5	Bencana alam	2023	Puting Beliung	--1M	Berat
6	Bencana alam	2023	Puting Beliung	--100Juta	Ringan
7	Bencana alam	2023	Puting Beliung	--120Juta	Ringan
8	Bencana alam	2023	Puting Beliung	--50Juta	Ringan
9	Bencana alam	2023	kebakaran	--240Juta	Berat
10	Bencana alam	2023	kebakaran	--250Juta	Berat
11	Bencana alam	2023	kebakaran	--200Juta	Berat
12	Bencana buatan	2023	Korban hanyut	--50Juta	Ringan
13	Bencana buatan	2023	Korban hanyut	--100Juta	Ringan
14	Bencana alam	2023	kebakaran	--150Juta	Berat
15	Bencana alam	2023	kebakaran	--170Juta	Berat
16	Bencana alam	2023	kebakaran	--50Juta	Ringan
17	Bencana alam	2023	kebakaran	--15Juta	Ringan
18	Bencana alam	2023	kebakaran	--130Juta	Ringan
19	Bencana alam	2023	kebakaran	--170Juta	Berat
20	Bencana alam	2023	kebakaran	--130Juta	Ringan
21	Bencana alam	2023	Tanah longsor	--130Juta	Ringan
22	Bencana alam	2023	Puting Beliung	--20Juta	Ringan
23	Bencana alam	2023	kebakaran	--20Juta	Ringan
24	Bencana alam	2023	kebakaran	--15Juta	Ringan
25	Bencana alam	2023	kebakaran	--500Juta	Berat
26	Bencana alam	2023	kebakaran	--700Juta	Berat
27	Bencana alam	2023	kebakaran	--350Juta	Berat

Tabel 7. Kelas Label Data Training

Kategori	Kriteria	Ringan	Berat
Jenis Bencana	Bencana Alam	16	13
	Bencana Buatan	4	2
Tahun	2022	2	0
	2023	18	15
Bencana	Banjir	4	1
	Puting beliung	5	1
	Kebakaran	6	12
	Korban Hanyut	6	0
	Tanah Longsor	1	0
Total Kerugian	≤ 500 jt	20	12
	≥ 500 jt ≤ 1M	0	3

Tabel diatas merupakan hasil transformasi dari data training yang diubah kedalam bentuk tabel kelas data training agar pengolahan data pada tahap perhitungan lebih mudah untuk dilakukan. Selanjutnya adalah tahapan-tahapan dalam perhitungan untuk klasifikasi kedalam metode Naive Bayes dengan menggunakan 2. Berikut merupakan langkah-langkah menghitung menggunakan algoritma *naive bayes*.

- a. Menghitung probabilitas class/label
 1. $P(Y=Ringan)=20/35$ "jumlah data Ringan dibagi dengan keseluruhan data" = 0,57
 2. $P(Y=Berat)=15/35$ "jumlah data Berat dibagi dengan keseluruhan data" = 0,42
- b. Menghitung probabilitas kategori
 1. Ringan

$$P(\text{Jenis Bencana} = \text{Bencana Alam} | \text{Ringan}) = \frac{16}{20} = 0,8$$

$$P(\text{Jenis Bencana} = \text{Bencana Buatan} | \text{Ringan}) = \frac{4}{20} = 0,2$$

$$P(\text{Tahun} = 2022 | \text{Ringan}) = \frac{2}{20} = 0,1$$

$$P(\text{Tahun} = 2023 | \text{Ringan}) = \frac{18}{20} = 0,9$$

$$P(\text{Bencana} = \text{Banjir} | \text{Ringan}) = 4/22 = 0,18$$

$$P(\text{Bencana} = \text{Puting Beliung} | \text{Ringan}) = 5/22 = 0,22$$

$$P(\text{Bencana} = \text{Kebakaran} | \text{Ringan}) = 6/22 = 0,27$$

$$P(\text{Bencana} = \text{Korban Hanyut} | \text{Ringan}) = 6/22 = 0,27$$

$$P(\text{Bencana} = \text{Tanah Longsor} | \text{Ringan}) = 1/22 = 0,04$$

$$P(\text{Total Kerugian} = \leq 500 \text{ jt} | \text{Ringan}) = 20/20 = 1$$

$$P(\text{Total Kerugian} = \geq 500 \text{ jt} \leq 1 \text{M Ringan}) = 0/20 = 0$$

2. Berat

$$P(\text{Jenis Bencana} = \text{Bencana Alam} | \text{Berat}) = 13/15 = 0,86$$

$$P(\text{Jenis Bencana} = \text{Bencana Buatan} | \text{Berat}) = 2/15 = 0,13$$

$$P(\text{Tahun} = 2022 | \text{Berat}) = 0/15 = 0$$

$$P(\text{Tahun} = 2023 | \text{Berat}) = 15/15 = 1$$

$$P(\text{Bencana} = \text{Banjir} | \text{Berat}) = 1/14 = 0,07$$

$$P(\text{Bencana} = \text{Puting Beliung} | \text{Berat}) = 1/14 = 0,07$$

$$P(\text{Bencana} = \text{Kebakaran} | \text{Berat}) = 12/14 = 0,85$$

$$P(\text{Bencana} = \text{Korban Hanyut} | \text{Berat}) = 0/14 = 0$$

$$P(\text{Bencana} = \text{Tanah Longsor} | \text{Berat}) = 0/14 = 0$$

$$P(\text{Total Kerugian} = \leq 500 \text{ jt} | \text{Berat}) = 12/15 = 0,8$$

$$P(\text{Total Kerugian} = \geq 500 \text{ jt} \leq 1 \text{M} | \text{Berat}) = 3/15 = 0,2$$

Tabel 8. Hasil Perhitungan Probabilitas Kategori

Kategori	Kriteria	Ringan	Berat
Jenis Bencana	Bencana Alam	0,8	0,86
	Bencana Buatan	0,2	0,13
Tahun	2022	0,1	0
	2023	0,9	1
Bencana	Banjir	0,18	0,07
	Puting beliung	0,22	0,07
	Kebakaran	0,27	0,85
	Korban Hanyut	0,27	0
	Tanah Longsor	0,04	0
Total Kerugian	$\leq 250 \text{ jt}$	1	0,8
	$\geq 250 \text{ jt} \leq 500 \text{ jt}$	0	0,2

c. Kalikan semua variabel Ringan, Sedang dan Berat

1. Likelihood of Ringan

$$P(\text{Keputusan} | \text{Ringan}) * P(\text{Jenis Bencana} = \text{Bencana Alam} | \text{Ringan}) * P(\text{Tahun} = 2022 | \text{Ringan}) * P(\text{Bencana} = \text{Banjir} | \text{Ringan}) * P(\text{Total Kerugian} = \leq 250 \text{ jt} | \text{Ringan})$$

$$= 0,57 * 0,8 * 0,1 * 0,18 * 1$$

$$= 0,008208$$

2. Likelihood of Berat

$$P(\text{Keputusan} | \text{Berat}) * P(\text{Jenis Bencana} = \text{Bencana Alam} | \text{Berat}) * P(\text{Tahun} = 2022 | \text{Berat}) * P(\text{Bencana} = \text{Banjir} | \text{Berat}) * P(\text{Total Kerugian} = \leq 250 \text{ jt} | \text{Berat})$$

$$= 0,42 * 0,86 * 0 * 0,07 * 0,8$$

$$= 0$$

d. Bandingkan hasil nilai kelas Ringan dan Berat. Dari hasil diatas, terlihat bahwa nilai probabilitas tertinggi ada pada kelas (P[Ringan]) dengan nilai 0,008208 disimpulkan bahwa status dengan berpotensi bencana ringan.

e. Bandingkan hasil nilai kelas Ringan dan Berat menggunakan perhitungan klasifikasi *Naive Bayes Classifier*. Hasil klasifikasi dengan *Naive Bayes Classifier* (NBC) dapat dilihat pada tabel 9.

Tabel 9. Perhitungan NBC Pada Data Testing

No	Ringan	Berat	Data Training	Data Uji	Prediksi
1.	0,008208	0	Ringan	Ringan	Sesuai
2.	0	0,01204	Berat	Berat	Sesuai
3.	0,1296	0,12384	Ringan	Ringan	Sesuai
4.	0,1296	0,04816	Ringan	Ringan	Sesuai
5.	0	0,01204	Berat	Berat	Sesuai
6.	0,1584	0,15136	Ringan	Ringan	Sesuai
7.	0,1584	0,15136	Ringan	Ringan	Sesuai
8.	0,1584	0,15136	Ringan	Ringan	Sesuai
9.	0,1944	0,5848	Berat	Berat	Sesuai
10.	0,1944	0,5848	Berat	Berat	Sesuai
11.	0,1944	0,5848	Berat	Berat	Sesuai
12.	0,0486	0	Ringan	Ringan	Sesuai
13.	0,0486	0	Ringan	Ringan	Sesuai
14.	0,1944	0,5848	Berat	Berat	Sesuai
15.	0,1944	0,5848	Berat	Berat	Sesuai
16.	0,1944	0,5848	Ringan	Berat	Tidak Sesuai
17.	0,1944	0,5848	Ringan	Berat	Tidak Sesuai
18.	0,1944	0,5848	Ringan	Berat	Tidak Sesuai
19.	0,1944	0,5848	Berat	Berat	Sesuai
20.	0,1944	0,5848	Ringan	Berat	Tidak Sesuai
21.	0,0288	0	Ringan	Ringan	Sesuai
22.	0,1584	0,15136	Ringan	Ringan	Sesuai
23.	0,1944	0,5848	Ringan	Berat	Tidak Sesuai
24.	0,1944	0,5848	Ringan	Berat	Tidak Sesuai
25.	0	0,1462	Berat	Berat	Sesuai
26.	0	0,1462	Berat	Berat	Sesuai
27.	0	0,1462	Berat	Berat	Sesuai
28.	0,1944	0,5848	Berat	Berat	Sesuai
29.	0,0486	0	Ringan	Ringan	Sesuai
30.	0,0054	0	Ringan	Ringan	Sesuai
31.	0,1296	0,04816	Ringan	Ringan	Sesuai
32.	0,1584	0,15136	Ringan	Ringan	Sesuai
33.	0	0,1462	Berat	Berat	Sesuai
34.	0,0486	0	Ringan	Ringan	Sesuai
35.	0,0486	0	Ringan	Ringan	Sesuai

Dari hasil tabel diatas dapat diketahui bahwa akurasi perhitungan kecocokan NBC dengan Data Faktual adalah sebesar 87,88%.

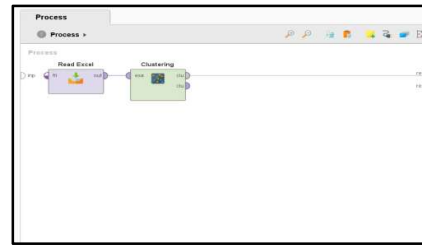
Implementasi Sistem Algoritma K-Means

Berikut merupakan data uji yang akan dimasukkan ke dalam aplikasi. Data uji dapat dilihat pada tabel 10.

Tabel 10. Data Uji K-Means

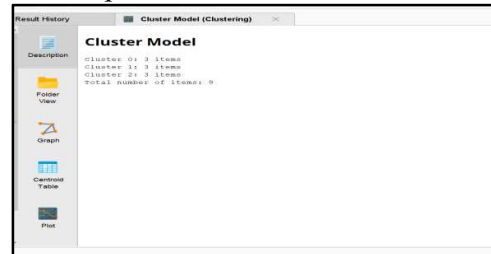
No	Kecamatan	Banjir	Puting beling	Kebakaran	Korban Hanyut	Tanah Longsor
1	Bilah Barat	1	1	3	-	1
2	Bilah Hilir	1	-	2	1	-
3	Bilah Hulu	1	1	4	1	-
4	Pangkatan	1	-	-	-	-
5	Pantai Hilir	-	-	2	1	-
6	Pantai Hulu	-	-	-	-	-
7	Pantai Tengah	-	-	1	1	-
8	Rantau Utara	-	2	4	1	-
9	Rantau Selatan	1	2	1	1	-

Secara keseluruhan proses yang dilakukan dengan menggunakan RapidMiner adalah sebagai berikut :



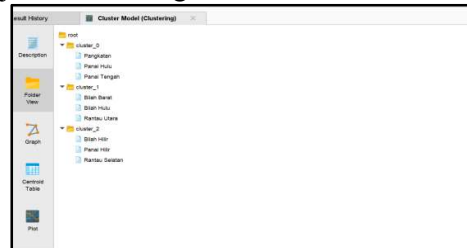
Gambar 3. Proses Rapidminer Menggunakan Algoritma K-Means

Berikut merupakan hasil cluster yang di dapat dari data uji yang sudah diolah oleh aplikasi.



Gambar 4. cluster model

Dari gambar diatas didapat hasil dengan 3 cluster yaitu cluster 0, cluster 1 dan cluster 2. Cluster 0 merupakan kelompok daerah yang paling sedikit bencana terjadi. Cluster 1 merupakan kelompok daerah yang paling banyak terjadi bencana dan menjadi daerah rawan bencana. Cluster 2 merupakan kelompok daerah dengan bencana jumlah menengah.



Gambar 5. Cluster Daerah Bencana

Dari gambar diatas didapatkan hasil beberapa daerah yang sudah dikelompokkan berdasarkan cluster nya. Cluster 0 dengan jumlah bencana paling sedikit terdapat di daerah pangkatan,

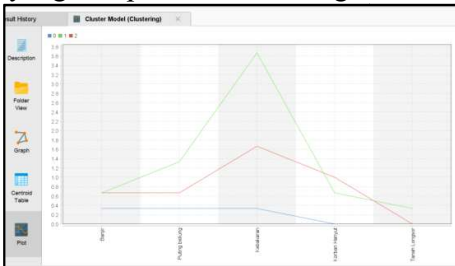
panai hulu, dan panai tengah. Cluster 1 dengan jumlah bencana paling banyak terdapat di daerah bilah barat, bilah hulu dan rantau utara. Cluster 2 dengan jumlah bencana mengengah terdapat di daerah bilah hilir, panai hilir dan rantau selatan.

Attributes	cluster_0	cluster_1	cluster_2
Bayu	0.333	0.667	0.667
Pung buking	0.333	1.333	0.667
Kebakaran	0.333	0.667	1.667
Kudam hanyut	0	0.667	1
Tanah longsor	0	0.333	0

Gambar 6. Bencana Dengan Cluster Daerah Bencana

Dari gambar diatas diketahui nilai rata – rata terjadi nya bencana pada cluster daerah masing – masing sesuai dengan potensi bencana yang terjadi. Dari hasil diatas dapat dilihat dan dinilai bahwa bencana yang paling sering terjadi adalah kebakaran di daerah cluster 1 dengan nilai rata – rata sebesar 3.667.

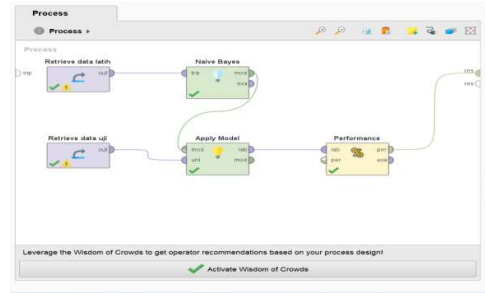
Berikut merupakan diagram yang didapatkan dari hasil gambar 7.



Gambar 7. Diagram Cluster Daerah Bencana

Algorithm Naïve Bayes

Secara keseluruhan proses yang dilakukan dengan menggunakan RapidMiner adalah sebagai berikut



Gambar 8. Proses Rapidminer Menggunakan Algoritma Naïve Bayes

Berikut ini merupakan hasil akurasi yang di dapatkan dari kedua data yang sudah di uji

Criterion	Table View	Plot View
accuracy	accuracy: 87.88%	
precision		
recall		
AUC (optimistic)	true Ringan	true Berat
AUC	10	0
AUC (pessimistic)	prec: Ringan	class precision
	4	19
	class recall	
	71.43%	100.00%

Gambar 9. Hasil Akurasi Data Training Dan Data Uji

Dari gambar diatas didapatkan hasil akurasi dari kedua yang sudah diuji data adalah sebesar 87.88%. Dengan deskripsi seperti gambar di bawah ini dan dipatkan hasil dengan class precision sebesar 82.61% dan class recall sebesar 100% dengan prediksi bencana dengan potensi berat.

```

PerformanceVector
PerformanceVector:
accuracy: 87.88%
ConfusionMatrix:
True: Ringan Berat
Ringan: 10 0
Berat: 4 19
precision: 82.61% (positive class: Berat)
ConfusionMatrix:
True: Ringan Berat
Ringan: 10 0
Berat: 4 19
recall: 100.00% (positive class: Berat)
ConfusionMatrix:
True: Ringan Berat
Ringan: 10 0
Berat: 4 19
AUC (optimistic): 0.980 (positive class: Berat)
AUC: 0.980 (positive class: Berat)
AUC (pessimistic): 0.980 (positive class: Berat)
    
```

Gambar 10. Deskripsi Naïve bayes

V. Kesimpulan dan Saran

Kesimpulan

Dari hasil pembahasan diatas maka didapatkan kesimpulan sebagai berikut:

1. Pada Algoritma *K-Means* didapatkan hasil *Cluster 0* dengan jumlah bencana paling sedikit terdapat di daerah pangkatan, panai hulu, dan panai tengah. *Cluster 1* dengan jumlah bencana paling banyak terdapat di daerah bilah barat, bilah hulu dan rantau utara. *Cluster 2* dengan jumlah bencana mengengah terdapat di daerah bilah hilir, panai hilir dan rantau selatan dan bencana yang paling sering terjadi adalah kebakaran di daerah cluster 1 dengan nilai rata – rata sebesar 3.667.
2. Pada algoritma *Naïve Bayes* didapatkan hasil didapatkan hasil akurasi dari kedua yang sudah diuji data adalah sebesar 87.88%. Dengan deskripsi seperti gambar di bawah ini dan dipatkan hasil dengan class precision sebesar 82.61% dan class recall sebesar 100% dengan prediksi bencana dengan potensi berat
3. Dari hasil kedua algoritma di atas didapatkan perbedaan antar kedua algoritma tersebut yaitu algortima *K-Means* digunakan untuk mengelompokkan bencana sesuai dengan cluster nya masing – masing. Algoritma *Naïve Bayes* dapat dilakukan untuk memprediksi bencana yang akan terjadi di masa depan dengan melihat kejadian masa lampau dengan memprediksi bencana yang terjadi dengan

potensi berat kerugian yang besar

Saran

Dengan adanya penelitian ini diharapkan kepada para masyarakat Kabupaten Labuhanbatu untuk lebih memperhatikan lingkungan sekitar dengan memperlihatkan bencana yang sering terjadi dengan potensi besar di beberapa daerah yang sudah diprediksi menjadi daerah rawan bencana. Kepada pihak berwajib dalam menanggulangi bencana untuk lebih memperhatikan daerah-daerah rawan bencana dengan memberikan beberapa edukasi untuk pencegahan maupun penanggulangan terhadap bencana. Semoga penelitian ini boleh lebih bermanfaat bagi penulis untuk lebih memahami mengenai data mining dan penerapan nya di dalam sebuah data uji, kepada Universitas maupun masyarakat di Kabupaten Labuhanbatu.

VI. Daftar Pustaka

- A. Adiyanto and Y. Arie Wijaya, “Penerapan Algoritma *K-Means* Pada Pengelompokan Data Set Bahan Pangan Indonesia Tahun 2022-2023,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 2, pp. 1344–1350, 2023, doi: 10.36040/jati.v7i2.6849.
- A. H. Kadarusno, Maryana, and A. Husein, “Konsep Dasar Manajemen Bencana,” e – J. Ris. Manaj., p. 146, 2022.
- D. A. Wiloso and S. Vienastra, “Mitigasi Bencana Tsunami di SDN Tirtohargo Dusun Baros, Desa Tirtohargo, Kecamatan Kretek, Kabupaten Bantul, Provinsi Daerah Istimewa

- Yogyakarta,” *J. Gaung Inform.*, vol. 11, no. 1, pp. 55–66, 2018.
- D. P. Utomo and M. Mesran, “Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung,” *J. Media Inform. Budidarma*, vol. 4, no. 2, p. 437, 2020, doi: 10.30865/mib.v4i2.2080.
- E. T. Ena Tasia and M. Afdal, “Perbandingan Algoritma K-Means Dan K-Medoids Untuk Clustering Daerah Rawan Banjir Di Kabupaten Rokan Hilir,” *Indones. J. Inform. Res. Softw. Eng.*, vol. 3, no. 1, pp. 65–73, 2023, doi: 10.57152/ijirse.v3i1.523.
- G. Ramadhan et al., “Penerapan Data Mining Menggunakan Algoritma C4.5 Dalam Mengukur Tingkat Kepuasan Pasien BPJS,” *Pros. Semin. Nas. Ris. Dan Inf. Sci.*, vol. 2, pp. 376–385, 2020.
- H. Apriyani and K. Kurniati, “Perbandingan Metode Naïve Bayes Dan Support Vector Machine Dalam Klasifikasi Penyakit Diabetes Melitus,” *J. Inf. Technol. Ampera*, vol. 1, no. 3, pp. 133–143, 2020, doi: 10.51519/journalita.volume1.issue3.year2020.page133-143.
- I. Fahrur Rozi, A. Taufika Firdausi, and K. Islamiyah, “Analisis Sentimen Pada Twitter Mengenai Pasca Bencana Menggunakan Metode Naïve Bayes Dengan Fitur N-Gram,” *J. Inform. Polinema*, vol. 6, no. 2, pp. 33–39, 2020, doi: 10.33795/jip.v6i2.316.
- N. Dwitri, J. A. Tampubolon, S. Prayoga, F. I. R.H Zer, and D. Hartama, “Penerapan Algoritma K-Means Dalam Menentukan Tingkat Penyebaran Pandemi Covid-19 Di Indonesia,” *J. Teknol. Inf.*, vol. 4, no. 1, pp. 128–132, 2020, doi: 10.36294/jurti.v4i1.1266.
- N. Hidayat and E. W. Santoso, “Gempa Bumi Dan Mekanismenya,” *Alami: Jurnal Teknologi Reduksi Resiko Bencana*, vol. 2, no. 3, p. 50, 1997.
- N. Nuryadi and S. Agustiarini, “Analisis Rawan Kekeringan Lahan Padi Kabupaten Banyuwangi Jawa Timur,” *J. Meteorol. Klimatologi dan Geofis.*, vol. 5, no. 2, pp. 29–36, 2019, doi: 10.36754/jmkg.v5i2.56.
- N. Syafri, Edi; Endrizal, “Manajemen Mitigasi Bencana,” *J. Chem. Inf. Model.*, vol. 53, no. 9, pp. 1689–1699, 2020.
- O. H. A. R. & P. H. Herni Tandi Sarapang, “Analisis Kerentanan Bencana Tsunami Di Kota Palu,” *J. Spasial*, vol. 6, no. 2, pp. 432–439, 2019, [Online]. Available: <http://repository.ub.ac.id/id/eprint/8056>
- T. Akhir and P. Studi, “BAB II TINJAUAN UMUM MUSEUM dan VULKANOLOGI,” pp. 1–25, 2009.
- Y. Mardi, “Data Mining: Klasifikasi Menggunakan Algoritma C4.5,” *Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2017, doi: 10.22202/ei.2016.v2i2.1465.