

ANALISIS PROBABILITAS KEBENCANAAN DAERAH MENGGUNAKAN ALGORITMA K-MEANS DAN LINEAR REGRESSION**Muhammad Ridwan Syamsuddin¹, Wirdayanti² Dwi Shinta Angreni³ Yusuf Anshori⁴**muhammad.ridwan001.mr@gmail.com¹, wirda_arbie@untad.ac.id², ds.angreni@untad.ac.id³,yusuf.anshori@untad.ac.id⁴

Universitas Tadulako Sulawesi Tengah

Informasi Artikel

Diterima : 21 Apr 2024

Direview : 9 Mei 2024

Disetujui : 15 Jun 2024

Kata Kunci*BPBD, Data Mining, K-Means, Linear Regression, Python, Weka***Abstrak**

Bencana alam, gempa bumi, angin puting beliung, kekeringan, gelombang tinggi, tanah longsor dan banjir terjadi setiap tahunnya di Sulawesi tengah. Bencana alam ini dapat menimbulkan kerugian yang besar baik materiil dan juga korban jiwa. Upaya mitigasi dan adaptasi diperlukan untuk mengurangi dampak bencana alam. Salah satu cara mitigasi dan adaptasi terhadap bencana alam adalah penggunaan teknologi teknik data mining. Metode ini menggunakan CRISP-DM yang mencakup pengumpulan data, data cluster, data preprocessing, data prediksi dan menguji model. Penelitian ini menggunakan algoritma *K-Means* untuk mengelompokkan data bencana dan algoritma *Linear Regression* untuk menentukan probabilitas terjadinya bencana. Hasil perhitungan tersebut didapatkan wilayah yang mengalami probabilitas peningkatan pada bencana banjir yaitu di daerah Kabupaten Buol, probabilitas peningkatan bencana puting beliung di daerah Kabupaten Toli-toli, dan untuk probabilitas peningkatan bencana gelombang tinggi wilayah Kabupaten Banggai Kepulauan. Hasil evaluasi pengujian model linear regression bahwa nilai RMSE 0.4277 memiliki satuan yang sama dengan nilai sebenarnya sehingga lebih mudah untuk diinterpretasikan. Oleh karena itu, penelitian dapat membantu memprediksi probabilitas kejadian bencana di masa depan dan juga mengembangkan strategi penanggulangan bencana yang lebih efektif.

Keywords*BPBD, Data Mining, K-Means, Linear Regression, Python, Weka***Abstract**

Natural disasters, earthquakes, tornadoes, droughts, high waves, landslides and floods occur every year in Central Sulawesi. This natural disaster can cause large losses both material and fatalities. Mitigation and adaptation efforts are needed to reduce the impact of natural disasters. One way to mitigate and adapt to natural disasters is the use of data mining engineering technology. This method uses CRISP-DM which includes data collection, cluster data, data preprocessing, prediction data and testing the model. This research uses the K-Means algorithm to group disaster data and the Linear Regression algorithm to determine the probability of a disaster occurring. The results of these calculations show that the areas experiencing an increased probability of flood disasters are in the Buol Regency area, the probability of an increased tornado disaster in the Toli-toli Regency area, and the probability of an increase in high wave disasters in the Banggai Islands Regency area. The evaluation results of the linear regression model test show that the RMSE value of 0.4277 has the same units as the actual value so it is easier to interpret. Therefore, research can help predict the probability of future disaster events and also develop more effective disaster management strategies.

A. Pendahuluan

Bencana alam, gempa bumi, angin puting beliung, kekeringan, gelombang tinggi, tanah longsor dan banjir terjadi setiap tahunnya di Sulawesi tengah[1]. Bencana alam ini dapat menimbulkan kerugian yang besar baik materil dan juga korban jiwa Upaya mitigasi dan adaptasi diperlukan untuk mengurangi dampak bencana alam. Upaya mitigasi bertujuan untuk mencegah terjadinya bencana alam, sedangkan upaya adaptasi bertujuan untuk mengurangi dampak bencana alam ketika terjadi[2]. Salah satu cara mitigasi dan adaptasi terhadap bencana alam adalah dengan teknologi yang menggunakan teknik data mining[3].

Penelitian ini menganalisis data bencana dari Badan Penanggulangan Bencana Daerah (BPBD) Sulawesi Tengah dengan menggunakan teknik data mining, yang melibatkan penggalian informasi dari kumpulan data yang besar dan kompleks. Teknik ini dapat digunakan untuk memahami karakteristik dan pola bencana alam[4].

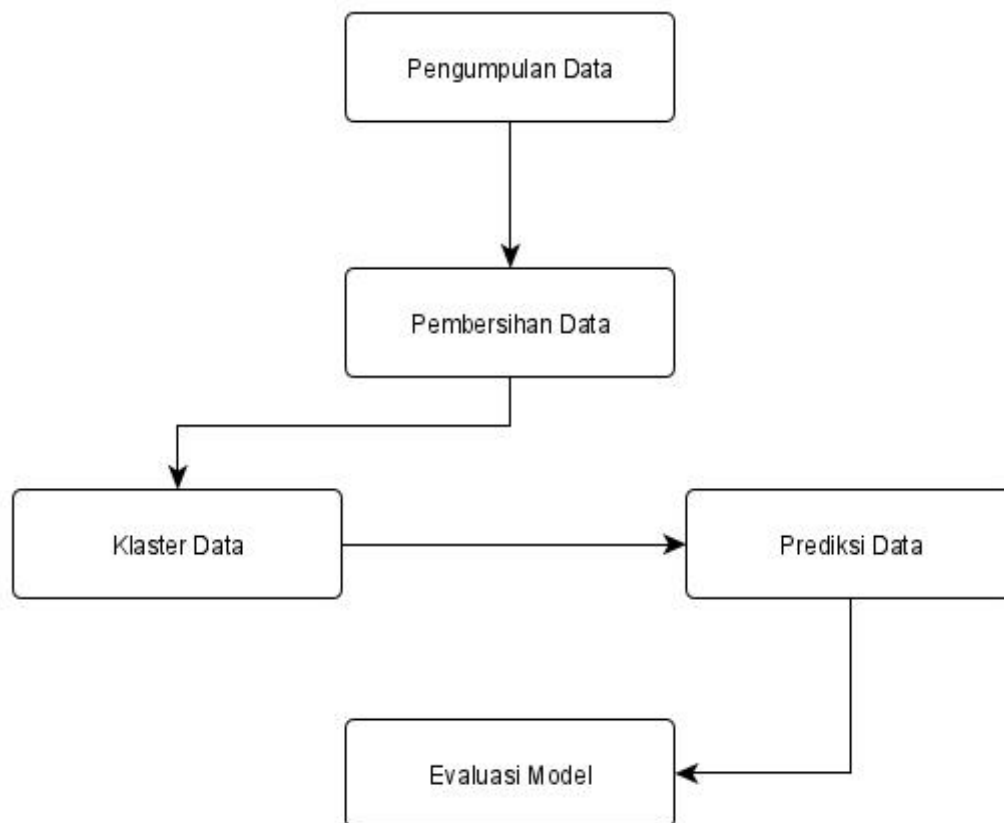
Meskipun BPBD memiliki kumpulan data yang luas tentang bencana, seringkali sulit untuk mengekstrak informasi yang bermanfaat secara efektif. Tantangan utama yang dihadapi adalah kompleksitas dan volume besar data, serta kebutuhan untuk mengidentifikasi pola dan tren yang dapat mendukung pengambilan keputusan. Akibatnya, diperlukan metode analisis yang dapat menangani data BPBD skala besar dengan lebih efisien. Tujuan penelitian ini adalah untuk menerapkan teknik data mining, khususnya algoritma clustering dan linear regresi K-means untuk membuat prediksi pada analisis data bencana BPBD[5]. Algoritma K-means membantu mengelompokkan data bencana ke dalam cluster-cluster dengan karakteristik yang sebanding[6]. Selanjutnya, gunakan Linear Regression untuk memahami hubungan antara variabel tertentu dan dampak atau kerugian bencana[7].

Penelitian ini juga merujuk pada beberapa studi sebelumnya tentang subjek yang sama misalnya, penelitian penggunaan data mining untuk memprediksi bencana alam menggunakan teknik K-means dan algoritma regresi linier, Penelitian ini berhasil mengumpulkan data bencana untuk setiap provinsi di Indonesia yang dimiliki oleh BNPB. Selain itu, algoritma K-Means dan algoritma regresi linier digunakan untuk memproyeksikan data bencana lima tahun ke depan. Algoritma K-Means dimaksudkan untuk mengelompokkan data bencana yang berkaitan dengan negara dari tahun 2005 hingga 2015[8]. Pada penelitian berikutnya analisis kemungkinan bencana alam dengan data mining menggunakan K-Means dan Regresi Linier. Penelitian ini bertujuan untuk menentukan kemungkinan bencana yang sering terjadi di beberapa wilayah di Indonesia[9]. Penelitian yang selanjutnya yaitu menerapkan pengelompokan K-Means untuk mengelompokkan wilayah rawan kebakaran menggunakan model CRISP-DM. Hal ini dimaksudkan untuk menciptakan kelompok wilayah yang mempunyai potensi kebakaran tinggi. Ini akan membantu Anda mewaspadaai segala bahaya yang mungkin terjadi di daerah tersebut[10]. Penelitian berikutnya yaitu Berdasarkan data penjualan air di PAM Kerta Raharja, algoritma K-means digunakan untuk mengklasifikasikan pelanggan berdasarkan kubikasi air yang dijual oleh Weka. Tujuannya adalah untuk memberikan informasi tentang kelompok pengguna air[11].

B. Metode Penelitian

a. Metode

Metode penelitian ini adalah menggunakan metode data minig dari CRISP-DM, Tahap awal dalam Analisa data dilakukan dengan 4 cara, yaitu pengumpulan data. Berupa data bencana alam 12 daerah Kabupaten dan 1 wilayah Kota di daerah Sulawesi Tengah dari tahun 2020 hingga tahun 2023 yang diperoleh dari BPBD Sulawesi Tengah, langkah kedua adalah Preprocessing data langkah ini melakukan transformasi data dari bentuk nominal diubah ke dalam bentuk numeric atau angka. Hal ini bertujuan untuk membersihkan data dan memudahkan proses clustering data. Langkah ketiga dalam analisis data adalah mengelompokkan data menggunakan K-means . Tujuan K-means adalah untuk mengelompokkan data. Mengolah seluruh data tersebut untuk mendapatkan hasil prediksi data bencana yang terjadi ditahun akan datang[12]. Prediksi bencana dilakukan dengan menggunakan algoritma LinearRegression. Python merupakan bahasa pemrograman tingkat tinggi yang mudah dibaca dan ditulis dan digunakan untuk melakukan analisis data. Python juga dapat digunakan untuk mengembangkan model machine learning, seperti klasifikasi, regresi, dan clustering untuk menjalankan kode Python. Penelitian ini menggunakan google colaboratory teks editor. Colaboratory adalah layanan komputasi cloud gratis yang memungkinkan Anda menjalankan kode Python secara langsung di browser. Tahapan keseluruhan penelitian ini diilustrasikan pada gambar 1



Gambar 1. Tahapan Alur Metode Penelitian

b. Dataset

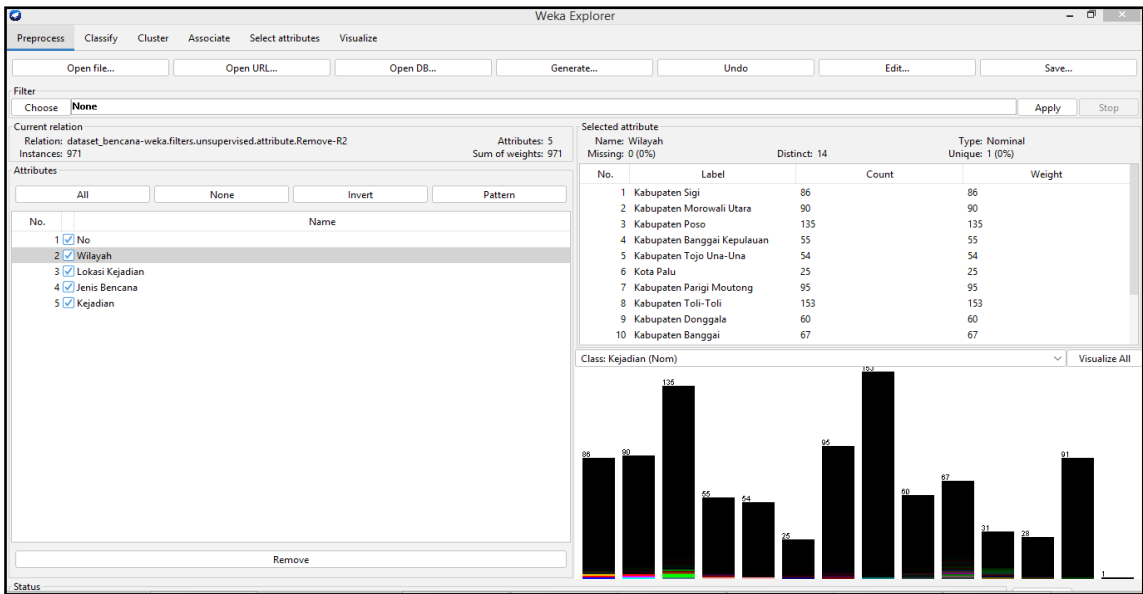
Data diperoleh dari BPBD dengan menghitung angka bencana alam dari tahun 2020 hingga 2023 di berbagai daerah di Sulawesi Tengah. Data ini mencakup jenis bencana alam rata-rata yang terjadi setiap tahun. Pada daerah diantaranya Kabupaten Banggai, Banggai Kepulauan, Banggai Laut, Toli-toli, Parigi Moutong, Morowali, Morowali Utara, Donggala, Sigi, Poso, Tojo Una-Una, Buol, dan Kota Palu. terdapat pada tabel 1 memiliki 971 data bencana.

Tabel 1. Dataset Kejadian

No	Kode	Wilayah	Lokasi Kejadian	Jenis Bencana
1	IB01528	Kabupaten Sigi	Omu	Banjir
2	IB01527	Kabupaten Sigi	Sambo	Banjir
3	IB01526	Kabupaten Morut	Tingkeao	Karhutla
4	IB01523	Kabupaten Poso	Bariri	Putting Beliang
--	-----		-----	---
970	IB00657	Kabupaten Buol	Paleleh	Tanah longsor
971	IB00233	Kabupaten Parimo	Mepanga	Banjir

c. Preprocessing Data

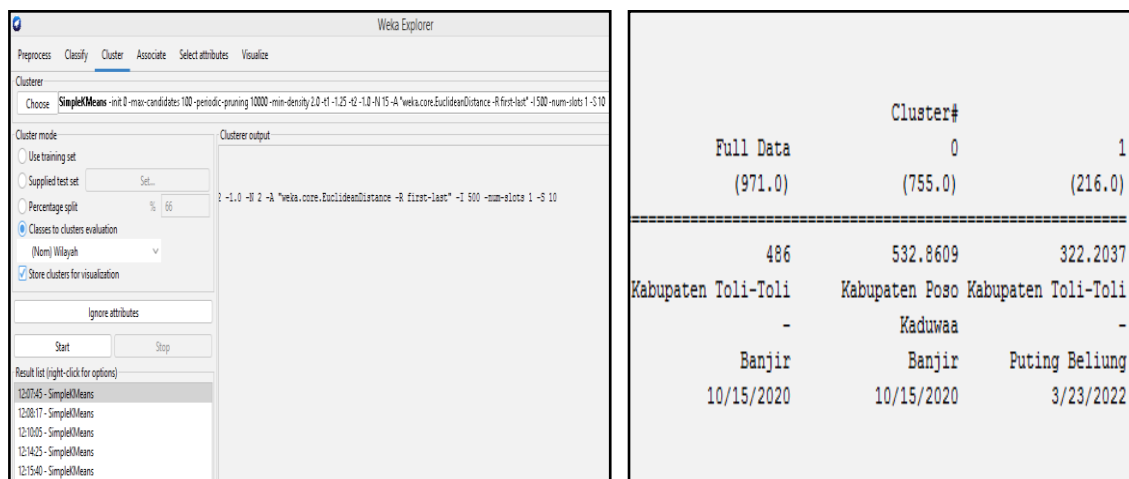
Pada tahap preprocessing (pembersihan) dilakukan pendataan bencana di 13 kabupaten di Provinsi Sulawesi Tengah. Dataset bencana alam dapat diimpor ke dalam aplikasi Weka sebelum melakukan preprocessing. dalam bentuk file excel csv. Setelah berhasil diimpor, lalu dapat menjalankan proses cleaning masing-masing atribut yaitu bertujuan untuk mengisi nilai value yang kosong atau hilang serta mentransformasi atribut menjadi nominal dan juga menghapus atribut yang tidak digunakan Setelah file data bencana ditampilkan di aplikasi Weka, dilakukan pra-pemrosesan data menggunakan fungsi penggantian nilai yang hilang pada alat Weka untuk menghapus atribut yang tidak terpakai. Selain melakukan preprocessing data, tujuan proses ini adalah untuk membersihkan kumpulan data yang digunakan untuk proses clustering data. Proses ini juga melakukan transformasi data dari numerik ke nominal[13]. Tahapan preprocessing ini terdapat pada gambar 2.



Gambar 2. Tampilan data preprocessing pada weka

d. Clustering Data

Analisis data cluster ialah algoritma membagi data menjadi beberapa kelompok yang berdasarkan data mempunyai kesamaan, Clustering yang digunakan adalah algoritma K-means yang bertujuan untuk menetapkan cluster data yang akan digunakan untuk memprediksi data bencana nantinya[14]. hasil clustering K-means dan dibuat menjadi 3 cluster untuk mengelompokan wilayah yang memiliki bencana yang sama, Cluster pertama mengelompokan untuk daerah yang terdampak bencana banjir dan cluster kedua mengelompokan daerah bencana puting beliung dan cluster 3 sebagai tambahan mengelompokan daerah bencana gelombang tinggi. Dapat dilihat pada gambar tersebut wilayah cluster pertama kabupaten Poso memiliki bencana Banjir yang tinggi dan cluster kedua kabupaten Toli-toli memiliki bencana Puting Beliung yang paling tinggi serta cluster ketiga kabupaten Banggai Kepulauan memiliki bencana Gelombang Tinggi yang paling tinggi. Tahapan cluster data dilihat pada gambar 3.



Gambar 3. Hasil data cluster pada weka

e. Prediksi Data

Prediksi data ini dilakukan dengan menggunakan algoritma Linear Regression pada tools Google Colaboratory menggunakan sintaks python untuk menentukan probabilitas terjadinya bencana Puting beliung, Banjir, dan Gelombang Tinggi, setelah mendapatkan cluster pada clustering menggunakan K-means didapatkan 3 bencana yang kemungkinan akan terjadi. lalu dibuatkan 3 atribut untuk mengelompokan cluster berdasarkan jenis bencana yang sering terjadi. terdapat pada tabel 2.

Tabel 2. Dataset Setelah Cluster

No	Wilayah	Jumlah Kejadian	Jenis_bencana Terjadi	bencana0	bencana1	bencana2
1	Kabupaten Sigi	86	Banjir	1	0	0
2	Kabupaten Morowali utara	90	Banjir	1	0	0
3	Kabupaten Poso	135	Puting Beliung	0	1	0
4	Kabupaten Banggai Kepulauan	55	Gelombang	0	0	1

	Bangai kepulauan		Tinggi			
5	Kabupaten Tojo Una-una	54	Puting Beling	0	0	0
6	Kota Palu Kabupaten.	25	Banjir	1	0	0
7	Parigi Moutong	96	Banjir	1	0	0
8	Kabupaten Toli-toli	153	Puting Beliung	0	1	0
9	Kabupaten Donggala	60	Banjir	1	0	0
10	Kabupaten Banggai	67	Banjir	1	0	0
11	Kabupaten Morowali	31	Puting Beliung	0	1	0
12	Kabupaten Banggai Laut	28	Banjir	1	0	0
13	Kabupaten Buol	91	Banjir	1	0	0

Setelah melakukan *cluster* dan mendapatkan data, selanjutnya melakukan prediksi untuk mendapatkan *Probabilitas* pada data bencana ke *goole coloboratory* dengan cara menghitung semua nilai total kejadian bencana dengan menjumlahkan semua nilai atribut jumlah kejadian pada dataset bencana, selanjutnya untuk menghitung probabilitas dengan membagi nilai jumlah kejadian untuk setiap jenis bencana dengan total kejadian bencana terdapat pada gambar 4.

```

[22] import pandas as pd
import numpy as np
from sklearn.linear_model import LinearRegression
import matplotlib.pyplot as plt

[45] bencana_count = pd.DataFrame(columns=['jenis_bencana', 'jumlah_kejadian'])

for index, row in data.iterrows():
    if row['jenis_bencana_terjadi'] == 'Banjir':
        bencana_count.loc[len(bencana_count)] = ['Banjir', row['jumlah_kejadian']]
    elif row['jenis_bencana_terjadi'] == 'Puting Beliung':
        bencana_count.loc[len(bencana_count)] = ['Puting Beliung', row['jumlah_kejadian']]
    elif row['jenis_bencana_terjadi'] == 'Tanah Longsor':
        bencana_count.loc[len(bencana_count)] = ['Tanah Longsor', row['jumlah_kejadian']]

[46] total_kejadian = data['jumlah_kejadian'].sum()

bencana_prob = bencana_count.apply(lambda x: x['jumlah_kejadian'] / total_kejadian, axis=1)
bencana_prob.name = 'probabilitas'

bencana_count = pd.concat([bencana_count, bencana_prob], axis=1)

print(bencana_count)

```

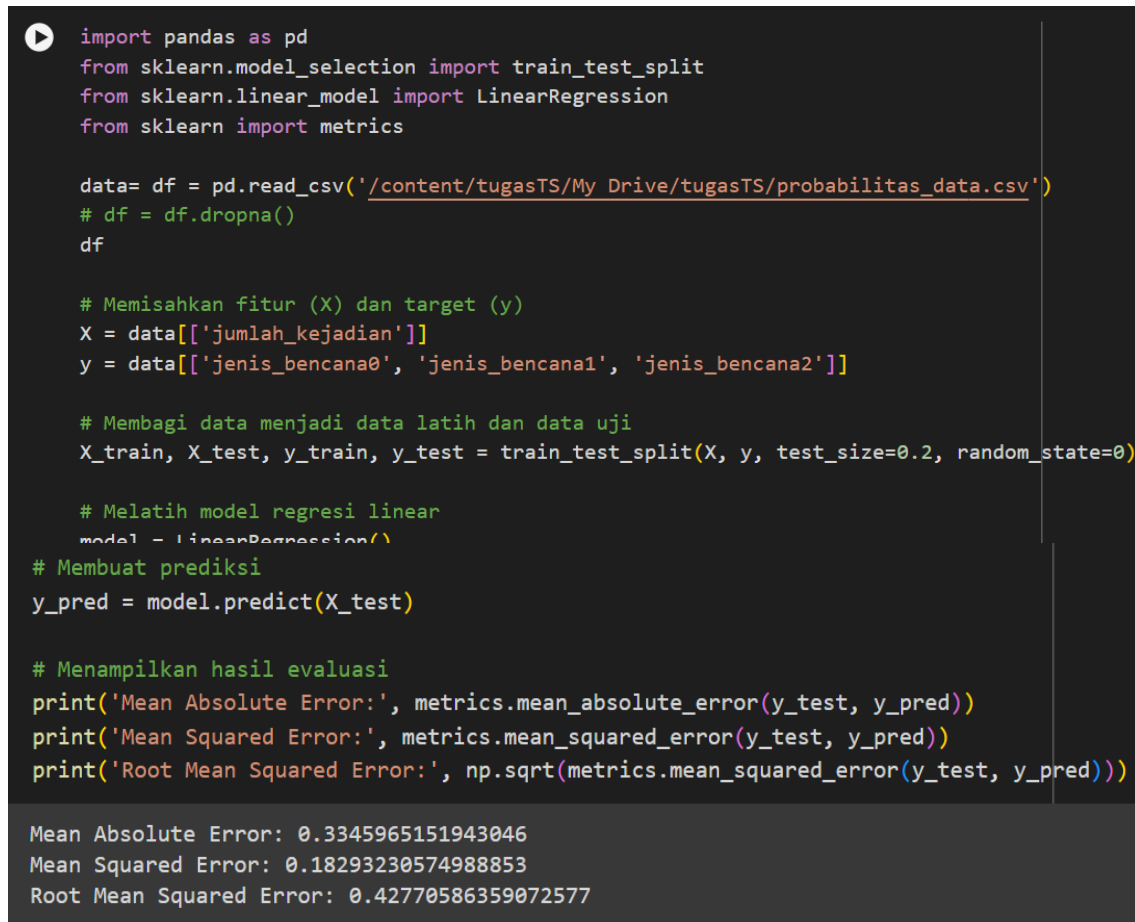
	jenis_bencana	jumlah_kejadian	probabilitas
0	Banjir	86	0.088568
1	Banjir	90	0.092688
2	Puting Beliung	135	0.139032
3	Puting Beliung	54	0.055613
4	Banjir	25	0.025747
5	Banjir	96	0.098867
6	Puting Beliung	153	0.157570
7	Banjir	60	0.061792
8	Banjir	67	0.069001
9	Puting Beliung	31	0.031926
10	Banjir	28	0.028836
11	Banjir	91	0.093718

Gambar 4. Tampilan Hasil Prediksi Data

f. Evaluasi Mode

Evaluasi kinerja model regresi linier untuk menentukan apakah model prediksi cukup akurat untuk digunakan sesuai dengan tujuan penelitian[15]. Nilai rata-rata kesalahan absolut antara nilai prediksi dan nilai aktual ditunjukkan oleh nilai MAE sebesar 0.3345, yang menunjukkan bahwa semakin kecil nilai MAE, semakin akurat model prediksinya[16]. Nilai mean squared error (MSE) adalah

0,1829, yang menunjukkan seberapa akurat model prediksinya. Nilai MSE yang lebih rendah menunjukkan seberapa akurat nilai prediksi dibandingkan dengan nilai sebenarnya.. Nilai RMSE sebesar 0.4277. Nilai RMSE (Root Mean Squared Error) memiliki satuan yang sama dengan nilai sebenarnya sehingga lebih mudah untuk diinterpretasikan[17] dilihat pada gambar 5.



```

import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn import metrics

data= df = pd.read_csv('/content/tugasTS/My Drive/tugasTS/probabilitas_data.csv')
# df = df.dropna()
df

# Memisahkan fitur (X) dan target (y)
X = data[['jumlah_kejadian']]
y = data[['jenis_bencana0', 'jenis_bencana1', 'jenis_bencana2']]

# Membagi data menjadi data latih dan data uji
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=0)

# Melatih model regresi linear
model = LinearRegression()

# Membuat prediksi
y_pred = model.predict(X_test)

# Menampilkan hasil evaluasi
print('Mean Absolute Error:', metrics.mean_absolute_error(y_test, y_pred))
print('Mean Squared Error:', metrics.mean_squared_error(y_test, y_pred))
print('Root Mean Squared Error:', np.sqrt(metrics.mean_squared_error(y_test, y_pred)))

Mean Absolute Error: 0.3345965151943046
Mean Squared Error: 0.18293230574988853
Root Mean Squared Error: 0.42770586359072577

```

Gambar 5. Hasil evaluasi pengujian model prediksi

C. Hasil dan Pembahasan

Hasil penelitian menunjukkan bahwa algoritma K-Means dan Linear Regression dapat memprediksi jumlah kejadian bencana yang terjadi di daerah dari 12 Kabupaten dan 1 Kota yang digunakan untuk membangun model dan menguji model prediksi. Hasil dari analisis ini dapat diperoleh hasil probabilitas terjadinya bencana di masa akan datang di berbagai wilayah Sulawesi Tengah dilihat pada tabel berikut :

Tabel 4. Hasil akhir dari analisis data bencana

No	Wilayah	Jumlah Kejadian	Kejadian paling terjadi	Probabilitas
1	Kab. Sigi	86	Banjir	0.088568
2	Kab. Morowali utara	90	Banjir	0.092688
3	Kab. Poso	135	Puting Beliung	0.139032
4	Kab. Banggai kepulauan	55	Gelombang Tinggi	0.056643

5	Kab. Tojo Una-una	54	Putting Beling	0.055613
6	Kota Palu	25	Banjir	0.025747
7	Kab. Parigi Moutong	96	Banjir	0.098867
8	Kab. Toli-toli	153	Putting Beliung	0.157570
9	Kab. Donggala	60	Banjir	0.061792
10	Kab. Banggai	67	Banjir	0.069001
11	Kab. Morowali	31	Putting Beliung	0.031926
12	Kab. Banggai Laut	28	Banjir	0.028836
13	Kab. Buol	91	Banjir	0.093718

D. Simpulan

Kesimpulan dari penerapan data mining untuk mendapatkan data probabilitas terjadinya bencana pada daerah di Sulawesi Tengah. Setelah melakukan clustering, data diklasifikasikan menjadi 3 cluster yaitu cluster banjir, putting beliung dan Gelombang Tinggi, masing-masing cluster tersebut memiliki entitas seperti banjir (Kabupaten Sigi, Morowali, Kota Palu, Parigi Moutong, Donggala, Banggai, Banggai laut dan Buol), putting beliung (Kabupaten Poso, Tojo Una-una, Toli-toli, Morowali), Gelombang Tinggi hanya memiliki 1 entitas wilayah yaitu kabupaten Banggai kepulauan. Menentukan probabilitas dari penelitian tersebut dilakukan dengan pemodelan Linear Regression serta menguji model menggunakan bahasa pemrograman python. dari hasil perhitungan tersebut didapatkan wilayah yang mengalami peningkatan probabilitas bencana banjir di wilayah Kabupaten Buol, sehingga probabilitas bencana meningkat putting beliung wilayah Kabupaten Parigi Moutong, dan untuk probabilitas peningkatan bencana gelombang tinggi wilayah Kabupaten Banggai Kepulauan. Hasil evaluasi pengujian model Linear Regression bahwa nilai RMSE 0.4277 memiliki satuan yang sama dengan nilai sebenarnya sehingga lebih mudah untuk diinterpretasikan. Penelitian tersebut dapat membantu dalam memprediksi probabilitas kejadian bencana di masa depan dan juga mengembangkan strategi penanggulangan bencana yang lebih efektif.

E. Ucapan Terima kasih

Penulis ingin mengucapkan terima kasih kepada Allah SWT, orang tua dan juga berterimakasih kepada rekan-rekan Universitas Tadulako, serta pihak Badan Penanggulangan Bencana Daerah (BPBD) Provinsi Sulawesi Tengah yang telah memberikan dukungan dan bantuan kepada penulis dalam proses menyelesaikan naskah ini.

F. Referensi

- [1] Nisa, F. (2014). Manajemen Penanggulangan Bencana Banjir, Putting Beliung, dan Tanah Longsor di Kabupaten Jombang. *JKMP (Jurnal Kebijakan Dan Manajemen Publik)*, 2(2), 103–116. <https://doi.org/10.21070/jkmp.v2i2.432>
- [2] Penelitian Dan Pengembangan, J., Mitigasi Bencana Tanah Longsor Berdasarkan Permendagri, K., Yan Prastowo, L., Setyo Wahyuningsih, A., Ilmu Kesehatan Masyarakat, J., Ilmu Keolahragaan, F., & Negeri Semarang, U. (2006). *60 JPPKMI 1 (1) (2020) Abstrak. 1(33), 60–71.*

- <https://journal.unnes.ac.id/sju/index.php/jppkmi>URL:<https://journal.unnes.ac.id/sju/index.php/jppkmi/article/view/41420/17343>
- [3] Hidayat, I., Darnila, E., & Afrillia, Y. (2023). Clustering Zonasi Daerah Rawan Bencana Alam di Kabupaten Mandailing Natal menggunakan Algoritma K-Means. *G-Tech: Jurnal Teknologi Terapan*, 7(3), 1218–1226. <https://doi.org/10.33379/gtech.v7i3.2880>
 - [4] Sugianto, C. A., & Astita, M. N. (2017). Implementasi Data Mining Dalam Data Bencana Tanah Longsor Di Jawa Barat Menggunakan Algoritma Fp-Growth. *Techno.Com*, 17(1), 91–102. <https://doi.org/10.33633/tc.v17i1.1601>
 - [5] Maulana, A., Danar Dana, R., & Nuris, N. D. (2024). Implementasi Algoritma K-Means Clustering Dalam Pengelompokan Data Kerusakan Rumah Akibat Bencana Alam Di Kabupaten Cirebon. *Jurnal Mahasiswa Teknik Informatika*, 8(2), 1417–1424.
 - [6] Rohman, D., Annisa, R., Indriyana Efendi, D., & Solahudin, D. (2024). Clustering Bencana Alam Menggunakan K-Means Pada Wilayah Jawa Barat. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 493–500. <https://doi.org/10.36040/jati.v8i1.8409>
 - [7] Somantri, O., & Maharrani, R. H. (2022). Metode Penilaian Kekuatan Gempa Menggunakan Model Feature Selection M5-Prime Dan Linear Regression. *Jurnal Informatika Polinema*, 9(1), 45–50. <https://doi.org/10.33795/jip.v9i1.989>
 - [8] Iqbal Ramadhan, M. (2017). Penerapan Data Mining untuk Analisis Data Bencana Milik Bnpb Menggunakan Algoritma K-Means dan Linear Regression. *Jurnal Informatika Dan Komputer*, 22(1), 57–65.
 - [9] Althaf, M., & Perkasa, P. (2023). 1, 2 1,2. 2.
 - [10] Dhewayani, F. N., Amelia, D., Alifah, D. N., Sari, B. N., & Jajuli, M. (2022). Implementasi K-Means Clustering untuk Pengelompokan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM. *Jurnal Teknologi Dan Informasi*, 12(1), 64–77. <https://doi.org/10.34010/jati.v12i1.6674>
 - [11] Pangestu, A., & Ridwan, T. (2022). Penerapan Data Mining Menggunakan Algoritma K-Means Pengelompokan Pelanggan Berdasarkan Kubikasi Air Terjual Menggunakan Weka. *JUST IT: Jurnal Sistem Informasi, Teknologi Informasi Dan Komputer*, 11(3), 67–71. <https://jurnal.umj.ac.id/index.php/just-it/article/view/11591>
 - [12] Wulandari, D., & Rumini, R. (2023). Pemodelan dan Prediksi Produksi Padi Menggunakan Regresi Linear. *Smart Comp: Jurnalnya Orang Pintar Komputer*, 12(4). <https://doi.org/10.30591/smartcomp.v12i4.5905>
 - [13] Sapuan, I., Fauzan, M. H., & Julianne, C. (2022). Implementasi Data Mining untuk Klasterisasi dan Prediksi Kelompok Keluarga. *JTERA (Jurnal Teknologi Rekayasa)*, 7(1), 149. <https://doi.org/10.31544/jtera.v7.i1.2022.149-156>
 - [14] Setianingsih, D., & Hakim, R. F. (2015). Penerapan Data Mining dalam Analisis Kejadian Tanah Longsor di Indonesia dengan Menggunakan Association Rule Algoritma Apriori. *Prosiding Seminar Nasional Matematika Dan Pendidikan Matematika UMS 2015*, 731–741.
 - [15] Durrotul. (2017). Implementasi Data Mining Dalam Prediksi Performance Software Engineer PT.Emerio Menggunakan Decision Tree. *Jurnal Ilmiah Informatika Komputer Universitas Gunadarma*, 22(1), 31–43.

- [16] Pebralia, J. (2022). Analisis Curah Hujan Menggunakan Machine Learning Metode Regresi Linier Berganda Berbasis Python dan Jupyter Notebook. *Jurnal Ilmu Fisika Dan Pembelajarannya (JIFP)*, 6(2), 23–30. <https://doi.org/10.19109/jifp.v6i2.13958>
- [17] Farhan, N. M., & Setiaji, B. (2023). Indonesian Journal of Computer Science. *Indonesian Journal of Computer Science*, 12(2), 284–301. <http://ijcs.stmikindonesia.ac.id/ijcs/index.php/ijcs/article/view/3135>