

**MODEL PEMBELAJARAN DAN LAPORAN AKHIR
PROJECT-BASED LEARNING
MATA KULIAH DATA MINING I
KELAS A**



**"KLASTERISASI DAERAH RAWAN BANJIR DI PROVINSI DKI JAKARTA
MENGGUNAKAN METODE HIERARCHICAL CLUSTERING"**

DISUSUN OLEH KELOMPOK II :

- | | |
|----------------------------|---------------------------|
| 1. ANGELA LISANTHONI | (21083010032) - KETUA |
| 2. RHEINKA ELYANA SUPRAPTO | (21083010021) - ANGGOTA |
| 3. EDINA ALANA NABILA | (21083010022) - ANGGOTA |
| 4. ELLEXIA LEONIE GUNAWAN | (21083010027) - ANGGOTA |
| 5. GEMA KHUSNUL MA'RIFAH | (21083010034) - ANGGOTA |
| 6. ALYSSA AMORITA AZZAH | (21083010057) - ANGGOTA |

DOSEN PENGAMPU:

TRESNA MAULANA FAHRUDIN, S.ST., MT (20219930501200)
TRIMONO, S.Si., M.Si (199509082022031003)

PROGRAM STUDI SAINS DATA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"
JAWA TIMUR
2023

KATA PENGANTAR

Puji syukur kami panjatkan kepada Tuhan Yang Maha Esa, dengan limpah karunia-Nya kelompok kami dapat menyelesaikan penyusunan Laporan Project Based Learning ini terselesaikan tepat pada waktunya.

Laporan ini bertujuan agar dapat menambah pengetahuan dan kreativitas dalam pembelajaran mata kuliah data mining khususnya dengan menggunakan metode hierarchical clustering. Selain itu, laporan ini juga bertujuan sebagai pemaparan penerapan klasterisasi terhadap daerah banjir di wilayah DKI Jakarta.

Dalam projek ini, kami memperoleh beberapa kesulitan, namun, berkat dukungan dari beberapa pihak, akhirnya projek ini dapat diselesaikan. Untuk itu, kami sepantasnya berterima kasih kepada:

1. Bapak Tresna Maulana Fahrudin S.ST., MT selaku dosen pengampu mata kuliah data mining.
2. Bapak Trimono, S.Si., M.Si selaku dosen pengampu mata kuliah data mining.

Laporan ini disusun berdasarkan materi-materi yang ada. Laporan yang kami buat ini masih jauh dari kesempurnaan. Oleh karena itu, kami mengharapkan kritik dan saran agar laporan ini lebih baik serta berguna dimasa yang akan datang.

DAFTAR ISI

KATA PENGANTAR.....	II
DAFTAR ISI	III
DAFTAR GAMBAR.....	V
DAFTAR TABEL	V
1. BAB I: PENDAHULUAN	1
1.1 LATAR BELAKANG	1
1.2 PERMASALAHAN	2
1.3 TUJUAN	2
1.4 MANFAAT	2
1.4.1. <i>Manfaat bagi Masyarakat</i>	2
1.4.2. <i>Manfaat bagi Industri</i>	3
1.4.3. <i>Manfaat bagi Pengembangan Ilmu</i>	3
2. BAB II: TINJAUAN PUSTAKA.....	4
2.1 TEORI PENUNJANG	4
2.1.1 <i>Provinsi DKI Jakarta</i>	4
2.1.2 <i>Hierarchical Clustering</i>	4
2.1.3 <i>Silhouette Score</i>	5
2.1.4 <i>Open Data</i>	6
2.1.5 <i>Python</i>	6
2.2 PENELITIAN TERKAIT.....	6
2.2.1. <i>Klasterisasi Menggunakan Agglomerative Hierarchical Clustering untuk Memodelkan Wilayah Banjir (Pratikto & Damastuti, 2021)</i>	6
2.2.2. <i>Pemetaan Zona Rawan Banjir di Jakarta Menggunakan Analytic Hierarchy Process (AHP) (Aldimasqie, Saputra, & Oktarina, 2022)</i>	7
2.2.3. <i>Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Dampak Bencana Banjir Tahun 2017-2020 Menggunakan K-Medoids (Islamy, Nursidah, Narendra, Anshori, & Widodo, 2022)</i>	7
3. BAB III: METODOLOGI PENELITIAN	8
3.1 PENGUMPULAN DATA	8
3.2 PREPROCESSING.....	8
3.2.1 <i>Menghilangkan Duplikasi Data</i>	8
3.2.2 <i>Menghapus Missing Value</i>	8
3.2.3 <i>Mengubah Format Data</i>	9
3.2.4 <i>Mengubah Kesalahan Penulisan Data</i>	9
3.2.5 <i>Reset Index</i>	9
3.3 FEATURE EXTRACTION	9
3.4 EXPLORATORY DATA ANALYSIS (EDA)	9

3.5	PEMODELAN	10
3.6	EVALUASI.....	10
3.7	VISUALISASI.....	10
3.8	DASHBOARD	10
4.	BAB IV: HASIL DAN PEMBAHASAN	11
4.1	PENGUMPULAN DATA	11
4.2	PREPROCESSING.....	11
4.2.1.	<i>Menghilangkan Duplikasi Data.....</i>	<i>11</i>
4.2.2.	<i>Menghapus missing value</i>	<i>11</i>
4.2.3.	<i>Mengubah Format Data</i>	<i>11</i>
4.2.4.	<i>Mengubah Kesalahan Penulisan Data.....</i>	<i>11</i>
4.2.5.	<i>Reset Index.....</i>	<i>12</i>
4.3	FEATURE EXTRACTION	12
4.4	EXPLORATORY DATA ANALYSIS (EDA)	12
4.5	PEMODELAN	14
4.6	EVALUASI.....	16
4.7	VISUALISASI.....	16
4.8	DASHBOARD	21
5.	BAB V: KESIMPULAN	37
6.	DAFTAR PUSTAKA.....	38
7.	LAMPIRAN	40

FINAL

DAFTAR GAMBAR

Gambar 1 Diagram Alur Penelitian	8
Gambar 2 Informasi Pada Data	13
Gambar 3 Statistika Deskriptif untuk Data Tipe Numerik	13
Gambar 4 Statistika Deskriptif untuk Data Tipe Objek	14
Gambar 5 Visualisasi Distribusi Variabel Prediktor	17
Gambar 6 Visualisasi Rata-Rata Jumlah Terdampak Jiwa Berdasarkan Kecamatan dan Bulan.....	17
Gambar 7 Visualisasi Terdampak Jiwa Setiap Kecamatan.....	18
Gambar 8 Visualisasi Terdampak Jiwa Kecamatan Pulo Gadung Setiap Bulan	18
Gambar 9 Visualisasi Terdampak Jiwa Kecamatan Makasar Setiap Bulan	19
Gambar 10 Visualisasi Matrix Correlation	19
Gambar 11 Visualisasi Hierarchical Clustering Dendrogram	20
Gambar 12 Visualisasi Scatter Plot Hasil Clustering	20
Gambar 13 Visualisasi Hasil Klasterisasi Menggunakan GeoPandas.....	21
Gambar 14 Dashboard yang Menunjukkan Hasil Klasterisasi Menggunakan GeoPandas.	22
Gambar 15 Dashboard yang Menunjukkan Bar Chart dan Line Chart.....	22
Gambar 16 Kode Script Penelitian.....	36

DAFTAR TABEL

Tabel 1 Hasil Klasterisasi berdasarkan Kecamatan	15
--	----

1. BAB I: PENDAHULUAN

1.1 Latar Belakang

Provinsi DKI Jakarta merupakan ibu kota negara dan kota terbesar di Indonesia. DKI Jakarta menjadi pusat dari kegiatan ekonomi nasional, politik, dan kebudayaan. Jakarta memiliki jumlah penduduk sebanyak 9041 juta jiwa dengan kepadatan penduduk 13667,01 jiwa per km². Jakarta berlokasi di sebelah Utara Pulau Jawa. Jakarta memiliki luas daratan 661,52 km² dan lautan seluas 6977,5 km² (RI, n.d.). Jakarta merupakan wilayah dataran rendah dengan ketinggian rata-rata 7 meter diatas permukaan laut (perkim, 2020). Jakarta merupakan daerah dataran rendah dengan titik tertinggi Jakarta adalah 91 meter diatas permukaan laut dan titik terendahnya yaitu -1 meter diatas permukaan laut dan juga wilayah Jakarta bagian Utara mengalami fenomena penurunan tanah sejak lama (Okta, Widyatama, & Utomo, 2022).

Dengan keadaan kepadatan penduduk yang sangat padat dan kondisi geologis yang berada pada wilayah dataran rendah, Provinsi DKI Jakarta memiliki berbagai permasalahan dan tantangan seperti kemacetan lalu lintas dan rawannya bencana banjir. Bencana banjir adalah salah satu bencana yang paling sering melanda Provinsi DKI Jakarta. Hal tersebut dikarenakan kurangnya lahan pemukiman dan lahan hijau di daerah Provinsi DKI Jakarta. Karena kurangnya lahan pemukiman, banyak penduduk yang mendirikan tempat tinggal di pinggiran sungai yang menyebabkan banyak limbah dan sampah yang terbuang di sungai. Selain itu, kurangnya lahan hijau juga menjadi penyebab Provinsi DKI Jakarta sering dilanda bencana banjir. Lahan hijau memiliki banyak peran penting dalam menjaga iklim dunia. Dalam kasus bencana banjir, lahan hijau berperan sebagai penyerap air dan menyimpan air hujan. Apabila lahan hijau berkurang dan digantikan dengan lahan pemukiman, maka bencana banjir akan sering melanda daerah tersebut.

Banjir dapat mengakibatkan beberapa kerugian. Kerugian yang ditimbulkan akibat dari adanya bencana banjir yaitu selain kerugian materil yang ditanggung oleh korban yang daerahnya terdampak oleh banjir dan juga kerugian lainnya seperti kerugian immateril dimana dapat menimbulkan adanya korban jiwa. Contoh kasus banjir DKI Jakarta yaitu pada pergantian tahun 2019 menuju 2020 dimana tercatat sebanyak 24 korban meninggal dunia, 724 wilayah

terdampak pemadaman listrik, arus lalu lintas tidak dapat beroperasi pada titik tertentu, dan terdapat kendaraan bermotor yang terbawa oleh arus air (Dwi & Nurhadi, 2022). Contoh lainnya yaitu pada tanggal 6/10/2022 banjir yang menimpa MTsN 19 di Pondok Labu, Jakarta Selatan menelan korban jiwa sebanyak 3 orang siswa yang disebabkan oleh runtuhnya dinding sekolah (Hakim, 2022).

Oleh karena itu, pada projek kali ini kelompok kami akan mengangkat kasus banjir pada wilayah DKI Jakarta sebagai permasalahan yang akan diselesaikan dengan menggunakan bantuan Machine Learning yaitu Clustering pada Unsupervised Learning dengan menggunakan metode Hierarchical clustering dalam melakukan klasterisasi terhadap daerah pada wilayah DKI Jakarta yang rawan mengalami banjir.

1.2 Permasalahan

- a. Sulitnya melakukan klasterisasi daerah rawan banjir di Provinsi DKI Jakarta berdasarkan indikator jumlah terdampak pada Rukun Warga (RW), Rukun Tetangga (RT), Kepala Keluarga (KK), dan jiwa.
- b. Belum tersedianya analisis klasterisasi daerah rawan banjir di Provinsi DKI Jakarta.

1.3 Tujuan

- a. Melakukan klasterisasi daerah rawan banjir di Provinsi DKI Jakarta berdasarkan indikator jumlah terdampak pada Rukun Warga (RW), Rukun Tetangga (RT), Kepala Keluarga (KK), dan jiwa.
- b. Tersedianya analisis klasterisasi daerah rawan bencana banjir di Provinsi DKI Jakarta menggunakan metode Hierarchical Clustering.

1.4 Manfaat

1.4.1. Manfaat bagi Masyarakat

Masyarakat khususnya warga DKI Jakarta dapat memperoleh manfaat yang signifikan dengan klasterisasi daerah rawan banjir di provinsi Jakarta menggunakan metode hierarchical clustering. Dengan menggunakan metode ini, masyarakat dapat memahami pola dan karakteristik banjir pada setiap daerah serta dapat mengetahui daerah yang memiliki kemungkinan tinggi untuk terkena

banjir dengan lebih akurat. Hal ini memungkinkan masyarakat DKI Jakarta untuk memperoleh informasi yang spesifik tentang risiko banjir yang mereka hadapi. Selain itu, hasil klasterisasi dapat digunakan sebagai dasar untuk pengembangan sistem peringatan dini yang lebih terarah. Dengan mengetahui daerah-daerah yang rentan, sistem peringatan dini dapat difokuskan pada wilayah tersebut sehingga masyarakat dapat menerima peringatan yang lebih cepat dan tepat.

1.4.2. Manfaat bagi Industri

Bagi pengusaha sektor infrastruktur pembangunan, seperti perusahaan real estate, manfaatnya adalah perusahaan dapat menyesuaikan rancangan pembangunan sesuai dengan kondisi daerah yang dikategorikan sebagai daerah rawan banjir. Dengan menggunakan metode ini, perusahaan dapat mengidentifikasi pola dan karakteristik daerah yang memiliki tingkat risiko banjir yang tinggi. Hal ini memungkinkan perusahaan untuk mengembangkan strategi pembangunan yang lebih efektif dan sesuai dengan kondisi setiap wilayah. Selain itu, klasterisasi daerah rawan banjir juga dapat membantu perusahaan dalam menentukan lokasi yang memiliki potensial untuk mengembangkan properti baru dengan mempertimbangkan faktor risiko banjir dan permintaan pasar. Hal tersebut dapat membantu perusahaan untuk mengadopsi solusi desain yang memperhitungkan faktor risiko banjir, seperti membangun drainase yang efektif, menggunakan material konstruksi yang tahan air, dan memilih lokasi yang lebih aman. Hal ini juga dapat dijadikan sebagai pengingat bagi perusahaan untuk mengolah dan memperhatikan limbah pabrik supaya pembuangannya sesuai dengan prosedur yang berlaku sehingga tidak terdapat pembuangan limbah yang menuju sungai dalam mencegah terjadinya banjir.

1.4.3. Manfaat bagi Pengembangan Ilmu

Universitas dapat menghasilkan peta klaster daerah rawan banjir yang lebih akurat dan terperinci. Hasil klasterisasi tersebut dapat menjadi sumber data yang berharga bagi universitas. Data yang dikumpulkan dan dianalisis dapat digunakan untuk penelitian lebih lanjut dan pemahaman yang lebih mendalam tentang faktor-faktor yang mempengaruhi banjir. Universitas dapat mempublikasikan temuan mereka, berbagi pengetahuan, dan berkontribusi dalam literatur ilmiah yang berkaitan dengan manajemen banjir dan mitigasi risiko.

2. BAB II: TINJAUAN PUSTAKA

2.1 Teori Penunjang

2.1.1 Provinsi DKI Jakarta

Provinsi DKI Jakarta merupakan provinsi yang memiliki posisi geografis pada $5^{\circ} 10' 00''$ LS - $6^{\circ} 22' 21,5''$ LS dan $106^{\circ} 41' 12,5''$ BT - $106^{\circ} 58' 24,2''$ BT . Selain itu, Provinsi DKI Jakarta dibagi menjadi 6 bagian menurut pembagian wilayah administrasi yaitu Jakarta Selatan dengan luas wilayah mencapai $154,32\text{ km}^2$, Jakarta Timur dengan luas daratan $182,70\text{ km}^2$, Jakarta Pusat dengan luas daratan $52,38\text{ km}^2$, Jakarta Utara dengan luas daratan $139,99\text{ km}^2$, Jakarta Barat dengan luas daratan $124,44\text{ km}^2$, dan yang terakhir adalah Kepulauan Seribu dengan luas wilayah $10,18\text{ km}^2$ (RI, n.d.).

Secara rincian administrasi, Provinsi DKI Jakarta memiliki 1 kabupaten, 5 kota, 44 kecamatan, 267 kelurahan, 2.741 rukun warga (RW), dan 30.470 rukun tetangga (RT). Selain itu, Provinsi DKI Jakarta memiliki 79 pulau di Kepulauan Seribu Utara dan 31 pulau di Kepulauan Seribu Sekatan. Dengan melihat luas dan banyaknya jumlah wilayah administrasi, Provinsi DKI Jakarta dikategorikan sebagai provinsi yang memiliki kepadatan penduduk paling tinggi. Dan apabila dihitung dengan rumus kepadatan penduduk, Provinsi DKI Jakarta memiliki jumlah kepadatan penduduk sebesar $13.6671,01$ jiwa per km^2 (Unit Pengelola Statistik Dinas Komunikasi, 2021).

2.1.2 Hierarchical Clustering

Hierarchical clustering adalah sebuah metode pengelompokan data yang dilakukan dengan cara hirarki. Pengelompokan data dilakukan dengan cara memvisualisasikan data dengan dendrogram, dimana data yang sama akan dikelompokkan dalam hirarki yang berdekatan. Dalam hierarchical clustering dibagi menjadi 2 jenis yaitu agglomerative atau pemusatan dan divisive atau pemisahan (Mustika, et al., 2021).

Dalam setiap setiap metode pasti memiliki kelemahan dan kelebihan masing-masing. Untuk hierarchical sendiri memiliki kelebihan sebagai berikut :

- a. Tidak perlu menentukan jumlah klaster yang diinginkan.
- b. Menghasilkan visualisasi yang mudah dipahami oleh pembaca.
- c. Tidak terpengaruh oleh inisialisasi yang acak atau penentuan centroid secara acak.

Untuk kekurangan dari hierarchical clustering adalah sebagai berikut :

- a. Hierarchical clustering tidak efisien untuk ukuran data yang sangat besar.
- b. Sulit mengolah data dengan skala yang berbeda.
- c. Tidak terbiasa terhadap noise dan outliers.
- d. Ketergantungan pada linkage method.

Secara manual, hierarchical clustering memiliki langkah-langkah dan rumus tertentu (Supianto, 2014). Langkah-langkah hierarchical clustering dengan agglomerative clustering adalah sebagai berikut :

- a. Membentuk matriks jarak, dengan rumus manhattan distance dan euclidean distance.

$$D_{man}(x, y) = \sum_{j=1}^d |x_j - y_j| \quad [1]$$

$$D(x_2, x_1) = \sqrt{\sum_{j=1}^d |x_j - y_j|^2} \quad [2]$$

- b. Menentukan jarak-jarak tertentu yaitu single linkage, complete linkage, dan average linkage.

Single Linkage

$$d_{uv} = \min\{d_{uv}\}, d_{uv} \in D \quad [3]$$

Complete Linkage

$$d_{uv} = \max\{d_{uv}\}, d_{uv} \in D \quad [4]$$

Average Linkage

$$d_{uv} = \text{average}\{d_{uv}\}, d_{uv} \in D \quad [5]$$

2.1.3 Silhouette Score

Silhouette score adalah metode machine learning yang digunakan untuk melihat seberapa baik data tersebut dikelompokkan dalam cluster terpilih. Selain itu, metode ini juga menilai kualitas sebuah kluster. Indikator penilaian kualitas baik atau buruknya suatu cluster adalah apabila nilai silhouette score mendekati 1 maka sampel tersebut baik dan cocok ditempatkan dalam klaster terpilih. Dan apabila silhouette score mendekati 0 maka sampel tersebut berada pada batas dua klaster yang membuat klaster data tersebut kurang jelas (Dewi & Pramita, 2019). Untuk nilai silhouette score mendekati -1 sampel tersebut tidak memiliki kecocokan dengan kluster terpilih, dan data tersebut lebih cocok ditempatkan pada cluster lainnya. Untuk rumus umum dari silhouette score adalah sebagai berikut (Wardy, Putra, & Rusjayanti, 2022):

$$Silhouette Score = \frac{(b-a)}{\max(a,b)}$$
 [6]

2.1.4 Open Data

Open data merupakan sebuah platform yang menyimpan berbagai judul dan jenis dataset. Dataset yang tersimpan dalam open data dapat diakses secara terbuka, dan sumber dataset tersebut sangat terpercaya karena berasal dari data pemerintah, lembaga riset, dan organisasi terkait. Pada penelitian ini, kami mengambil dataset bencana banjir di Provinsi DKI Jakarta Tahun 2020 yang diunggah oleh Badan Penanggulangan Bencana Daerah atau BPBD pada platform data.jakarta.go.id.

2.1.5 Python

Python adalah sebuah bahasa pemrograman tingkat tinggi yang mudah dipahami. Python memang didesain secara sederhana agar mudah dipahami oleh pemula dalam mempelajari sebuah pemrograman komputer. Python dinilai sederhana dikarenakan memiliki banyak library dan framework yang mendukung program yang akan diterapkan. Contoh dari library dari python adalah pandas, numpy, dan matplotlib.

Dalam data mining python juga sangat berperan penting untuk pemrograman pengolahan data. Seluruh library python sangat diperlukan dalam data mining, Pada pandas digunakan untuk manipulasi data, numpy digunakan untuk perhitungan numerik, dan matplotlib digunakan untuk visualisasi data.

2.2 Penelitian Terkait

2.2.1. Klasterisasi Menggunakan Agglomerative Hierarchical Clustering untuk Memodelkan Wilayah Banjir (Pratikto & Damastuti, 2021)

Penelitian ini dilakukan untuk membagi daerah terdampak banjir menjadi 3 kelompok dengan tujuan untuk mengetahui karakteristik dampak banjir yang tinggi, sedang, dan rendah yang kemudian hasilnya digunakan sebagai acuan untuk lokasi di Jawa Timur untuk ditampilkan ke dalam sistem informasi geografis. Selain itu, penelitian ini dilakukan untuk mengetahui ketepatan cluster optimal yang dihasilkan oleh metode Agglomerative Hierarchical Clustering dengan uji performansi menggunakan metode cophenetic correlation coefficient. Hasil dari penelitian ini menunjukkan bahwa metode average linkage

memberikan solusi cluster yang lebih baik dibandingkan dengan metode AHC lainnya yakni sebesar 0,92.

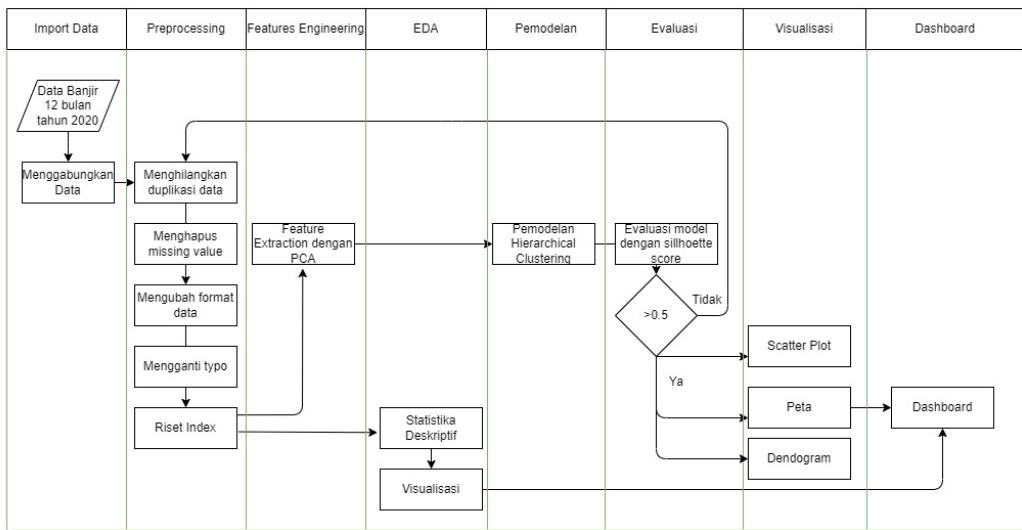
2.2.2. Pemetaan Zona Rawan Banjir di Jakarta Menggunakan Analytic Hierarchy Process (AHP) (Aldimasqie, Saputra, & Oktarina, 2022)

Penelitian ini dilakukan untuk mengetahui faktor yang mempengaruhi peristiwa banjir di Jakarta. Selain itu, penelitian ini juga bertujuan untuk menentukan wilayah rawan banjir di kota Jakarta menggunakan metode Analytic Hierarchy Process. Penelitian ini menunjukkan bahwa faktor yang paling berpengaruh adalah curah hujan dengan persentase sekitar 40,54%. Sementara faktor yang kurang berpengaruh adalah tata guna lahan dengan persentase sebesar 5,2%. Wilayah Jakarta memiliki kerawanan sedang terhadap banjir dimana wilayah yang paling rawan terkena banjir adalah Jakarta Selatan dan Jakarta Timur.

2.2.3. Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Dampak Bencana Banjir Tahun 2017-2020 Menggunakan K-Medoids (Islamy, Nursidah, Narendra, Anshori, & Widodo, 2022)

Penelitian ini bertujuan untuk mengelompokkan provinsi-provinsi di Indonesia berdasarkan indikator dampak bencana banjir menggunakan metode K-Medoids atau Partitioning Around Method. Hasil dari penelitian tersebut didapatkan dua klaster yang memiliki karakteristik tersendiri untuk mengelompokkan provinsi-provinsi di Indonesia. Klaster pertama terdiri dari 12 provinsi yang berisi daerah dampak banjir yang paling menonjol pada hampir tiap variabelnya. Sedangkan 22 provinsi lainnya berada di klaster kedua dengan karakteristik dampak banjir paling menonjol terdapat pada korban luka-luka.

3. BAB III: METODOLOGI PENELITIAN



Gambar 1 Diagram Alur Penelitian

3.1 Pengumpulan Data

Penelitian ini menggunakan data banjir DKI Jakarta bulanan sebanyak 12 bulan pada tahun 2020. Seluruh data kemudian digabungkan menjadi satu sehingga data yang sebelumnya sejumlah 12 file menjadi 1 file saja.

3.2 Preprocessing

Data yang telah digabungkan kemudian perlu dilakukan preprocessing. Preprocessing dilakukan melalui beberapa tahapan yaitu menghilangkan adanya data yang duplikasi, menghapus adanya missing value, mengubah format data yang belum sesuai, melakukan penggantian terhadap data yang mengalami kesalahan penulisan, dan melakukan riset index.

3.2.1 Menghilangkan Duplikasi Data

Langkah menghilangkan duplikasi data pada pre-processing data adalah proses penghapusan data yang memiliki duplikasi. Data yang terduplicasi perlu dihapus karena dapat memengaruhi hasil analisis menjadi tidak akurat. Dengan menghilangkan atau menghapus data yang terduplicasi, dataset yang digunakan untuk dianalisis menjadi konsisten dan mewakili variasi yang sebenarnya dari data yang ada.

3.2.2 Menghapus Missing Value

Missing value adalah keadaan dimana sebuah dataset memiliki nilai yang hilang atau tidak ada nilai yang tercatat untuk suatu variabel dan pengamatan.

Missing value terjadi dikarenakan beberapa alasan seperti kesalahan dalam pengumpulan data, data yang tidak lengkap, dan pengamatan yang hilang. Dalam sebuah analisis diperlukan untuk memeriksa apakah sebuah dataset memiliki sebuah *missing value*. Karena sebuah *missing value* dapat mengakibatkan bias, kesalahan estimasi, atau hasil yang tidak akurat.

3.2.3 Mengubah Format Data

Mengubah format tipe data adalah hal yang umum dilakukan dimana proses ini dilakukan dengan tujuan untuk mengubah struktur atau tipe data yang ada dalam dataset agar sesuai dengan kebutuhan analisis dan pemodelan yang akan dilakukan.

3.2.4 Mengubah Kesalahan Penulisan Data

Kesalahan penulisan data perlu dilakukan pengubahan karena data tersebut dapat memengaruhi hasil analisis.

3.2.5 Reset Index

Reset index dilakukan untuk mengatur ulang indeks dari sebuah dataset. Karena ketika melakukan manipulasi data seringkali indeks pada dataframe menjadi tidak teratur atau tidak sesuai dengan kebutuhan analisis. Dataframe yang telah diatur ulang indexnya akan lebih teratur, konsisten, dan sesuai dengan kebutuhan analisis selanjutnya.

3.3 Feature Extraction

Data selanjutnya melalui tahap feature extraction. Feature extraction adalah teknik untuk melakukan reduksi dimensi pada dataset. Pada penelitian ini, data melalui tahap feature extraction dengan melakukan pengurangan dimensi data menggunakan metode PCA (Principal Component Analysis).

3.4 Exploratory Data Analysis (EDA)

Pada tahap EDA, data yang sebelumnya telah melewati tahap reset index selanjutnya dilakukan analisis berupa statistika deskriptif yang selanjutnya dilakukan visualisasi data. Hasil visualisasi data kemudian selanjutnya akan digunakan pada dashboard. Exploratory Data Analysis (EDA) adalah metode analisis data yang digunakan untuk memahami suatu dataset. Tujuan dari EDA adalah untuk memperoleh informasi awal tentang dataset yang akan dianalisis

sebelum melakukan analisis statistik yang lebih mendalam atau membangun model prediksi. Untuk melakukan EDA dapat menggunakan analisis statistika deskriptif dengan mencari mean, jumlah data, standar deviasi, nilai kuartil, nilai maksimum, dan nilai minimum.

3.5 Pemodelan

Setelah data melalui tahap feature extraction maka selanjutnya dilakukan pemodelan terhadap data. Pemodelan terhadap data disini yaitu pemodelan dengan menggunakan metode agglomerative hierarchical clustering. Klasterisasi yang dilakukan pada tahap ini yaitu klasterisasi terhadap 4 indikator yaitu jumlah terdampak pada Rukun Warga (RW), Rukun Tetangga (RT), Kepala Keluarga (KK), dan jiwa.

3.6 Evaluasi

Setelah melalui tahap pemodelan selanjutnya melakukan evaluasi terhadap pemodelan yang telah dibuat yaitu dengan menggunakan silhouette score untuk melihat hasil evaluasi model. Apabila hasil evaluasi menunjukkan $>0,5$ maka selanjutnya akan dilakukan visualisasi, namun jika hasil evaluasi menunjukkan $<0,5$ maka data akan kembali kepada tahap preprocessing dan melakukan tahapan selanjutnya sesuai dengan bagan. Hal ini akan dilakukan berulang sampai hasil evaluasi menunjukkan $>0,5$.

3.7 Visualisasi

Visualisasi data menggunakan tiga jenis visualisasi yaitu visualisasi menggunakan scatter plot, peta, dan juga visualisasi menggunakan dendogram. Visualisasi peta menggunakan geopandas. Visualisasi menggunakan peta selanjutnya akan digunakan pada dashboard.

3.8 Dashboard

Dashboard dimaksudkan untuk menyampaikan hasil analisis yang telah dilakukan supaya lebih mudah dipahami dengan menggunakan visualisasi data. Dashboard disini akan menampilkan visualisasi data yang telah dihasilkan pada tahap EDA dan juga visualisasi peta yang dihasilkan sebelumnya.

4. BAB IV: HASIL DAN PEMBAHASAN

4.1 Pengumpulan Data

Tahap pengumpulan data dilakukan dengan mendownload data kejadian banjir di Provinsi DKI Jakarta tahun 2020 melalui Jakarta Open Data. Jakarta Open Data adalah sebuah opensource website yang dimana masyarakat dapat mendownload data secara bebas. Data yang diambil mencakup rentang waktu dari bulan Januari hingga Desember. Setelah proses pengunduhan selesai, data tersebut kemudian digabungkan menjadi satu file dalam format *.csv. Dalam file tersebut, terdapat baris-baris data yang mencatat kejadian banjir selama periode yang disebutkan. Data terdiri dari 1006 baris dan 18 kolom.

4.2 Preprocessing

4.2.1. Menghilangkan Duplikasi Data

Pada tahap ini melakukan penghapusan terhadap data yang terduplikasi sehingga tidak terjadi redundan yang dapat memengaruhi hasil analisis. Berdasarkan pada proses ini, penelitian kami ditemukan adanya ada yang terduplicat dan terjadi perubahan terhadap jumlah data yang sebelumnya sebanyak 1006 baris data menjadi 916 baris data dengan banyaknya data yang terduplicat yaitu sebanyak 90 baris data.

4.2.2. Menghapus missing value

Pada tahap ini melakukan penghapusan terhadap missing value yang terdapat dalam dataset guna menghindari bias, kesalahan estimasi, atau hasil yang tidak akurat. Pada penelitian kami tidak ditemukan adanya missing value setelah dilakukan pengecekan terhadap nilai null pada setiap kolom.

4.2.3. Mengubah Format Data

Tipe data yang diubah pada dataset kejadian banjir di Provinsi DKI Jakarta adalah variabel “jumlah_terdampak_kk” diubah menjadi numerik yang semula bertipe object menjadi integer.

4.2.4. Mengubah Kesalahan Penulisan Data

Data yang diubah kesalahan penulisannya adalah data yang ada pada variabel kecamatan. Data-data tersebut adalah Kecamatan Cilincing, Kecamatan Pulo Gadung, dan Kecamatan Mampang Prapatan. Selain data pada variabel

kecamatan, pengubahan kesalahan penulisan juga dilakukan pada variabel kota_administrasi dengan data Jakarta Utara.

4.2.5. Reset Index

Pada tahap ini merupakan proses untuk mengatur ulang indeks pada dataset setelah dilakukannya manipulasi terhadap data supaya indeks lebih teratur, konsisten, dan sesuai dengan kebutuhan analisis. Reset index ini dilakukan dengan menambahkan variabel baru yaitu index sehingga data yang sebelumnya terdapat 18 kolom menjadi 19 kolom data.

4.3 Feature Extraction

Tahap ini dilakukan reduksi terhadap data menggunakan Principal Component Analysis (PCA) dimana mengurangi dimensi data menjadi 2 dimensi saja. Pengurangan dimensi tidak diberlakukan terhadap seluruh variabel, melainkan hanya pada variabel tertentu saja yang selanjutnya akan digunakan untuk analisis lebih lanjut. Variabel-variabel tersebut yaitu variabel jumlah_terdampak_rw, jumlah_terdampak_rt, jumlah_terdampak_kk, dan jumlah_terdampak_jiwa.

4.4 Exploratory Data Analysis (EDA)

Berikut merupakan informasi detail mengenai data yang digunakan yaitu mengenai jumlah baris data, nama-nama kolom beserta jumlah data, tipe data, dan lain sebagainya. Dapat dilihat bahwa terdapat 916 entri dengan index dimulai dari 0 sampai 915, kemudian terdapat kolom sebanyak 19 kolom dilengkapi dengan info setiap kolom yaitu banyaknya data serta tipe datanya. Terdapat juga penjelasan mengenai jumlah tipe data yang ada yaitu int32 sebanyak 1 data, int64 sebanyak 13 data, dan object sebanyak 5 data. Terdapat informasi mengenai jumlah memori yang digunakan yaitu sebanyak 132,5+ KB.

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 916 entries, 0 to 915
Data columns (total 19 columns):
 #   Column           Non-Null Count Dtype  
 --- 
 0   index            916 non-null    int64  
 1   kota_administrasi 916 non-null    object  
 2   kecamatan         916 non-null    object  
 3   kelurahan         916 non-null    object  
 4   rw                916 non-null    object  
 5   jumlah_terdampak_rw 916 non-null    int64  
 6   jumlah_terdampak_rt 916 non-null    int64  
 7   jumlah_terdampak_kk 916 non-null    int32  
 8   jumlah_terdampak_jiwa 916 non-null    int64  
 9   ketinggian_air     916 non-null    object  
 10  tanggal_kejadian   916 non-null    int64  
 11  lama_genangan     916 non-null    int64  
 12  jumlah_meninggal   916 non-null    int64  
 13  jumlah_hilang      916 non-null    int64  
 14  jumlah_luka_berat 916 non-null    int64  
 15  jumlah_luka_ringan 916 non-null    int64  
 16  jumlah_pengungsi_tertinggi 916 non-null    int64  
 17  jumlah_tempat_pengungsian 916 non-null    int64  
 18  nilai_kerugian    916 non-null    int64  
dtypes: int32(1), int64(13), object(5)
memory usage: 132.5+ KB

```

Gambar 2 Informasi Pada Data

Berikut merupakan hasil dari perhitungan statistika deskriptif dimana hanya menampilkan kolom dengan tipe numerik. Terdapat perhitungan mengenai jumlah data, mean, standar deviasi, nilai minimal, kuartil bawah, kuartil tengah, kuartil atas, dan nilai maksimal.

	index	jumlah_terdampak_rw	jumlah_terdampak_rt	jumlah_terdampak_kk	jumlah_terdampak_jiwa	tanggal_kejadian	lama_genangan
count	916.000000	916.000000	916.000000	916.000000	916.000000	916.000000	916.000000
mean	474.426856	2.037118	3.629913	46.293668	165.136463	6.301310	0.020742
std	286.535797	2.389123	5.559174	206.817498	711.542445	4.118841	0.157181
min	0.000000	0.000000	1.000000	0.000000	0.000000	1.000000	0.000000
25%	228.750000	1.000000	1.000000	0.000000	0.000000	2.000000	0.000000
50%	457.500000	1.000000	1.000000	0.000000	0.000000	8.000000	0.000000
75%	719.250000	3.000000	4.000000	8.000000	27.250000	10.000000	0.000000
max	1004.000000	15.000000	46.000000	3762.000000	13450.000000	12.000000	2.000000
	jumlah_meninggal	jumlah_hilang	jumlah_luka_berat	jumlah_luka_ringan	jumlah_pengungsi_tertinggi	jumlah_tempat_pengungsian	nilai_kerugian
	916.0	916.0	916.0	916.000000	916.000000	916.000000	916.0
	0.0	0.0	0.0	98.530568	1.454148	0.013100	0.0
	0.0	0.0	0.0	375.088452	11.138624	0.147262	0.0
	0.0	0.0	0.0	0.000000	0.000000	0.000000	0.0
	0.0	0.0	0.0	0.000000	0.000000	0.000000	0.0
	0.0	0.0	0.0	0.000000	0.000000	0.000000	0.0
	0.0	0.0	0.0	4461.000000	271.000000	3.000000	0.0

Gambar 3 Statistika Deskriptif untuk Data Tipe Numerik

Selain itu, dilakukan perhitungan statistika deskriptif terhadap data dengan tipe object dimana yang ditampilkan informasi dari setiap kolom berupa jumlah data, banyaknya data unik, nilai yang paling sering muncul, dan frekuensi dari nilai yang paling sering muncul.

	kota_administrasi	kecamatan	kelurahan	rw	ketinggian_air
count	916	916	916	916	916
unique	6	44	220	356	183
top	Jakarta Timur	JATINEGARA	KAMPUNG MELAYU	0	20 cm
freq	342	84	57	180	64

Gambar 4 Statistika Deskriptif untuk Data Tipe Objek

Pada penelitian ini tidak diperlukan adanya pengecekan terhadap data apakah data berdistribusi normal atau tidak dikarenakan untuk melakukan hierarchical clustering tidak terdapat asumsi yang memerlukan data berdistribusi normal. Hierarchical clustering adalah metode yang digunakan untuk mengelompokkan objek berdasarkan kesamaan jarak atau kesamaan atributnya sehingga metode ini tidak bergantung pada distribusi normal data.

4.5 Pemodelan

Pemodelan terhadap data menggunakan metode agglomerative hierarchical clustering. Pada pemodelan terdapat fungsi yang digunakan untuk menghitung jarak antar titik berdasarkan Euclidean, perhitungan untuk single linkage, complete linkage, dan average linkage, kemudian terdapat fungsi dimana digunakan untuk memilih tipe perhitungan yang diinginkan, jika 0 maka perhitungan single linkage, jika 1 maka perhitungan complete linkage, dan selain 0 dan 1 merupakan perhitungan average linkage. Pada penelitian ini digunakan banyaknya klaster yaitu sebanyak 4 klaster dengan jenis perhitungan yaitu 2 yang artinya jenis perhitungan average linkage. Klaster 0 diartikan sebagai daerah sangat tidak rawan banjir, 1 sebagai daerah tidak rawan banjir, 2 sebagai daerah rawan banjir, dan 3 sebagai daerah sangat rawan terjadi banjir.

Berikut merupakan hasil klasterisasi berdasarkan pemodelan agglomerative hierarchical clustering:

Tabel 1 Hasil Klasterisasi berdasarkan Kecamatan

No	terdampak_jiwa	terdampak_kk	Klaster	Kecamatan
1.	7.295867	5.550171	0	Ciracas
2.	10.465539	-9.320051	0	Duren Sawit
3.	-16.208618	-4.268114	0	Kalideres
4.	-70.366236	-2.238483	0	Kebayoran Lama
5.	-61.336669	-2.059699	0	Tebet
6.	-41.785376	-2.483966	0	Jagakarsa
7.	-42.799660	-3.603340	0	Sawah Besar
8.	-40.700808	-4.159634	0	Kelapa Gading
9.	-45.525369	-7.135819	0	Mampang Prapatan
10.	-138.973612	0.007610	0	Cilandak
11.	-134.864593	-0.130612	0	Pesanggrahan
12.	-143.762546	0.178453	0	Grogol Petamburan
13.	-142.659643	0.139524	0	Johar Baru
14.	-141.215299	0.642372	0	Kebon Jeruk
15.	-144.982037	0.155823	0	Kebayoran Baru
16.	-145.751280	0.245373	0	Pademangan
17.	-154.326834	1.891663	0	Cipayung
18.	-154.116076	3.151491	0	Pasar Rebo
19.	-150.321454	-2.381085	0	Koja
20.	-171.498368	1.168179	0	Kepulauan Seribu Selatan
21.	-171.498368	1.168179	0	Menteng
22.	-171.498368	1.168179	0	Kepulauan Seribu Selatan
23.	-171.498368	1.168179	0	Menteng
24.	-171.501048	1.162526	0	Cempaka Putih
25.	-171.495072	1.181310	0	Kemayoran
26.	-171.498003	1.183166	0	Kepulauan Seribu Utara
27.	-171.504640	1.151072	0	Senen
28.	-171.506080	1.150457	0	Tambora

29.	-171.503290	1.155857	0	Tanjung Priok
30.	-171.507775	1.142519	0	Gambir
31.	-171.488120	1.231476	0	Taman Sari
32.	-166.032417	0.955274	0	Palmerah
33.	-102.756256	-1.734810	0	Penjaringan
34.	-100.645736	-1.316103	0	Setiabudi
35.	-106.139874	0.981587	0	Kramat Jati
36.	-116.336244	-0.648887	0	Kembangan
37.	-114.571001	1.593302	0	Pasar Minggu
38.	141.571614	-0.128639	1	Cakung
39.	177.392134	-11.237573	1	Cengkareng
40.	220.786776	5.964019	1	Jatinegara
41.	210.153778	-0.858129	1	Matraman
42.	100.598021	2.210224	1	Cilincing
43.	89.256474	-5.933891	1	Tanah Abang
44.	57.283713	-6.212048	1	Pancoran
45.	487.967222	-5.622706	2	Makasar
46.	1093.525501	58.918450	3	Pulo Gadung

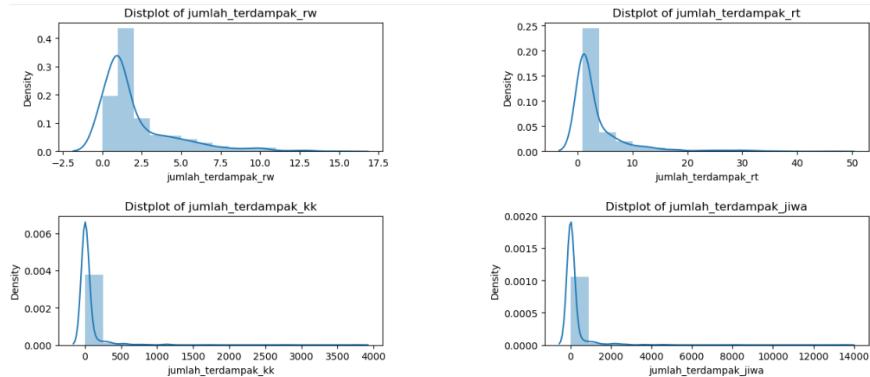
4.6 Evaluasi

Evaluasi performa menggunakan silhouette score dimana pada penelitian kami didapatkan hasil evaluasi performa sebesar 0,6875167194435193. Nilai tersebut menunjukkan bahwa klasterisasi yang dilakukan sudah relatif baik, dimana nilai silhouette score mendekati nilai 1. Hal tersebut mengindikasikan bahwa klasterisasi berhasil dengan baik dalam mengidentifikasi dan mengelompokkan daerah-daerah yang rawan banjir di DKI Jakarta kedalam klaster-klaster yang berbeda.

4.7 Visualisasi

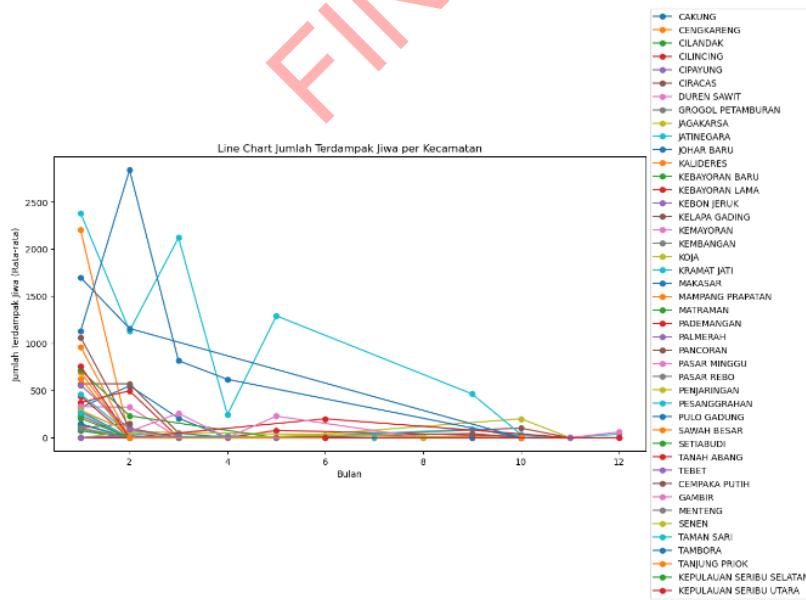
Grafik distribusi variabel prediktor digunakan untuk memvisualisasikan distribusi atau pola data pada variabel prediktor dalam suatu dataset. Tujuan utama adalah untuk memperoleh pemahaman tentang karakteristik distribusi data. Pada penelitian ini variabel prediktornya sebanyak 4 variabel yaitu

jumlah_terdampak_rw, jumlah_terdampak_rt, jumlah_terdampak_kk, dan jumlah_terdampak_jiwa. Berikut merupakan visualisasi distribusi variabel prediktor:



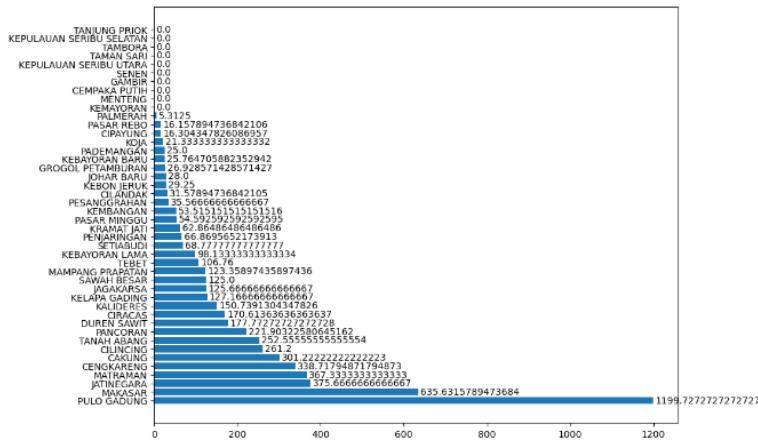
Gambar 5 Visualisasi Distribusi Variabel Prediktor

Visualisasi dibawah ini digunakan untuk mengetahui rata-rata jumlah terdampak jiwa setiap bulannya pada masing-masing kecamatan. Pada visualisasi ini dapat dilihat bahwa pada bulan-bulan tertentu terdapat jumlah terdampak jiwa yang tinggi seperti pada bulan Januari menuju bulan Februari. Selain itu, dapat dilihat pula kecamatan mana saja yang cenderung mengalami jumlah terdampak jiwa yang signifikan tinggi.



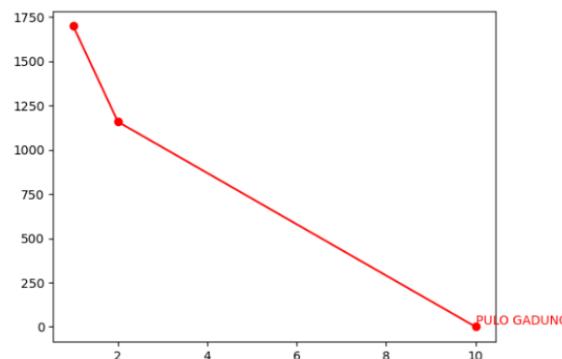
Gambar 6 Visualisasi Rata-Rata Jumlah Terdampak Jiwa Berdasarkan Kecamatan dan Bulan

Visualisasi dibawah ini digunakan untuk mengetahui frekuensi terdampak jiwa pada setiap kecamatan. Berdasarkan visualisasi tersebut dapat dilihat bahwa secara jumlah total jiwa yang terdampak pada peringkat pertama yaitu pada kecamatan Pulo Gadung dan pada peringkat kedua yaitu pada kecamatan Makasar. Hal tersebut berarti kedua daerah tersebut dapat dikatakan rawan berdasarkan total jiwa yang terdampak banjir.



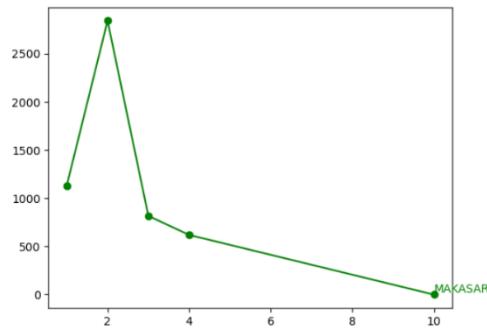
Gambar 7 Visualisasi Terdampak Jiwa Setiap Kecamatan

Berdasarkan hasil visualisasi sebelumnya yang menyatakan bahwa daerah kecamatan dengan frekuensi terdampak jiwa tertinggi urutan pertama yaitu pada kecamatan Pulo Gadung. Oleh karena itu, selanjutnya dapat dilakukan visualisasi mengenai jumlah terdampak jiwa pada kecamatan Pulo Gadung setiap bulannya dan berdasarkan pada hasil visualisasi didapatkan bahwa pada kecamatan Pulo Gadung, jumlah terdampak jiwa tertinggi yaitu terjadi pada bulan Januari dan bulan Februari kemudian terjadi penurunan pada bulan berikutnya.



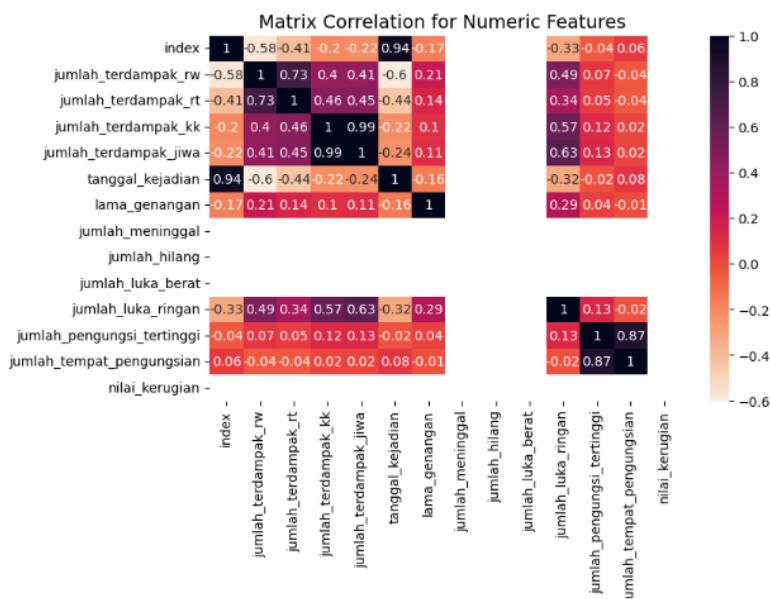
Gambar 8 Visualisasi Terdampak Jiwa Kecamatan Pulo Gadung Setiap Bulan

Berdasarkan hasil visualisasi sebelumnya yang menyatakan bahwa daerah kecamatan dengan frekuensi terdampak jiwa tertinggi urutan kedua yaitu pada kecamatan Makasar. Oleh karena itu, selanjutnya dapat dilakukan visualisasi mengenai jumlah terdampak jiwa pada kecamatan Makasar setiap bulannya dan berdasarkan pada hasil visualisasi didapatkan bahwa pada kecamatan Makasar, jumlah terdampak jiwa tertinggi yaitu terjadi pada bulan Februari dimana terjadi peningkatan dari bulan Januari menuju bulan Februari kemudian terjadi penurunan pada bulan berikutnya.



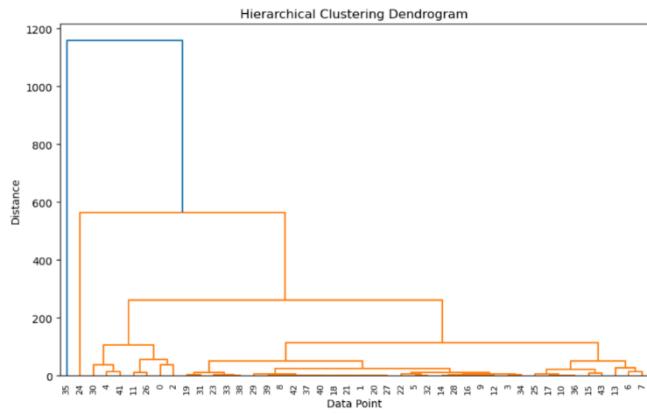
Gambar 9 Visualisasi Terdampak Jiwa Kecamatan Makasar Setiap Bulan

Visualisasi matriks korelasi adalah metode untuk memvisualisasikan hubungan korelasi antara berbagai variabel dalam sebuah dataset menggunakan matriks atau heatmap. Matriks korelasi menunjukkan sejauh mana dua variabel terkait satu sama lain, apakah hubungan tersebut positif (korelasi positif) atau negatif (korelasi negatif). Dapat dilihat bahwa korelasi tertinggi yaitu antara variabel jumlah_terdampak_kk dengan jumlah_terdampak_jiwa.



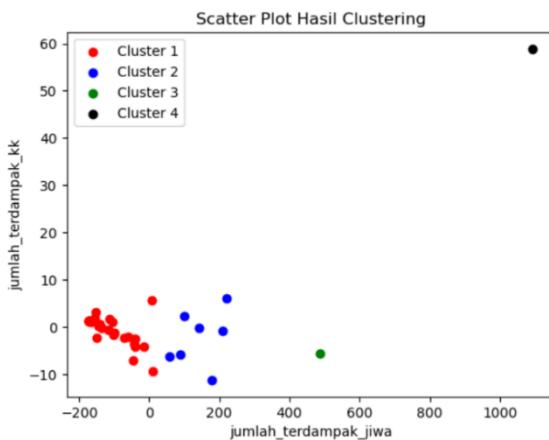
Gambar 10 Visualisasi Matrix Correlation

Visualisasi dendrogram digunakan untuk proses pengelompokan pada hierarchical clustering dimana kita dapat mengetahui visualisasi dari langkah-langkah dalam analisis clustering dengan menunjukkan bagaimana klaster terbentuk dan nilai koefisien jarak pada setiap langkah. Pada penelitian ini sumbu x menyatakan data poin dengan sumbu y menyatakan jarak antar data. Dendrogram disini menggunakan metode average linkage.



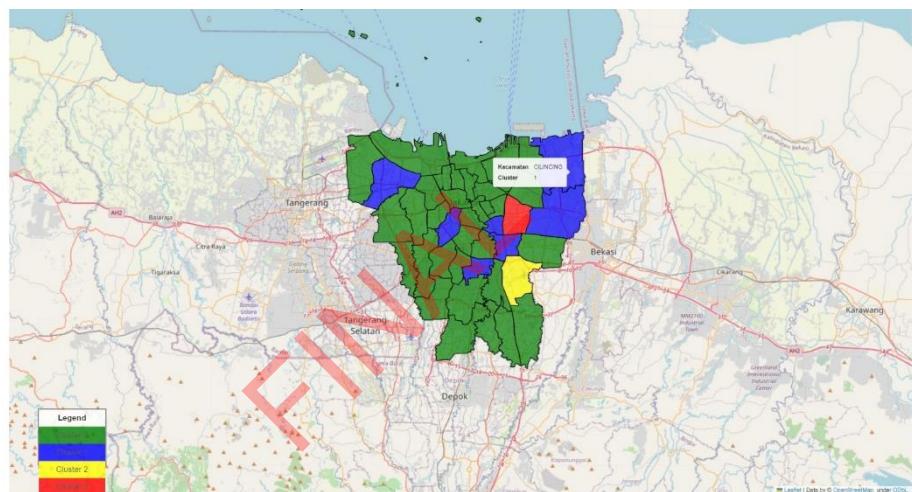
Gambar 11 Visualisasi Hierarchical Clustering Dendrogram

Visualisasi scatter plot adalah metode untuk memvisualisasikan hubungan antara dua variable numerik. Dengan scatter plot, kita dapat melihat hasil klasterisasi dan mengidentifikasi pola serta hubungan antara klaster yang terbentuk. Setiap titik data pada grafik menunjukkan sebuah kecamatan di DKI Jakarta, sedangkan warna menunjukkan klaster yang berbeda. Berdasarkan grafik scatter plot di bawah, dapat dilihat bahwa terdapat empat klaster yang terbentuk. Keempat klaster tersebut adalah klaster 1 yang ditandai dengan warna merah, klaster 2 yang ditandai dengan warna biru, klaster 3 yang ditandai dengan warna hijau, dan klaster 4 yang ditandai dengan warna hitam.



Gambar 12 Visualisasi Scatter Plot Hasil Clustering

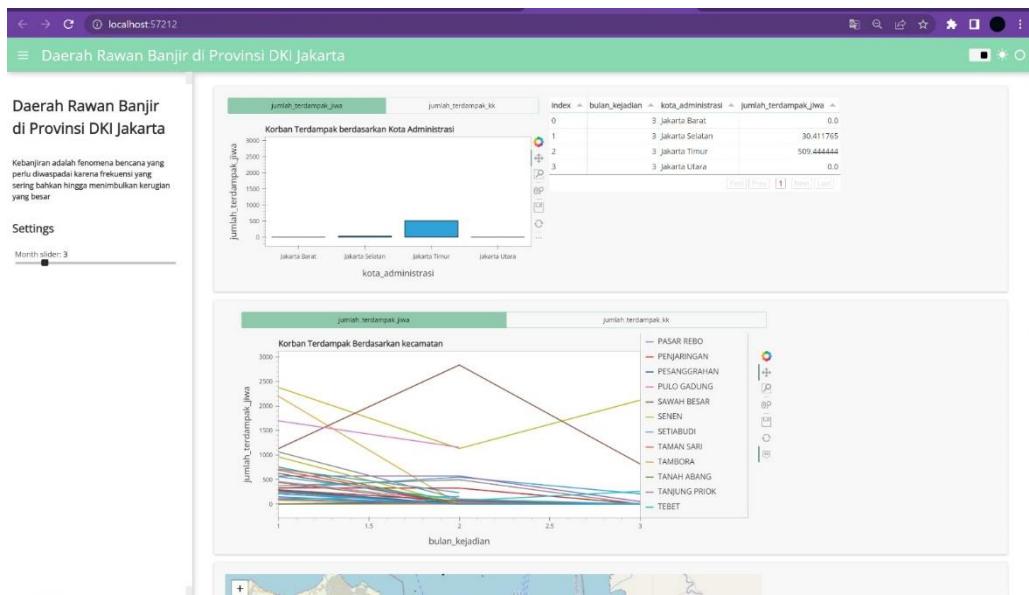
Visualisasi geopandas adalah metode untuk memvisualisasikan data geografis. Dalam visualisasi geopandas, data geografis, seperti peta, batas administratif, atau titik koordinat dapat diolah dan ditambilkan dalam bentuk visual. Pada analisis kali ini akan memvisualisasikan data kecamatan sesuai dengan frekuensi terjadinya bencana banjir. Berdasarkan visualisasi di atas, dapat terlihat bahwa terdapat empat klaster yang terbentuk. Klaster 0 ditandai dengan warna hijau yang bermakna bahwa daerah tersebut sangat tidak rawan banjir. Kemudian klaster 1 ditandai dengan warna biru yang bermakna bahwa daerah tersebut tidak rawan banjir. Selanjutnya klaster 2 ditandai warna kuning yang bermakna daerah tersebut rawan banjir. Terakhir klaster 3 ditandai warna merah yang bermakna bahwa daerah tersebut sangat rawan banjir.



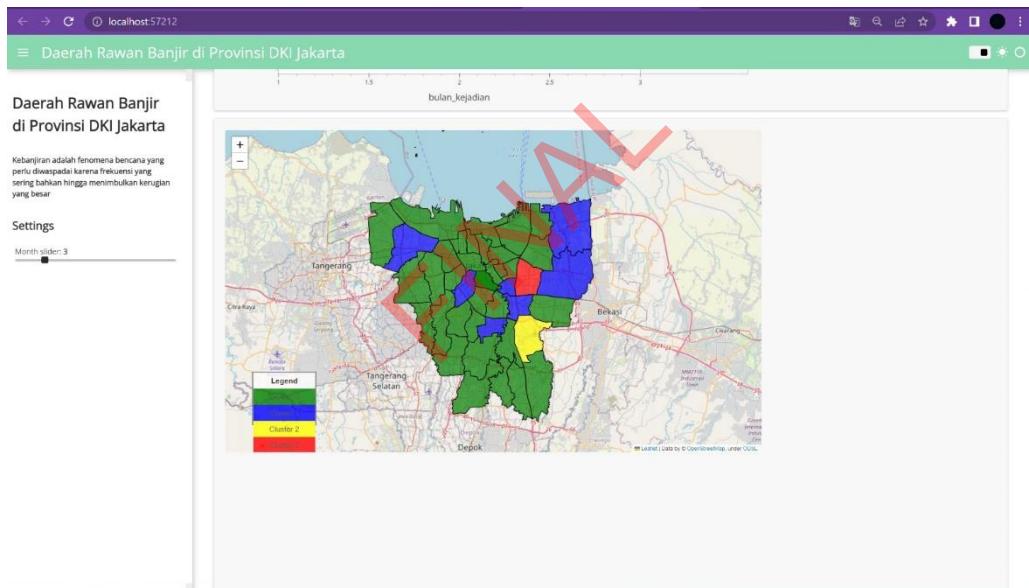
Gambar 13 Visualisasi Hasil Klasterisasi Menggunakan GeoPandas

4.8 Dashboard

Dashboard pada penelitian ini yaitu mencakup hasil visualisasi korban terdampak jiwa berdasarkan kota administrasi dengan bentuk visualisasi berupa bar chart. Kemudian terdapat hasil visualisasi korban terdampak jiwa berdasarkan kecamatan berupa line chart. Pada visualisasi bar chart maupun line chart dapat melihat visualisasi berdasarkan variabel jumlah_terdampak_jiwa dan jumlah_terdampak_kk. Terakhir terdapat visualisasi klasterisasi berupa peta yang menunjukkan tingkat kerawanan banjir pada setiap kecamatan. Terdapat fitur pada sebelah kiri yaitu untuk mengatur jumlah bulan yang akan muncul dalam visualisasi pada setiap slidenya.



Gambar 15 Dashboard yang Menunjukkan Bar Chart dan Line Chart



Gambar 14 Dashboard yang Menunjukkan Hasil Klasterisasi Menggunakan GeoPandas

```

=====PENGUMPULAN DATA=====
-----Import Library yang Digunakan-----
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.decomposition import PCA
from sklearn.metrics import silhouette_samples, silhouette_score
from sklearn.cluster import AgglomerativeClustering
from scipy.spatial.distance import pdist
from scipy.cluster.hierarchy import dendrogram, linkage
import math
import geopandas as gpd
-----Menggabungkan Dataset-----
januari = pd.read_csv('januari.csv')
januari['tanggal_kejadian'] = 1
januari.to_csv('januari.csv', index=False)

februari = pd.read_csv('februari.csv')
februari['tanggal_kejadian'] = 2
februari.to_csv('februari.csv', index=False)

maret = pd.read_csv('maret.csv')
maret['tanggal_kejadian'] = 3
maret.to_csv('maret.csv', index=False)

april = pd.read_csv('april.csv')
april['tanggal_kejadian'] = 4
april.to_csv('april.csv', index=False)

mei = pd.read_csv('mei.csv')
mei['tanggal_kejadian'] = 5
mei.to_csv('mei.csv', index=False)

juni = pd.read_csv('juni.csv')
juni['tanggal_kejadian'] = 6
juni.to_csv('juni.csv', index=False)

juli = pd.read_csv('juli.csv')
juli['tanggal_kejadian'] = 7
juli.to_csv('juli.csv', index=False)

agustus = pd.read_csv('agustus.csv')
agustus['tanggal_kejadian'] = 8
agustus.to_csv('agustus.csv', index=False)

september = pd.read_csv('september.csv')
september['tanggal_kejadian'] = 9

```

```

september.to_csv('september.csv', index=False)

oktober = pd.read_csv('oktober.csv')
oktober['tanggal_kejadian'] = 10
oktober.to_csv('oktober.csv', index=False)

november = pd.read_csv('november.csv')
november['tanggal_kejadian'] = 11
november.to_csv('november.csv', index=False)

desember = pd.read_csv('desember.csv')
desember['tanggal_kejadian'] = 12
desember.to_csv('desember.csv', index=False)

-----Import Dataset-----
# Daftar nama file CSV yang akan digabungkan
file_names = ['januari.csv', 'februari.csv', 'maret.csv', 'april.csv', 'mei.csv', 'juni.csv',
'juli.csv', 'agustus.csv', 'september.csv', 'oktober.csv', 'november.csv', 'desember.csv']

# Membuat daftar kosong untuk menyimpan data dari setiap file CSV
dataframes = []

# Membaca setiap file CSV dan menyimpannya sebagai dataframe
for file_name in file_names:
    dataframe = pd.read_csv(file_name)
    dataframes.append(dataframe)

# Menggabungkan data dari semua dataframe menjadi satu dataframe
merged_dataframe = pd.concat(dataframes)

# Menyimpan dataframe hasil penggabungan ke file CSV baru
merged_dataframe.to_csv('gabungan.csv', index=False)

df = pd.read_csv('gabungan.csv')
df

=====PREPROCESSING=====
-----Duplikasi Data-----
df.duplicated().sum()

df.drop_duplicates(inplace=True)

df.shape

df

-----Missing Value-----
#Cek nilai Null
print(df.isna().sum())

```

```

-----Mengubah Format Data yang Sesuai-----
df.info()

df['jumlah_terdampak_kk'] = pd.to_numeric(df['jumlah_terdampak_kk'],
errors='coerce').fillna(0).astype(int)
df.info()

-----Mengubah Kesalahan Penulisan Data-----
df['kecamatan'].unique()

df.loc[df['kecamatan'] == 'CILIINCING', 'kecamatan'] = 'CILINCING'
df.loc[df['kecamatan'] == 'PULOGADUNG', 'kecamatan'] = 'PULO GADUNG'
df.loc[df['kecamatan'] == 'MAMPANG', 'kecamatan'] = 'MAMPANG PRAPATAN'

df['kecamatan'].unique()

df['kota_administrasi'].unique()

df.loc[df['kota_administrasi'] == 'Jakarta Urata', 'kota_administrasi'] = 'Jakarta Utara'
df['kota_administrasi'].unique()

-----Reset Index-----
df = df.reset_index()
df

=====FEATURE ENGINEERING=====
# Menerapkan PCA untuk mengurangi dimensi menjadi 2
pca = PCA(n_components=2)
X_pca =
pca.fit_transform(df[['jumlah_terdampak_rw','jumlah_terdampak_rt','jumlah_terda-
mpak_kk','jumlah_terdampak_jiwa']])

# Mendapatkan loadings dari PCA
loadings = pd.DataFrame(pca.components_.T, columns=['PC1', 'PC2'],
index=['jumlah_terdampak_rw','jumlah_terdampak_rt','jumlah_terdampak_kk','jumlah_
terdampak_jiwa'])

# Menampilkan loadings
print(loadings)

# Membuat DataFrame baru dengan fitur baru dan kolom Klaster
df_reduction = pd.DataFrame({'jumlah_terdampak_jiwa': X_pca[:, 0],
'jumlah_terdampak_kk': X_pca[:, 1]})
df_reduction

df2=df[['kecamatan']]

df_merged1 = pd.concat([df2, df_reduction], axis=1)

```

```

df_merged1

# Menghitung jumlah_terdampak_jiwa sesuai kecamatan yang koresponden
df_final = df_merged1.groupby('kecamatan').mean().reset_index()

# Menampilkan hasil
df_final

df_dampak1 = df_final[['jumlah_terdampak_jiwa','jumlah_terdampak_kk']]
df_dampak1

=====EXPLORATORY DATA ANALYSIS=====
df.info()

df.shape

df.describe()

df.describe(include='object')

=====VISUALISASI DATA=====
-----Grafik Distribusi Variabel Prediktor-----
# grafik distribusi variabel prediktor
plt.figure(1 , figsize = (15 , 6))
graph = 0
# for loop
for x in ['jumlah_terdampak_rw' , 'jumlah_terdampak_rt' , 'jumlah_terdampak_kk',
'jumlah_terdampak_jiwa']:
    graph += 1

    # plotting graph
    plt.subplot(2 , 2 , graph)
    plt.subplots_adjust(hspace = 0.5 , wspace = 0.5)
    sns.distplot(df[x] , bins = 15)
    plt.title('Distplot of {}'.format(x))
# showing the graph
plt.show()

-----Rata-Rata Jumlah Terdampak Jiwa per Kecamatan dan Bulan-----
# Menghitung rata-rata jumlah_terdampak_jiwa per kecamatan dan bulan
df_grouped = df.groupby(['tanggal_kejadian', 'kecamatan']).mean().reset_index()

# Membuat line chart
plt.figure(figsize=(12, 6)) # Ukuran plot

# Loop melalui setiap kecamatan
for kecamatan in df_grouped['kecamatan'].unique():
    # Filter data untuk kecamatan tertentu
    data_kecamatan = df_grouped[df_grouped['kecamatan'] == kecamatan]

```

```

# Mengatur sumbu x dan y
x = data_kecamatan['tanggal_kejadian']
y = data_kecamatan['jumlah_terdampak_jiwa']

# Menggambar line chart untuk setiap kecamatan
plt.plot(x, y, marker='o', label=kecamatan)

# Memberi label sumbu x dan y, serta judul plot
plt.xlabel('Bulan')
plt.ylabel('Jumlah Terdampak Jiwa (Rata-rata)')
plt.title('Line Chart Jumlah Terdampak Jiwa per Kecamatan')

# Menampilkan legenda di sisi kanan plot
plt.legend(loc='center left', bbox_to_anchor=(1, 0.5))

# Menampilkan plot
plt.show()

-----Jumlah Terdampak Jiwa Setiap Kecamatan-----
# Menghitung jumlah terdampak jiwa tiap kecamatan
jumlah_terdampak =
df.groupby('kecamatan')['jumlah_terdampak_jiwa'].mean().reset_index()

# Mengurutkan kecamatan berdasarkan jumlah terdampak jiwa secara menurun
jumlah_terdampak = jumlah_terdampak.sort_values('jumlah_terdampak_jiwa',
ascending=False)

# Membuat visualisasi barh
plt.figure(figsize=(10, 8)) # Ukuran plot

# Menggambar barh untuk tiap kecamatan
plt.barh(jumlah_terdampak['kecamatan'],
jumlah_terdampak['jumlah_terdampak_jiwa'])

# Menambahkan label nilai di atas setiap barh
for i, v in enumerate(jumlah_terdampak['jumlah_terdampak_jiwa']):
    plt.text(v + 5, i, str(v), ha='left', va='center')

-----Jumlah Terdampak Jiwa Maksimum Setiap Kecamatan-----
# Menghitung jumlah_terdampak_jiwa maksimum per kecamatan
df_max = df.groupby('kecamatan')['jumlah_terdampak_jiwa'].mean().reset_index()
df_max = df_max.sort_values('jumlah_terdampak_jiwa', ascending=False)

# Mengambil kecamatan dengan jumlah_terdampak_jiwa maksimum
kecamatan_tertinggi = df_max.loc[df_max['jumlah_terdampak_jiwa'].idxmax(),
'kecamatan']

# Menandai kecamatan dengan jumlah_terdampak_jiwa maksimum pada line chart

```

```

data_kecamatan_tertinggi = df_grouped[df_grouped['kecamatan'] == kecamatan_tertinggi]
plt.plot(data_kecamatan_tertinggi['tanggal_kejadian'],
          data_kecamatan_tertinggi['jumlah_terdampak_jiwa'],
          marker='o', label=kecamatan_tertinggi, color='red')

# Menampilkan teks kecamatan tertinggi
plt.text(data_kecamatan_tertinggi['tanggal_kejadian'].iloc[-1],
          data_kecamatan_tertinggi['jumlah_terdampak_jiwa'].iloc[-1],
          kecamatan_tertinggi,
          ha='left', va='bottom', color='red')

# Menampilkan plot
plt.show()

-----Rata-Rata Jumlah Terdampak Jiwa Setiap Kecamatan-----
# Menghitung rata-rata jumlah_terdampak_jiwa per kecamatan
df_mean = df.groupby('kecamatan')['jumlah_terdampak_jiwa'].mean().reset_index()
df_mean_sorted = df_mean.sort_values('jumlah_terdampak_jiwa', ascending=False)

# Mengambil kecamatan dengan rata-rata jumlah_terdampak_jiwa tertinggi kedua
kecamatan_tertinggi_kedua = df_mean_sorted.iloc[1]['kecamatan']

# Menandai kecamatan dengan rata-rata jumlah_terdampak_jiwa tertinggi kedua pada line chart
data_kecamatan_tertinggi_kedua = df_grouped[df_grouped['kecamatan'] == kecamatan_tertinggi_kedua]
plt.plot(data_kecamatan_tertinggi_kedua['tanggal_kejadian'],
          data_kecamatan_tertinggi_kedua['jumlah_terdampak_jiwa'],
          marker='o', label=kecamatan_tertinggi_kedua, color='green')

# Menampilkan teks kecamatan tertinggi kedua
plt.text(data_kecamatan_tertinggi_kedua['tanggal_kejadian'].iloc[-1],
          data_kecamatan_tertinggi_kedua['jumlah_terdampak_jiwa'].iloc[-1],
          kecamatan_tertinggi_kedua,
          ha='left', va='bottom', color='green')

# Menampilkan plot
plt.show()

-----Matrix Correlation-----
plt.figure(figsize = (8,5))
correlation_matrix = df.corr().round(2)

sns.heatmap(data = correlation_matrix, annot = True, cmap = 'rocket_r')
plt.title("Matrix Correlation for Numeric Features ", size = 14)

=====PEMODELAN=====
def distance(p, q): #menghitung jarak antar titik berdasarkan ecludiean

```

```

return math.sqrt(sum([(pi - qi)**2 for pi, qi in zip(p, q)]))

def single_link(ci, cj): #menghitung berbasis single linkage
    return min([distance(vi, vj) for vi in ci for vj in cj])

def complete_link(ci, cj): #menghitung berbasis complete_link
    return max([distance(vi, vj) for vi in ci for vj in cj])

def average_link(ci, cj): #menghitung berbasis average_link
    distances = [distance(vi, vj) for vi in ci for vj in cj]
    return sum(distances) / len(distances)

def get_distance_measure(M): #user dapat memilih tipe perhitungan mana
    if M == 0:
        return single_link
    elif M == 1:
        return complete_link
    else:
        return average_link

class AgglomerativeHierarchicalClustering:
    def __init__(self, data, K, M):
        self.data = data
        self.N = len(data)
        self.K = K
        self.measure = get_distance_measure(M)
        self.clusters = self.init_clusters()

    def init_clusters(self): #setiap point adalah clusternya sendiri
        return {data_id: [data_point] for data_id, data_point in enumerate(self.data)}

    def find_closest_clusters(self): #mencari cluster lain yang terdekat
        min_dist = math.inf
        closest_clusters = None

        clusters_ids = list(self.clusters.keys())

        for i, cluster_i in enumerate(clusters_ids[:-1]):
            for j, cluster_j in enumerate(clusters_ids[i+1:]):
                dist = self.measure(self.clusters[cluster_i], self.clusters[cluster_j])
                if dist < min_dist:
                    min_dist, closest_clusters = dist, (cluster_i, cluster_j)
        return closest_clusters

    def merge_and_form_new_clusters(self, ci_id, cj_id): #gabungkan cluster2 yang
        berdekatan menjadi satu
        new_clusters = {0: self.clusters[ci_id] + self.clusters[cj_id]}

        for cluster_id in self.clusters.keys():

```

```

if (cluster_id == ci_id) | (cluster_id == cj_id):
    continue
    new_clusters[len(new_clusters.keys())] = self.clusters[cluster_id]
return new_clusters

def run_algorithm(self): #menjalankan algoritma
    while len(self.clusters.keys()) > self.K:
        closest_clusters = self.find_closest_clusters()
        self.clusters = self.merge_and_form_new_clusters(*closest_clusters)

def print(self): #menampilkan hasil clusterisasi
    for id, points in self.clusters.items():
        print("Cluster: {}".format(id))
        for point in points:
            print("  {}".format(point))

def data_frame(self):
    # Membuat list untuk menyimpan nilai kolom
    kolom_1 = []
    kolom_2 = []
    #kolom_3 = []
    #kolom_4 = []
    cluster = []

    # Mengisi list dengan nilai dari hasil klasterisasi
    for cluster_id, points in self.clusters.items():
        for point in points:
            kolom_1.append(point[0])
            kolom_2.append(point[1])
            #kolom_3.append(point[2])
            #kolom_4.append(point[3])
            cluster.append(cluster_id)

    # Membuat DataFrame dari list
    df_baru = pd.DataFrame({
        'jumlah_terdampak_jiwa': kolom_1,
        'jumlah_terdampak_kk': kolom_2,
        #'jumlah_terdampak_kk': kolom_3,
        #'jumlah_terdampak_jiwa': kolom_4,
        'cluster': cluster
    })

    # Menampilkan DataFrame
    return df_baru

dataset = df_dampak1.values

K = 4 #banyak kluster yang dipilih
M = 2 #jenis perhitungan yang dipilih

```

```

# 0 adalah single linkage
# 1 adalah complete linkage
# 2 adalah average linkage

agg_hierarchical_clustering = AgglomerativeHierarchicalClustering(dataset, K, M)
#memanggil class
agg_hierarchical_clustering.run_algorithm()
agg_hierarchical_clustering.print()

# 3 paling rawan
# 2 cukup rawan
# 1 sedang - sedang aja
# 0 tidak rawan

df_baru = agg_hierarchical_clustering.data_frame()
df_baru

gabungan = pd.merge(df_baru, df_final, on=['jumlah_terdampak_jiwa',
'jumlah_terdampak_kk'])
gabungan

=====EVALUASI=====
# Menghitung nilai Silhouette score
silhouette_avg_Hierarchical =
silhouette_score(gabungan[['jumlah_terdampak_jiwa','jumlah_terdampak_kk']],
gabungan[['cluster']])
print("Nilai Silhouette score:", silhouette_avg_Hierarchical)

# Menghitung shadow silhouette untuk setiap sampel
sample_silhouette_Hierarchical =
silhouette_samples(gabungan[['jumlah_terdampak_jiwa','jumlah_terdampak_kk']],
gabungan[['cluster']])
sample_silhouette_Hierarchical

-----Hierarchical Clustering Dendogram-----
# Menghitung matriks jarak antar data
dist_matrix = pdist(df_dampak1)

# Menggunakan metode linkage untuk menghubungkan klaster
linkage_matrix = linkage(dist_matrix, method='average')

# Membuat dendrogram
plt.figure(figsize=(10, 6))
dendrogram(linkage_matrix)
plt.title('Hierarchical Clustering Dendrogram')
plt.xlabel('Data Point')
plt.ylabel('Distance')
plt.show()

```

```

-----Scatter Plot Hasil Hierarchical Clustering-----
# Menentukan warna untuk setiap cluster
colors = ['red', 'blue', 'green','black'] # Sesuaikan dengan jumlah cluster Anda

clusters = [1,2,3,4]
# Membuat scatter plot untuk setiap cluster
for cluster_id, color in zip(range(len(clusters)), colors):
    cluster_data = gabungan[gabungan['cluster'] == cluster_id]
    plt.scatter(cluster_data['jumlah_terdampak_jiwa'],
                cluster_data['jumlah_terdampak_kk'], color=color, label=f'Cluster {cluster_id + 1}')

# Menambahkan label dan judul plot
plt.xlabel('jumlah_terdampak_jiwa')
plt.ylabel('jumlah_terdampak_kk')
plt.title('Scatter Plot Hasil Clustering')

# Menampilkan legenda
plt.legend()

# Menampilkan plot
plt.show()

-----Visualisasi GeoPandas-----
# Membaca file geopackage
data = gpd.read_file('Kecamatan-DKI-Jakarta.gpkg')

# Menampilkan informasi data
data.head()

# Mengubah nilai dalam kolom menjadi huruf kapital
data['name'] = data['name'].str.upper()
data.head()

# mengecek apakah nama dari kecamatannya sudah sama
sorted(df['kecamatan'].unique()) == sorted(data['name'].unique())

#ubah nama kolom
data = data.rename(columns={'name': 'kecamatan'})
data.head()

#agar kolom cluster muncul di dalam dataframe yang sama
peta_gabungan = data.merge(gabungan[['cluster','kecamatan']], on='kecamatan')

import folium

# Kolom 'cluster' berisi hasil klasterisasi
# Ubah nilai kolom 'cluster' menjadi tipe data string untuk memastikan konsistensi
tipe data
peta_gabungan['cluster'] = peta_gabungan['cluster'].astype(str)

```

```

# Membuat peta interaktif menggunakan library Folium
m = folium.Map(location=[-6.2088, 106.8456], zoom_start=11) # Menentukan
koordinat awal dan tingkat zoom

# Menentukan skema warna
color_scale = ['green', 'blue', 'yellow', 'red'] # Warna yang akan digunakan

# Menambahkan layer peta kecamatan ke peta interaktif dengan skema warna
folium.GeoJson(
    peta_gabungan,
    name='Kecamatan',
    style_function=lambda feature: {
        'fillColor': color_scale[int(feature['properties']['cluster'])]],
        'color': 'black',
        'weight': 1,
        'fillOpacity': 0.7
    },
    tooltip=folium.features.GeoJsonTooltip(fields=['kecamatan', 'cluster'],
    aliases=['Kecamatan', 'Cluster'], labels=True)
).add_to(m)

# Menambahkan custom legend
legend_html = """
<div style="position: fixed;
bottom: 50px; left: 50px; width: 120px; height: 100px;
border:2px solid grey; z-index:9999; font-size:14px;
background-color: white;
opacity: 0.8;
">
<p style="margin: 0; padding: 5px; color: black; text-align:center;"><strong>Legend</strong></p>
<ul style="list-style-type: none; padding: 0; margin: 0; text-align:center;">
    <li style="background-color:green; padding: 5px;">Cluster 0</li>
    <li style="background-color:blue; padding: 5px;">Cluster 1</li>
    <li style="background-color:yellow; padding: 5px;">Cluster 2</li>
    <li style="background-color:red; padding: 5px;">Cluster 3</li>
</ul>
</div>
"""

m.get_root().html.add_child(folium.Element(legend_html))

# Menampilkan peta interaktif
m

=====
=====DASHBOARD=====
import panel as pn
import folium

```

```

import hvplot.pandas

def environment():
    try:
        get_ipython()
        return "notebook"
    except:
        return "server"
environment()

pn.extension('tabulator')

df1 = df.copy()

# Make DataFrame Pipeline Interactive
idf = df1.interactive()

idf

df1

df1.rename(columns={'tanggal_kejadian': 'bulan_kejadian'}, inplace=True)
df1

# Define Panel widgets
month_slider = pn.widgets.IntSlider(name='Month slider', start=1, end=12, step=1,
value=3)
print(month_slider)

month_slider

yaxis_kota_administrasi = pn.widgets.RadioButtonGroup(
    name='Y axis',
    options=['jumlah_terdampak_jiwa','jumlah_terdampak_kk'],
    button_type='success'
)

kota_administrasi = df1['kota_administrasi'].unique()
kota_administrasi_bar_pipeline = (
    idf[
        (idf.bulan_kejadian == month_slider) &
        (idf.kota_administrasi.isin(kota_administrasi))
    ]
    .groupby(['bulan_kejadian', 'kota_administrasi'])[yaxis_kota_administrasi].mean()
    .to_frame()
    .reset_index()
    .sort_values(by='bulan_kejadian')
    .reset_index(drop=True)
)

```

```

)
kota_administrasi_bar_pipeline

kota_administrasi_bar_plot = kota_administrasi_bar_pipeline.hvplot(kind='bar',
                    x='kota_administrasi',
                    y=yaxis_kota_administrasi,
                    title='Korban Terdampak berdasarkan Kota
Administrasi')
kota_administrasi_bar_plot

kota_administrasi_tabel = kota_administrasi_bar_pipeline.pipe(pn.widgets.Tabulator,
pagination='remote', page_size=10)
kota_administrasi_tabel

# Radio buttons for CO2 measures
yaxis_korbanjiwa = pn.widgets.RadioButtonGroup(
    name='Y axis',
    options=['jumlah_terdampak_jiwa', 'jumlah_terdampak_kk'],
    button_type='success'
)

kecamatan = df['kecamatan'].unique()

korbanjiwa_pipeline = (
    idf[
        (idf.bulan_kejadian <= month_slider) &
        (idf.kecamatan.isin(kecamatan))
    ]
    .groupby(['kecamatan', 'bulan_kejadian'])[yaxis_korbanjiwa].mean()
    .to_frame()
    .reset_index()
    .sort_values(by='bulan_kejadian')
    .reset_index(drop=True)
)
korbanjiwa_pipeline

bulan_plot = korbanjiwa_pipeline.hvplot(x = 'bulan_kejadian', by='kecamatan',
y=yaxis_korbanjiwa, line_width=2, title="Korban Terdampak Berdasarkan
kecamatan")
bulan_plot

month_itable = korbanjiwa_pipeline.pipe(pn.widgets.Tabulator, pagination='remote',
page_size=10)
month_itable

# Mengubah peta Folium menjadi HTML
html_map = m._repr_html_()

```

```

# Membuat komponen Panel untuk menampilkan peta interaktif
peta_interaktif = pn.pane.HTML(html_map, width=1000, height=1000)

import panel as pn

# Membuat template FastListTemplate
template = pn.template.FastListTemplate(
    title='Daerah Rawan Banjir di Provinsi DKI Jakarta',
    sidebar=[pn.pane.Markdown("# Daerah Rawan Banjir di Provinsi DKI Jakarta"),
             pn.pane.Markdown("####Kebanjiran adalah fenomena bencana yang perlu
diwaspadai karena frekuensi yang sering bahkan hingga menimbulkan kerugian yang
besar"),
             pn.pane.Markdown("## Settings"),
             month_slider],
    main=[pn.Row(pn.Column(yaxis_kota_administrasi,
                           kota_administrasi_bar_plot.panel(width=600)),
                 kota_administrasi_tabel.panel(width=300)),
          pn.Row(pn.Column(yaxis_korbanjiwa,
                           bulan_plot.panel(width=1000,height=400), margin=(0,25))),peta_interaktif],
    accent_base_color="#88d8b0",
    header_background="#88d8b0",
)
# Menampilkan template
template.show()

```

Gambar 16 Kode Script Penelitian

5. BAB V: KESIMPULAN

Klasterisasi merupakan sebuah metode yang digunakan untuk mengelompokkan data. Pada penelitian kali ini, klasterisasi dilakukan untuk mengetahui pengelompokan wilayah rawan banjir di DKI Jakarta. Project Based Learning ini menggunakan data kejadian banjir di DKI Jakarta pada tahun 2020 yang bersumber dari Jakarta Open Data dimana dataset yang digunakan sebanyak 12 data berdasarkan bulan yang kemudian digabungkan menjadi satu dengan jumlah total data sebanyak 1006 data. Kemudian data diolah melalui tahapan preprocessing seperti menghilangkan duplikasi data, menghapus missing value, mengubah format data, mengganti data dengan kesalahan penulisan, dan melakukan reset index. Setelah melalui tahap preprocecssing, penelitian ini selanjutnya melakukan feature extraction, Explaratory Data Analysis, Pemodelan menggunakan metode hierarchical clustering, evaluasi, visualisasi, dan dashboard. Berdasarkan pemodelan didapatkan hasil klasterisasi berupa sebanyak 37 data menyatakan bahwa kecamatan tersebut merupakan daerah sangat tidak rawan banjir, sebanyak 7 data menyatakan bahwa kecamatan tersebut merupakan daerah tidak rawan banjir, sebanyak 1 data menyatakan bahwa kecamatan tersebut merupakan daerah rawan banjir, dan sebanyak 1 data menyatakan bahwa kecamatan tersebut merupakan daerah sangat rawan banjir. Penelitian ini memiliki nilai evaluasi performa sebesar $0,6875167194435193$ yang artinya klasterisasi berhasil dengan baik dalam mengidentifikasi dan mengelompokkan daerah-daerah yang rawan banjir di DKI Jakarta. Pada penelitian ini menggunakan dashboard guna memudahkan dalam menyampaikan hasil analisis yang telah dilakukan berupa visualisasi data peta yang menggambarkan hasil klasterisasi daerah rawan banjir di DKI Jakarta.

6. DAFTAR PUSTAKA

- Aldimasqie, A. M., Saputra, A. H., & Oktarina, S. (2022, Oktober). Pemetaan Zona Rawan Banjir di Jakarta Menggunakan Analytic Hierarchy Process (AHP). *Jurnal Environmental Science*, 5(1), 1-14. Retrieved Mei 2023
- Dewi, D. I., & Pramita, D. K. (2019, November). Analisis Perbandingan Metode Elbow dan Sillhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali. *Jurnal Matrix*, 9(2), 102-109. Retrieved Mei 2023
- Dwi, A., & Nurhadi. (2022, Oktober 18). *Ini 10 Banjir Terparah di Jakarta Sepanjang Sejarah*. (tempo.com) Retrieved Mei 23, 2023, from <https://metro.tempo.co/read/1646650/ini-10-banjir-terparah-di-jakarta-sepanjang-sejarah>
- Hakim, R. N. (2022, Oktober 7). *Banjir Jakarta yang Kembali Telan Korban Jiwa*. (Kompas.com) Retrieved Mei 23, 2023, from <https://megapolitan.kompas.com/read/2022/10/07/07080311/banjir-jakarta-yang-kembali-telan-korban-jiwa?page=all>
- Islamy, U., Nursidah, D. R., Narendra, I. S., Anshori, M., & Widodo, E. (2022). Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Dampak Bencana Banjir Tahun 2017-2020 Menggunakan K-Medoids. *Bimaster (Buletin Ilmiah Math. Stat. dan Terapannya)*, 11(2), 381-388. Retrieved Mei 2023
- Mustika, Ardilla, Y., Manuhutu, A., Ahmad, N., Hasbi, I., Guntoro, . . . Ernawati, L. (2021). *Data Mining dan Aplikasinya*. Bandung: WIDINA BHAKTI PERSADA BANDUNG.
- Okta, E. V., Widyatama, A. A., & Utomo, B. (2022). Prediksi Penurunan Muka Air Tanah terhadap Banjir Rob di Kawasan Pesisir Kota Jakarta. *Jurnal Penataan Ruang*, 17(2), 98-109. Retrieved Mei 2023
- perkim. (2020, Juni 12). *Profil PKP Daerah Khusus Ibukota Jakarta*. Retrieved Mei 23, 2023, from perkim.id: <https://perkim.id/profil-pkp/profil-provinsi/profil-perumahan-dan-kawasan-permukiman-daerah-khusus-ibukota-jakarta/>
- Pratikto, R. O., & Damastuti, N. (2021). Klasterisasi Menggunakan Agglomerative Hierarchical Clustering Untuk Memodelkan Wilayah Banjir. *JOINTECS (Journal of Information Technology and Computer Science)*, 6(1), 13-20. Retrieved Mei 2023

- RI, B. (n.d.). *Pemerintah Provinsi DKI Jakarta*. (BPK Perwakilan Provinsi DKI Jakarta) Retrieved Mei 23, 2023, from jakarta.bpk.go.id: <https://jakarta.bpk.go.id/pemerintah-provinsi-dki-jakarta/>
- Supianto, A. (2014). *Pengenalan Pola Hierarchical Clustering*. Retrieved from <http://afif.lecture.ub.ac.id/files/2014/05/Slide-12-Klasterisasi-Hierarchical-Clustering.pdf>
- Unit Pengelola Statistik Dinas Komunikasi, I. d. (2021). *Statistik Sektoral Pemerintah Provinsi DKI Jakarta Tahun 2021 Volume 3 Tahun 2021*. DKI Jakarta: Unit Pengelola Statistik Dinas Komunikasi, Informatika dan Statistik Provinsi DKI Jakarta.
- Wardy, D. K., Putra, I. G., & Rusjayanthi, N. D. (2022). Clustering Artikel pada Portal Berita Online Menggunakan Metode K-Means. *JITTER - Jurnal Ilmiah Teknologi dan Komputer*, 3(1).

FINAL

7. LAMPIRAN

Dataset Bulan Januari

januari

kota_administrasi	kecamatan	kelurahan	nr	jumlah_terdamp	jumlah_terdamp	jumlah_terdamp	jumlah_terdamp	jumlah_ketingilan_air	tanggal_kejadian_lama	genangan_jumlah_meningg	jumlah_hilang	jumlah_luka_ber_jumlah_luka_rin	jumlah_pengung	jumlah_tempat_j	nilai_kerugian	
Jakarta Pusat	JOHAR BARU	JOHAR BARU	RV 01, 02, 06, C	4	4	0	35	140 10 s/d 70 cm	tgl 01 Januari	0	0	0	140	1	0	0
Jakarta Pusat	KEMAYORAN	GUNUNG SAHAH	RV 01, 02, 07, C	4	4	0	0	0 10 s/d 70 cm	tgl 01 Januari	0	0	0	0	0	0	0
Jakarta Pusat	KEMAYORAN	SERDANG	RV 01, 02, 06, C	4	4	0	0	0 10 s/d 30 cm	tgl 01 Januari	1	0	0	0	0	0	0
Jakarta Pusat	SAWAH BESAR PASAR BARA	RV 02, 03, 04, C	7	7	155	625	10 s/d 70 cm	tgl 01 Januari	0	0	0	625	5	0	0	
Jakarta Pusat	TANAH ABANG KARET TENGS	RV 02, 03, 04, C	2	2	52	195	10 s/d 70 cm	tgl 01 Januari	0	0	0	195	6	0	0	
Jakarta Pusat	TANAH ABANG PETAMBURAN	RV 02, 03, 04, C	2	2	0	0	0 40 s/d 60 cm	tgl 01 Januari	0	0	0	0	2070	8	0	
Jakarta Utara	CILINCING	CILINCING	RV 05, 06, 09, C	3	21	280	1121	10 s/d 35 cm	tgl 01 Januari	0	0	0	1121	14	0	0
Jakarta Utara	CILINCING	MARUNDA	RV 04, C	1	1	0	0	0 10 s/d 30 cm	tgl 01 Januari	0	0	0	0	0	0	0
Jakarta Utara	CILINCING	ROTOWAN	RV 01, 02, 03, C	5	5	36	150	10 s/d 150 cm	tgl 01, 24 Januari	1	0	0	0	338	3	0
Jakarta Utara	CILINCING	SEMPER BARA	RV 01, 02, 04, C	5	5	143	574	10 s/d 70 cm	tgl 01 Januari	0	0	0	574	7	0	0
Jakarta Utara	CILINCING	SEMPER TIMUR	RV 02, 03, 10, 1	4	6	30	120	10 s/d 150 cm	tgl 01 Januari	0	0	0	120	4	0	0
Jakarta Utara	KELAPA GADING BARAT	RV 02, 03, 09, C	9	9	70	278	10 s/d 150 cm	tgl 01 Januari	0	0	0	278	2	0	0	
Jakarta Utara	KELAPA GADING KELAPA GADING	RV 02, 03, 06, 1	5	5	0	0	0 10 s/d 30 cm	tgl 01 Januari	0	0	0	0	0	0	0	
Jakarta Utara	KELAPA GADING KELAPA GADING	RV 02, 03, 06, 16	2	2	25	105	10 s/d 150 cm	tgl 01 Januari	0	0	0	105	2	0	0	
Jakarta Utara	KELAPA GADING KELAPA GADING	RV 02, 03, 04, C	7	7	52	210	10 s/d 70 cm	tgl 01 Januari	0	0	0	210	3	0	0	
Jakarta Utara	KOJA	TUGU SELATAN	RV 04	1	1	0	0	0 20 s/d 40 cm	tgl 01 Januari	0	0	0	0	0	0	0
Jakarta Utara	PENJARINGAN KAPUK MUARA	RV 01, 02, 03, I	6	6	0	0	0 10 s/d 30 cm	tgl 01 Januari	0	0	0	0	0	0	0	
Jakarta Utara	PENJARINGAN PEAGALGAN	RV 08, 09, 10, I	7	7	292	1168	10 s/d 150 cm	tgl 01 Januari	0	0	0	1168	10	0	0	
Jakarta Utara	PENJARINGAN PENJARINGAN	RV 17	1	1	0	0	0 10 s/d 30 cm	tgl 01 Januari	0	0	0	0	0	0	0	
Jakarta Utara	PENJARINGAN PLUIT	RV 01, 11, 22	3	3	0	0	0 15 s/d 70 cm	tgl 24, 27, 28, 3	0	0	0	0	0	0	0	
Jakarta Utara	PADEMANANG ANCOL	RV 02, 05, 08, 1	4	4	0	0	0 10 s/d 30 cm	tgl 24 Januari	0	0	0	0	0	0	0	
Jakarta Barat	CENGKARENG CENGKARENG	RV 01, 03, 05, C	9	9	550	2220	31 s/d 70 cm	tgl 01, 24 Januari	0	0	0	2220	20	0	0	
Jakarta Barat	CENGKARENG CENGKARENG	RV 02, 03, 04, C	9	9	90	362	10 s/d 70 cm	tgl 01 Januari	1	0	0	362	6	0	0	
Jakarta Barat	CENGKARENG DUS SOSAMBHU	RV 01, 02, 03, C	15	15	307	120	71 s/d 150 cm	tgl 01 Januari	0	0	0	120	8	0	0	
Jakarta Barat	CENGKARENG KAPUK	RV 03, 04, C	13	13	1115	4401	31 s/d 70 cm	tgl 01, 24 Januari	1	0	0	4401	9	0	0	
Jakarta Barat	CENGKARENG KEDAUJA KALIW	RV 01, 02, 03, C	8	8	701	2804	31 s/d 70 cm	tgl 01 Januari	0	0	0	2804	31	0	0	
Jakarta Barat	CENGKARENG RAVIA RAYUA	RV 01, 02, 03, C	6	6	533	2133	31 s/d 70 cm	tgl 01 Januari	0	0	0	2133	7	0	0	
Jakarta Barat	GROGOL PETA JELAMBAR	RV 01, 02, 03, C	13	13	15	60	31 s/d 70 cm	tgl 01 Januari	0	0	0	60	1	0	0	
Jakarta Barat	GROGOL PETA TANJUNG DURI	RV 01, 02, 04, C	5	5	21	85	10 s/d 150 cm	tgl 01 Januari	0	0	0	85	3	0	0	
Jakarta Barat	GROGOL PETA TOMANG	RV 14	1	1	6	24	31 s/d 70 cm	tgl 01 Januari	0	0	0	24	1	0	0	
Jakarta Barat	KALIDERES KALIDERES	RV 04, 05, 06, C	12	12	37	208	31 s/d 70 cm	tgl 01 Januari	0	0	0	208	3	0	0	
Jakarta Barat	KALIDERES KALIDERES	RV 01, 02, 03, C	7	7	0	0	0 10 s/d 70 cm	tgl 01 Januari	0	0	0	0	147	1	0	
Jakarta Barat	KALIDERES PEGADUNGAN	RV 01, 03, 04, C	5	5	0	0	0 10 s/d 70 cm	tgl 01 Januari	0	0	0	0	0	0	0	

Dataset Bulan Februari

februari

kota_administrasi	kecamatan	kelurahan	nr	jumlah_terdamp	jumlah_terdamp	jumlah_terdamp	jumlah_terdamp	jumlah_ketingilan_air	tanggal_kejadian_lama	genangan_jumlah_meningg	jumlah_hilang	jumlah_luka_ber_jumlah_luka_rin	jumlah_pengung	jumlah_tempat_j	nilai_kerugian	
Jakarta Pusat	JOHAR BARU	JOHAR BARU	RV 05, 09, C	2	2	0	0	0 30 s/d 50 cm	23, 25 februari	0	0	0	0	0	0	0
Jakarta Pusat	KEMAYORAN	KEMAYORAN	RV 04, 09, C	2	2	0	0	0 20 s/d 100 cm	2, 8, 23 februari	0	0	0	0	0	0	0
Jakarta Pusat	KEMAYORAN	SERDANG	RV 06, 01	2	2	0	0	0 30 cm	23 februari	0	0	0	0	0	0	0
Jakarta Pusat	KEMAYORAN	KEBON BARU	RV 05, 07	2	7	0	0	0 20 s/d 50 cm	23 februari	0	0	0	0	0	0	0
Jakarta Pusat	KEMAYORAN	KEMAYORAN	RV 06, 08, 09	1	1	0	0	0 20 s/d 50 cm	23 februari	0	0	0	0	0	0	0
Jakarta Pusat	KEMAYORAN	KEBON KOSONG	RV 01, 02, 03, 04	10	10	0	0	0 10 s/d 50 cm	25 februari	0	0	0	0	0	0	0
Jakarta Pusat	KEMAYORAN	SUMIR BATU	RV 07	1	1	0	0	0 10 s/d 50 cm	25 februari	0	0	0	0	0	0	0
Jakarta Pusat	SAWAH BESAR PASAR BARU	RV 03, 04, 05, 0	4	6	0	0	0 10 s/d 40 cm	B, 21, 22, 23 feb	0	0	0	450	1	0	0	
Jakarta Pusat	SAWAH BESAR KARANG ANYA	RV 06	1	1	0	0	0 20 cm	23 februari	0	0	0	0	0	0	0	
Jakarta Pusat	SAWAH BESAR MANGGA DUA	RV 11	1	1	0	0	0 30 cm	25 februari	0	0	0	0	0	0	0	
Jakarta Pusat	SAWAH BESAR SUNGAI SAHA	RV 03, 07	2	2	0	0	0 30 cm	25 februari	0	0	0	0	0	0	0	
Jakarta Pusat	TANAH ABANG BENDUNGJAWA	RV 01, 03, 04, 06	5	12	0	0	0 10 s/d 60 cm	8, 23, 25 februari	0	0	0	475	2	0	0	
Jakarta Pusat	TANAH ABANG KETENGKENG	RV 07, 09	2	2	0	0	0 40 s/d 80 cm	22, 23 februari	0	0	0	1368	6	0	0	
Jakarta Pusat	TANAH ABANG PETAMBURAN	RV 03, 05, 09	2	10	0	0	0 10 s/d 60 cm	22, 23 feb	2	0	0	0	0	0	0	
Jakarta Pusat	TANAH ABANG KEBON MELETU	RV 07, 16	2	2	0	0	0 35 s/d 50 cm	23, 25 februari	0	0	0	0	0	0	0	
Jakarta Pusat	SENEN KRAMAT	RV 04, 06, 08, 0	3	3	0	0	0 10 s/d 50 cm	25, 23, 25 feb	0	0	0	0	0	0	0	
Jakarta Pusat	SENEN KENARI	RV 05	1	1	0	0	0 20 cm	23 februari	0	0	0	0	0	0	0	
Jakarta Pusat	MENTENG KEBON SIRIH	RV 02, 04, 08	3	3	0	0	0 20 s/d 90 cm	20, 23 februari	0	0	0	50	1	0	0	
Jakarta Pusat	MENTENG MENTENG	RV 05, 09, 11	3	3	0	0	0 10 s/d 60 cm	23, 25 februari	0	0	0	0	0	0	0	
Jakarta Pusat	CEMPcosa PUTI RAWASARI	RV 09	1	2	0	0	0 20 s/d 30 cm	23, 25 februari	0	0	0	0	0	0	0	
Jakarta Pusat	CEMPcosa PUTI CEMPcosa PUTI	RV 05, 06, 07, 08	4	4	0	0	0 30 cm	23, 25 februari	0	0	0	0	0	0	0	
Jakarta Pusat	CEMPcosa PUTI CEMPcosa PUTI	RV 03, 05, 08, 12	3	3	0	0	0 20 s/d 30 cm	08, 25 februari	0	0	0	0	0	0	0	
Jakarta Pusat	TANJUNG PRIO KEBON BAWAN	RV 02, 07	2	2	0	0	0 10 s/d 30 cm	20 feb	0	0	0	0	0	0	0	
Jakarta Pusat	TANJUNG PRIO SUNTER JAYA	RV 02, 04, 12	3	3	0	0	0 10 s/d 45 cm	8, 23 februari	0	0	0	0	0	0	0	
Jakarta Utara	TANJUNG PRIO WARAKAS	RV 01, 12, 13	3	3	0	0	0 20 cm	08, 25 februari	0	0	0	0	0	0	0	
Jakarta Utara	TANJUNG PRIO TANJUNG PRIO	RV 02, 07	2	2	0	0	0 10 s/d 30 cm	8 februari	0	0	0	0	0	0	0	
Jakarta Utara	TANJUNG PRIO KEBON BAWAN	RV 02, 07	2	2	0	0	0 10 s/d 30 cm	8 februari	0	0	0	0	0	0	0	
Jakarta Utara	KELAPA GADING KELAPA GADING	RV 01, 04, 05, 06	13	13	65	271	30 s/d 40 cm	0, 23, 25 februar	0	0	0	271	6	0	0	
Jakarta Utara	KELAPA GADING PEGANGSAMAN	RV 02, 03, 04, 06	9	9	44	182	10 s/d 70 cm	23 februari	0	0	0	182	3	0	0	
Jakarta Utara	KOJA TUGU UTARA	RV 02, 03, 04, 1C	9	12	22	192	10 s/d 80 cm	0, 23 februari	0	0	0	138	1	0	0	
Jakarta Utara	KOJA TUGU UTARA	RV 04, 08, 09, 11	4	4	0	0	0 15 s/d 60 cm	8, 23 februari	0	0	0	0	0	0	0	

Dataset Bulan Maret

Screenshot of an Excel spreadsheet titled "maret" showing data for March 2023. The table has 30 rows and 24 columns. The columns are labeled A through S.

kota_administrasi	kecamatan	kelurahan	nr		jumlah_terdamp_jumlah_terdamp_jumlah_terdamp_jumlah_terdamp_jumlah_ketinggian_air	tanggal_kejadian_lama_nenangan_jumlah_meningg_jumlah_hilang	jumlah_luka_berjumlah_luka_rina	jumlah_pengung_jumlah_tempat_nilar_kerugian
2 Jakarta Utara	CILINCING	ROROTAN	rv 04		1 1 0 0 30 cm	2. 18 maret	0 0 0 0 0	0 0 0 0 0
3 Jakarta Utara	PEKALONGAN	SELAHARU ADI	rv 01		1 1 0 0 10 cm	2. 19 maret	0 0 0 0 0	0 0 0 0 0
4 Jakarta Barat	KEMBANGAN	KEMBANGAN 5	nr 04, 09		3 6 0 0 20 s/d 40 cm	1. 2, 20 maret	0 0 0 0 0	0 0 0 0 0
5 Jakarta Barat	KEMBANGAN	KEMBANGAN 1	nr 04		1 1 0 0 15 s/d 35 cm	1. 2, 20 maret	0 0 0 0 0	0 0 0 0 0
6 Jakarta Barat	KEBON JERUK	KEDOYA SELAT	rv 03, 05		2 4 0 0 10 s/d 30 cm	2. 20 maret	0 0 0 0 0	0 0 0 0 0
7 Jakarta Barat	CENGKARENG RAVIA BLUYA	rv 01, 02, 03, 08	5		0 0 0 0 10 s/d 50 cm	22 Maret	0 0 0 0 0	0 0 0 0 0
8 Jakarta Barat	CENGKARENG	CENGKARENG	rv 03, 04, 06, 08		5 12 0 0 5 s/d 30 cm	22 Maret	0 0 0 0 0	0 0 0 0 0
9 Jakarta Barat	KALIDERES	KALIDERES	rv 03		1 1 0 0 15 cm	22 Maret	0 0 0 0 0	0 0 0 0 0
10 Jakarta Barat	KALIDERES	SEMANAN	rv 01, 03, 04, 1C		4 11 0 0 5 s/d 30 cm	22 Maret	0 0 0 0 0	0 0 0 0 0
11 Jakarta Barat	PALMERAH	KOTA BAMBU	rv 04		1 1 0 0 10 cm	22 Maret	0 0 0 0 0	0 0 0 0 0
12 Jakarta Barat	PALMERAH	PALMERAH	rv 14		1 1 0 0 20 cm	22 Maret	0 0 0 0 0	0 0 0 0 0
13 Jakarta Selatan	PASAR MINGGI CILANDAK TIM	rv 05, 06, 07, 08	5		1 1 0 0 10 s/d 15 cm	2. 19, 20, 26, 27	0 0 0 0 0	0 0 0 0 0
14 Jakarta Selatan	PASAR MINGGI CILANDAK TIM	rv 05, 06, 07, 13	2		0 0 0 0 10 s/d 100 cm	18, 19, 20, 26, 27	0 0 0 0 0	0 0 0 0 0
15 Jakarta Selatan	PANCORAN	PENGADEGAN	rv 01		1 1 0 0 90 cm	1. 28 maret	0 0 0 0 0	0 0 0 0 0
16 Jakarta Selatan	PANCORAN	BAWAJATI	rv 07		1 3 0 0 10 s/d 40 cm	1. 27 maret	0 0 0 0 0	0 0 0 0 0
17 Jakarta Selatan	PANCORAN	DUREN TIGA	rv 05		1 1 0 0 60 cm	20 Maret	0 0 0 0 0	0 0 0 0 0
18 Jakarta Selatan	TEBET	KEBON BARU	rv 05, 06, 10		3 3 0 0 5 s/d 60 cm	1. 16, 20 maret	0 0 0 0 0	0 0 0 0 0
19 Jakarta Selatan	TEBET	BUKIT DURI	rv 12		1 1 0 0 80 cm	1 Maret	0 0 0 0 0	0 0 0 0 0
20 Jakarta Selatan	KEBAYORAN L/ CIPULIR	rv 10			0 0 0 0 20 s/d 100 cm	2. 16, 18, 20, 28	0 0 0 0 0	0 0 0 0 0
21 Jakarta Selatan	KEBAYORAN L/ CIPULIR	rv 10			1 2 0 0 10 s/d 115 cm	2. 16, 17, 20, 28	0 0 0 0 0	0 0 0 0 0
22 Jakarta Selatan	PESANGRAN BINTARO	rv 05, 11, 10, 14	4		0 0 0 0 10 s/d 15 cm	2. 16, 20 maret	0 0 0 0 0	0 0 0 0 0
23 Jakarta Selatan	PESANGRAN PUS	rv 01, 02, 03, 04	1		0 0 0 0 10 cm	0 0 0 0 0	0 0 0 0 0	0 0 0 0 0
24 Jakarta Selatan	CILANDAK	CIPETE SELAT	rv 07		1 1 0 0 15 cm	16 Maret	0 0 0 0 0	0 0 0 0 0
25 Jakarta Selatan	CILANDAK	PONDOK LABU	rv 01, 03, 07, 1C		4 7 0 0 10 s/d 50 cm	16 Maret	0 0 0 0 0	0 0 0 0 0
26 Jakarta Selatan	MAMPANG PRA BANGKA	rv 02, 03, 04	3		0 0 0 0 20 s/d 40 cm	16 Maret	0 0 0 0 0	0 0 0 0 0
27 Jakarta Selatan	MAMPANG PRA KUNINGAN BAR	rv 05	1		0 0 0 0 40 cm	16 Maret	0 0 0 0 0	0 0 0 0 0
28 Jakarta Selatan	JAGAKARTA	CIPERAD	rv 04		1 1 0 0 70 cm	16 Maret	0 0 0 0 0	0 0 0 0 0
29 Jakarta Selatan	SETIABUDI	MENTENG ATA	rv 07, 08, 09, 14		5 5 0 0 10 s/d 60 cm	20 Maret	0 0 0 0 0	0 0 0 0 0
30 Jakarta Timur	JATINEGARA	CIPINANG MUA	rv 13		1 2 118 0 456 s/d 50 cm	1. 2, 16 maret	0 0 0 0 65	1 0 0 0 0
31 Jakarta Timur	JATINEGARA	KAMPUNG MELAYU	rv 04, 05, 07, 08		4 25 326 0 1068 s/d 110 cm	1. 2, 16, 18, 23	0 0 0 0 186	3 0 0 0 0
32 Jakarta Timur	JATINEGARA	KAMPUNG MELAYU	rv 04, 05, 06, 07, 09		5 12 1711 0 4300 s/d 140 cm	1. 2, 16, 18, 23	0 0 0 0 265	5 0 0 0 0
33 Jakarta Timur	KRAMAT JATI	CAHAYA	rv 01, 02, 03, 05		6 15 0 0 8 s/d 115 cm	1. 2, 16, 23, 25	0 0 0 0 0	0 0 0 0 0
34 Jakarta Timur	KRAMAT JATI	CILILITAN	rv 06, 07		2 6 10 0 43 s/d 100 cm	1. 16, 25, 28, ma	0 0 0 0 0	0 0 0 0 0
35 Jakarta Timur	KRAMAT JATI	TENGAH	rv 05		1 1 0 0 70 cm	16 Maret	0 0 0 0 0	0 0 0 0 0
36 Jakarta Timur	CIPAYUNG	LUBANG BUAYA	rv 12		1 1 0 0 20 cm	1 Maret	0 0 0 0 0	0 0 0 0 0

Dataset Bulan April

Screenshot of an Excel spreadsheet titled "april" showing data for April 2023. The table has 30 rows and 24 columns. The columns are labeled A through S.

kota_administrasi	kecamatan	kelurahan	nr		jumlah_terdamp_jumlah_terdamp_jumlah_terdamp_jumlah_terdamp_jumlah_ketinggian_air	tanggal_kejadian_lama_nenangan_jumlah_meningg_jumlah_hilang	jumlah_luka_berjumlah_luka_rina	jumlah_pengung_jumlah_tempat_nilar_kerugian
2 Jakarta Utara	PEREMPUAN	KRADIO	rv 01		1 3 0 0 10 s/d 30 cm	tgl 15 April	0 0 0 0 0	0 0 0 0 0
3 Jakarta Utara	REBOZ JERUK SUKABUMI SELAT	rv 01, 02, 04, C	6 14 0 0 5 s/d 90 cm	tgl 16 April	0 0 0 0 0	0 0 0 0 0		
4 Jakarta Barat	KEBON JERUK KEDOYA SELAT	rv 03, 04	2 6 0 0 10 s/d 30 cm	tgl 17 April	0 0 0 0 0	0 0 0 0 0		
5 Jakarta Barat	CENGKARENG	CENGKARENG RV 01	1 1 0 0 15 cm	tgl 06 April	0 0 0 0 0	0 0 0 0 0		
6 Jakarta Barat	CENGKARENG RAVIA BLUYA	rv 02, 04, 08	3 6 0 0 20 cm	tgl 06, 27 April	0 0 0 0 0	0 0 0 0 0		
7 Jakarta Barat	KEMBANGAN	KEMBANGAN 5	rv 01, 04, 05, C	4 5 20 0 65, 10 s/d 60 cm	tgl 06, 27 April	0 0 0 0 65	1 0 0 0 0	
8 Jakarta Barat	KEMBANGAN	GRENGSEN	rv 05, 08	2 4 0 0 15 s/d 60 cm	tgl 06 April	0 0 0 0 0	0 0 0 0 0	
9 Jakarta Barat	KEMBANGAN	KEMBANGAN 1	rv 01, 05	2 2 0 0 50 cm	tgl 06, 27 April	0 0 0 0 0	0 0 0 0 0	
10 Jakarta Selatan	JAGAKARTA	CIPERAD	rv 04, 05	2 2 0 0 50 cm	tgl 05 April	0 0 0 0 0	0 0 0 0 0	
11 Jakarta Selatan	JAGAKARTA	JAGAKARTA	rv 03	1 1 0 0 40 cm	tgl 10 April	0 0 0 0 0	0 0 0 0 0	
12 Jakarta Selatan	PASAR MINGGI CILANDAK TIM	rv 01, 04, 05, 07	2 3 0 0 10 s/d 80 cm	tgl 11, 12, 13, 2, 3	0 0 0 0 0	0 0 0 0 0		
13 Jakarta Selatan	PASAR MINGGI RAGUNGAN	rv 09	1 1 0 0 30 cm	tgl 13 April	0 0 0 0 0	0 0 0 0 0		
14 Jakarta Selatan	PASAR MINGGI PEATEN BARA	rv 04, 07, 08	3 12 0 0 10 s/d 30 cm	tgl 10, 19, 20 Ap	0 0 0 0 0	0 0 0 0 0		
15 Jakarta Selatan	PASAR MINGGI PADANG	rv 04, 05, 06, 07	2 4 0 0 30 s/d 40 cm	tgl 06, 26, 29 Ap	0 0 0 0 0	0 0 0 0 0		
16 Jakarta Selatan	KEBAYORAN L/ PONDOK PINA	rv 05, 08	2 2 0 0 30 cm	tgl 30 April	0 0 0 0 0	0 0 0 0 0		
17 Jakarta Selatan	KEBAYORAN L/ KEBAYORAN L/ CIPULIR	rv 03, 04	2 2 0 0 15 s/d 30 cm	tgl 06, 15, 25, Ap	0 0 0 0 0	0 0 0 0 0		
18 Jakarta Selatan	TEBET	KEBON BARU	rv 05, 06, 010	4 6 0 0 15 s/d 30 cm	tgl 29 April	0 0 0 0 0	0 0 0 0 0	
19 Jakarta Selatan	TEBET	TERBEK BARAT	rv 07	1 1 0 0 40 cm	tgl 06 April	0 0 0 0 0	0 0 0 0 0	
20 Jakarta Selatan	MAMPANG PRERA BANGKA	rv 01, 02, 03, C	2 2 0 0 10 s/d 80 cm	tgl 06 April	0 0 0 0 0	0 0 0 0 0		
21 Jakarta Selatan	MAMPANG PRERA	MAMPANG PRERA MAMPA	rv 01, 02, 03	4 6 0 0 20 s/d 50 cm	tgl 27 April	0 0 0 0 0	0 0 0 0 0	
22 Jakarta Selatan	CILANDAK	LEBAK BULUS	rv 04, 08	2 2 3 0 10, 50 s/d 70 cm	tgl 26 April	0 0 0 0 0	0 0 0 0 0	
23 Jakarta Selatan	CILANDAK	CIPETE SELAT	rv 07	1 1 0 0 50 cm	tgl 27 April	0 0 0 0 0	0 0 0 0 0	
24 Jakarta Selatan	CILANDAK	GANDARIA SELAT	rv 07	1 2 0 0 30 s/d 80 cm	tgl 27 April	0 0 0 0 0	0 0 0 0 0	
25 Jakarta Selatan	PANCORAN	DUREN TIGA	rv 05	1 1 0 0 20 s/d 50 cm	tgl 29, 30 April	0 0 0 0 0	0 0 0 0 0	
26 Jakarta Selatan	PANCORAN	GANDARIA	rv 04	1 1 0 0 15 cm	tgl 29 April	0 0 0 0 0	0 0 0 0 0	
27 Jakarta Selatan	PANCORAN	KALIBATA	rv 04, 08	2 2 0 0 10 s/d 20 cm	tgl 29, 30 April	0 0 0 0 0	0 0 0 0 0	
28 Jakarta Selatan	PANCORAN	KALIBATA II	rv 04	1 2 0 0 30 cm	tgl 29 April	0 0 0 0 0	0 0 0 0 0	
29 Jakarta Selatan	PESAWAR BHAYANGKARA	rv 01, 04, 05, C	4 7 0 0 60 cm	tgl 29 April	0 0 0 0 0	0 0 0 0 0		
30 Jakarta Selatan	KEBAYORAN BI	GANDARA UTRAL	rv 02	1 2 0 0 60 cm	tgl 29 April	0 0 0 0 0	0 0 0 0 0	
31 Jakarta Selatan	KEBAYORAN BI	PETOOGAN	rv 03	1 1 0 0 60 cm	tgl 30 April	0 0 0 0 0	0 0 0 0 0	
32 Jakarta Timur	JATINEGARA	CIPINANG MUA	rv 03	1 2 0 0 10 s/d 35 cm	tgl 06 April	0 0 0 0 0	0 0 0 0 0	
33 Jakarta Timur	JATINEGARA	CIPINANG MUA	rv 08	2 4 49 0 153 s/d 130 cm	tgl 18, 19 April	0 0 0 0 0	0 0 0 0 0	
34 Jakarta Timur	JATINEGARA	BIDARA CINA	rv 07, 011	1 1 0 0 15 cm	tgl 20 April	0 0 0 0 0	0 0 0 0 0	
35 Jakarta Timur	JATINEGARA	CIPINANG CEM	rv 08	1 1 0 0 15 cm	tgl 20 April	0 0 0 0 0	0 0 0 0 0	

Dataset Bulan Mei

Screenshot of a Microsoft Excel spreadsheet titled "mei". The table has 24 columns labeled A through R. The first column, A, contains the header "kota_administrasi" and the data for 24 different locations. The second column, B, contains the header "kecamatan" and the data for the same 24 locations. The third column, C, contains the header "kelurahan" and the data for the same 24 locations. The fourth column, D, contains the header "nr" and the data for the same 24 locations. The fifth column, E, contains the header "jumlah_terdamp" and the data for the same 24 locations. The sixth column, F, contains the header "jumlah_terdamp" and the data for the same 24 locations. The seventh column, G, contains the header "jumlah_terdamp" and the data for the same 24 locations. The eighth column, H, contains the header "jumlah_terdamp" and the data for the same 24 locations. The ninth column, I, contains the header "jumlah_terdamp" and the data for the same 24 locations. The tenth column, J, contains the header "jumlah_terdamp" and the data for the same 24 locations. The eleventh column, K, contains the header "jumlah_terdamp" and the data for the same 24 locations. The twelfth column, L, contains the header "jumlah_terdamp" and the data for the same 24 locations. The thirteenth column, M, contains the header "jumlah_terdamp" and the data for the same 24 locations. The fourteenth column, N, contains the header "jumlah_terdamp" and the data for the same 24 locations. The fifteenth column, O, contains the header "jumlah_terdamp" and the data for the same 24 locations. The sixteenth column, P, contains the header "jumlah_terdamp" and the data for the same 24 locations. The seventeenth column, Q, contains the header "jumlah_terdamp" and the data for the same 24 locations. The eighteenth column, R, contains the header "jumlah_terdamp" and the data for the same 24 locations. The nineteenth column, S, contains the header "jumlah_terdamp" and the data for the same 24 locations.

Dataset Bulan Juni

Screenshot of a Microsoft Excel spreadsheet titled "juni". The table has 24 columns labeled A through R. The first column, A, contains the header "kota_administrasi" and the data for 24 different locations. The second column, B, contains the header "kecamatan" and the data for the same 24 locations. The third column, C, contains the header "kelurahan" and the data for the same 24 locations. The fourth column, D, contains the header "nr" and the data for the same 24 locations. The fifth column, E, contains the header "jumlah_terdamp" and the data for the same 24 locations. The sixth column, F, contains the header "jumlah_terdamp" and the data for the same 24 locations. The seventh column, G, contains the header "jumlah_terdamp" and the data for the same 24 locations. The eighth column, H, contains the header "jumlah_terdamp" and the data for the same 24 locations. The ninth column, I, contains the header "jumlah_terdamp" and the data for the same 24 locations. The tenth column, J, contains the header "jumlah_terdamp" and the data for the same 24 locations. The eleventh column, K, contains the header "jumlah_terdamp" and the data for the same 24 locations. The twelfth column, L, contains the header "jumlah_terdamp" and the data for the same 24 locations. The thirteenth column, M, contains the header "jumlah_terdamp" and the data for the same 24 locations. The fourteenth column, N, contains the header "jumlah_terdamp" and the data for the same 24 locations. The fifteenth column, O, contains the header "jumlah_terdamp" and the data for the same 24 locations. The sixteenth column, P, contains the header "jumlah_terdamp" and the data for the same 24 locations. The seventeenth column, Q, contains the header "jumlah_terdamp" and the data for the same 24 locations. The eighteenth column, R, contains the header "jumlah_terdamp" and the data for the same 24 locations. The nineteenth column, S, contains the header "jumlah_terdamp" and the data for the same 24 locations.

Dataset Bulan Juli

Dataset Bulan Agustus

Dataset Bulan September

Dataset Bulan Oktober

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	kota administrasi	Kecamatan	kelurahan	rw	jumlah_terdampak	jumlah_terdampak	jumlah_terdampak	jumlah_terdampak	jumlah_terdampak	jumlah_terdampak							
2	Jakarta Pusat	CEMPaka PUTI	CEMPAKA PUTRI	RV 68	1	1	0	0	0	15 cm	16 Oktober	0	0	0	0	0	0
3	Jakarta Pusat	JOHAR BARU	JOHAR BARU	RV 06	1	1	0	0	0	20 cm	16 Oktober	0	0	0	0	0	0
4	Jakarta Utara	PROKTOR	PROKTO	RV 13	1	1	0	0	0	30 cm	16 Oktober	0	0	0	0	0	0
5	Jakarta Utara	TANJUNG PRIO	TANJUNG BAWAH	RV 01	0	0	1	0	0	30 cm	16 Oktober	0	0	0	0	0	0
6	Jakarta Utara	CILINCING	ROTORAN	RV 10	1	1	0	0	0	15 cm	26 Oktober	0	0	0	0	0	0
7	Jakarta Utara	KOJA	RAWA BADAK	RV 04	1	1	0	0	0	10 sd 60 cm	26 Oktober	0	0	0	0	0	0
8	Jakarta Utara	KOJA	RAWA BADAK	RV 08	1	1	0	0	0	10 sd 80 cm	26 Oktober	0	0	0	0	0	0
9	Jakarta Utara	KOJA	RAWA BADAK	RV 09	1	1	0	0	0	10 sd 80 cm	26 Oktober	0	0	0	0	0	0
10	Jakarta Utara	KOJA	RAWA BADAK	RV 10	1	1	0	0	0	10 sd 80 cm	26 Oktober	0	0	0	0	0	0
11	Jakarta Utara	KOJA	RAWA BADAK	RV 11	1	1	0	0	0	10 sd 80 cm	26 Oktober	0	0	0	0	0	0
12	Jakarta Barat	KEBON JERUK	KEDOYA SELATAN	03	1	1	0	0	0	20 cm	16 Oktober	0	0	0	0	0	0
13	Jakarta Barat	KEBON JERUK	KEDOYA SELATAN	0	0	1	0	0	20 cm	16 Oktober	0	0	0	0	0	0	
14	Jakarta Barat	KEBON JERUK	KEDOYA SELATAN	04	0	0	1	0	0	15 cm	11 Oktober	0	0	0	0	0	0
15	Jakarta Barat	KEBON JERUK	KEDOYA SELATAN	04	1	1	0	0	0	25 cm	16 Oktober	0	0	0	0	0	0
16	Jakarta Barat	KEBON JERUK	KEDOYA SELATAN	05	1	1	0	0	0	10 sd 30 cm	05.09.11 Oktober	0	0	0	0	9	1
17	Jakarta Barat	KEBON JERUK	KEDOYA SELATAN	05	0	0	1	0	0	20 sd 70 cm	05.11 Oktober	0	0	0	0	0	0
18	Jakarta Barat	KEBON JERUK	KEDOYA SELATAN	05	0	0	1	0	0	60 cm	11 Oktober	0	0	0	0	0	0
19	Jakarta Barat	KEBON JERUK	KEDOYA SELATAN	05	0	0	1	0	0	10 sd 100 cm	05.10.11.25 Oktober	0	0	0	0	0	0
20	Jakarta Barat	KEBON JERUK	KEDOYA SELATAN	05	0	0	1	0	0	10 sd 60 cm	05.09.11.16.25 Oktober	0	0	0	0	0	0
21	Jakarta Barat	KEBON JERUK	KEDOYA SELATAN	05	0	0	1	0	0	60 cm	11 Oktober	0	0	0	0	6	0
22	Jakarta Barat	KEBON JERUK	KEDOYA SELATAN	05	0	0	1	0	0	100 cm	11 Oktober	0	0	0	0	4	0
23	Jakarta Barat	KEBON JERUK	KEDOYA SELATAN	05	0	0	1	0	0	70 cm	11 Oktober	0	0	0	0	5	0
24	Jakarta Barat	KEBON JERUK	KEDOYA SELATAN	10	1	1	0	0	0	100 cm	11 Oktober	0	0	0	0	0	0
25	Jakarta Barat	KEBON JERUK	KEDOYA SELATAN	10	1	1	0	0	0	60 cm	11 Oktober	0	0	0	0	0	0
26	Jakarta Barat	KEBON JERUK	KEDOYA UTARA	02	0	0	1	0	0	20 cm sd 60 cm	10.11 Oktober	0	0	0	0	0	0
27	Jakarta Barat	KEBON JERUK	KEDOYA UTARA	02	0	0	1	0	0	20 cm sd 60 cm	10.11 Oktober	0	0	0	0	0	0
28	Jakarta Barat	KEBON JERUK	KEDOYA UTARA	02	0	0	1	0	0	60 cm	10.11 Oktober	0	0	0	0	0	0
29	Jakarta Barat	KEBON JERUK	KEDOYA UTARA	02	0	0	1	0	0	10 sd 60 cm	09.10.11 Oktober	0	0	0	0	0	0
30	Jakarta Barat	KEBON JERUK	SUKABUMI SEL	RV 08	1	1	0	0	0	30 cm	10.11.15 Oktober	0	0	0	0	0	0
31	Jakarta Barat	KEBON JERUK	SUKABUMI UTA	RV 02	1	1	0	0	0	20 cm	16 Oktober	0	0	0	0	0	0
32	Jakarta Barat	KEBON JERUK	SUKABUMI UTA	RV 05	1	1	0	0	0	50 cm	16 Oktober	0	0	0	0	0	0
33	Jakarta Barat	KEBON JERUK	SUKABUMI UTA	RV 05	0	0	1	0	0	40 sd 100 cm	16.10 Oktober	0	0	0	0	0	0
34	Jakarta Barat	KEBON JERUK	KEBON JERUK	RV 04	1	1	0	0	0	30 cm	16 Oktober	0	0	0	0	0	0
35	Jakarta Barat	KEBON JERUK	KEBON JERUK	RV 06	1	1	0	0	0	30 cm	16 Oktober	0	0	0	0	0	0
36	Jakarta Barat	KEBON JERUK	KEBON JERUK	RV 10	1	1	0	0	0	50 cm	16 Oktober	0	0	0	0	0	0

Dataset Bulan November

Screenshot of a Google Sheets document titled "november". The sheet contains data from column A1 to S. The columns represent various parameters such as administrative unit (kota_administrasi), location (kecamatan), street (kelurahan), house number (nr), and specific measurements like jumlah_terdampak, jumlah_ketinggian_air, and dates (tanggal_kejadian). The data spans across various districts in Jakarta, including Utara, Selatan, and Timur, covering multiple streets and house numbers. The data is presented in a tabular format with rows corresponding to individual entries.

Dataset Bulan Desember

Screenshot of a Google Sheets document titled "desember". The sheet contains data from column A1 to S. The columns represent various parameters such as administrative unit (kota_administrasi), location (kecamatan), street (kelurahan), house number (nr), and specific measurements like jumlah_terdampak, jumlah_ketinggian_air, and dates (tanggal_kejadian). The data spans across various districts in Jakarta, including Barat, Selatan, and Timur, covering multiple streets and house numbers. The data is presented in a tabular format with rows corresponding to individual entries.

Link Video Pembahasan

<https://youtu.be/wLtP68DVK10>