

**IMPLEMENTASI METODE *K-MEANS* DALAM PENGKLUSTERAN
WILAYAH RAWAN BANJIR DI KABUPATEN LAMONGAN**

SKRIPSI

**Oleh:
ADI CANDRA ARYA KUSUMA
NIM. 19650042**



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2024**

**IMPLEMENTASI METODE *K-MEANS* DALAM PENGKLUSTERAN
WILAYAH RAWAN BANJIR DI KABUPATEN LAMONGAN**

SKRIPSI

**Diajukan kepada:
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Komputer (S. Kom)**

**Oleh:
ADI CANDRA ARYA KUSUMA
NIM. 19650042**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2024**

HALAMAN PERSETUJUAN

**IMPLEMENTASI METODE *K-MEANS* DALAM PENGLUSTERAN
WILAYAH RAWAN BANJIR DI KABUPATEN LAMONGAN**

SKRIPSI

Oleh:
ADI CANDRA ARYA KUSUMA
NIM. 19650042

Telah Diperiksa dan Disetujui untuk Diuji:
Tanggal: 01 Mei 2024

Pembimbing I,



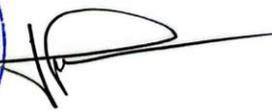
Dr. M. Faisal, M.T
NIP. 19740510 200501 1 007

Pembimbing II,



Dr. Fresy Nugroho, M.T
NIP. 19710722 201101 1 001

Mengetahui,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang



Dr. Fachrul Kurniawan, M.MT, IPM
NIP. 19771020 200912 1 001

HALAMAN PENGESAHAN

**IMPLEMENTASI METODE *K-MEANS* DALAM PENGKLUSTERAN
WILAYAH RAWAN BANJIR DI KABUPATEN LAMONGAN**

SKRIPSI

Oleh:
ADI CANDRA ARYA KUSUMA
NIM. 19650042

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Tanggal: 17 Mei 2024

Susunan Dewan Penguji

Ketua Penguji : Dr. Fachrul Kurniawan, M.MT, IPM
NIP. 19771020 200912 1 001

Anggota Penguji I : Hani Nurhayati, M.T
NIP. 19780625 200801 2 006

Anggota Penguji II : Dr. Muhammad Faisal, M.T
NIP. 19740510 200501 1 007

Anggota Penguji III : Dr. Fresy Nugroho, M.T
NIP. 19710722 201101 1 001



Mengetahui dan Mengesahkan,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang



Dr. Fachrul Kurniawan, M.MT, IPM
NIP. 19771020 200912 1 001

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Adi Candra Arya Kusuma
NIM : 19650042
Fakultas / Jurusan : Sains dan Teknologi / Teknik Informatika
Judul Skripsi : *Implementasi Metode K-Means* dalam pengklasteran wilayah rawan banjir di Kabupaten Lamongan

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini merupakan hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 17 Mei 2024
Yang membuat pernyataan,



Adi Candra Arya Kusuma
NIM. 19650042

HALAMAN MOTTO

“Never Try, Never Know”

HALAMAN PERSEMBAHAN

Puja dan puji syukur atas kehadiran Allah subhanahu wa ta'ala, serta shalawat dan salam bagi Rasul-Nya Penulis mempersembahkan hasil karya ini kepada:

Orang tua penulis yang sangat dicintai dan disayangi, Bapak Sumakin dan Ibu Luluk Mairawati, yang tak pernah berhenti memberikan dukungan, do'a, semangat serta motivasi kepada penulis.

Para dosen pembimbing penulis, Bapak Dr. Muhammad Faisal, M.T dan Bapak Dr. Fresy Nugroho, M.T yang senantiasa dengan ketelatenan dan penuh rasa kesabaran dalam memberikan saran, bimbingan, serta masukan dalam penyusunan karya ini.

Seluruh dosen dan jajaran civitas akademica jurusan Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang yang telah memberi ilmu dan memberi kelancaran dalam penyusunan karya ini, tak lupa seluruh guru-guru penulis yang telah memberikan banyak ilmu dan pengalaman yang sangat berharga kepada penulis.

Seluruh pihak-pihak yang tak dapat penulis sebutkan satu per satu yang telah selalu memberikan dukungan, do'a, semangat serta bantuan sehingga penulis dapat menyelesaikan skripsi dengan lancar dan tepat waktu.

KATA PENGANTAR

Assalamu'alaikum warahmatullahi wabarakatuh.

Alhamdulillah segala puji dan syukur senantiasa penulis panjatkan pada Allah SWT yang berkat rahmat, serta hidayah-Nya, sehingga penulis dapat menyempurnakan skripsi ini dengan tepat waktu. Sholawat serta salam tetap tercurahkan kepada Nabi Muhammad SAW yang telah menuntun umat manusia menuju jalan yang lebih baik.

Dengan kerendahan hati dan penuh rasa syukur bahagia ini, saya mempersembahkan skripsi yang telah saya susun ini untuk memenuhi salah satu persyaratan sarjana komputer (S.Kom) di Universitas Islam Negeri Maulana Malik Ibrahim Malang program studi Teknik Informatika Fakultas Sains dan Teknologi.

Penulis mengucapkan rasa terima kasih yang begitu besar kepada seluruh pihak yang memberikan dukungan dan membantu rampungnya skripsi ini. Ucapan terima kasih ini penulis disampaikan kepada:

1. Prof. Dr. M. Zainuddin, M.A., selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Prof. Dr. Hj. Sri Harini, M.Si., selaku dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Dr. Fachrul Kurniawan, M.MT, selaku ketua Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.

4. Dr. Muhammad Faisal, M.T, dan Dr. Fresy Nugroho, M.T, selaku dosen pembimbing I dan II yang telah membimbing dan memberikan arahan kepada peneliti sehingga dapat menyelesaikan skripsi ini.
5. Dr. Fachrul Kurniawan, M.MT dan Hani Nurhayati, M.T selaku dosen penguji I dan II yang telah memberikan, kritik serta saran kepada penulis hingga ujian skripsi dengan penuh kesabaran.
6. Seluruh dosen dan segenap staff program studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang yang telah memberikan segala ilmu dan wawasan semasa kuliah.
7. Kedua orang tua tercinta yang selalu memberikan semangat dan do'a pada penulis hingga rampungnya skripsi ini.
8. Tholaat Ahmad Novriyanto, Herliand Putra O T, serta teman-teman dekat penulis yang senantiasa memberikan dukungan penuh dan menjadi teman perjalanan menempuh pendidikan mulai dari menjadi mahasiswa baru hingga lulus satu persatu.
9. Seseorang yang tidak bisa disebutkan namanya yang telah memberikan dukungan dan menjadi penyemangat bagi penulis.
10. Kawan-kawan Alliance of Informatics Engineering (ALIEN) Angkatan 2019, yang senantiasa selalu memberikan semangat dan dukungan dalam berjuang bersama dalam mengejar gelar S.Kom dan pengalaman di Universitas yang sama.
11. Semua pihak yang tidak dapat penulis sebutkan satu persatu yang telah membantu penulis dalam menyelesaikan skripsi ini.

Penulis menyadari bahwa dalam penyusunan skripsi ini masih jauh dari kata sempurna. Maka dari itu penulis menerima saran, kritik dan masukan yang bersifat membangun sehingga dapat menjadi lebih baik kedepannya. Penulis berharap semoga skripsi ini dapat memberikan manfaat untuk kedepannya.

Wassalamu 'alaikum warahmatullahi wabarakatuh.

Malang, 15 Juli 2024

Penulis

DAFTAR ISI

HALAMAN JUDUL	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	iv
PERNYATAAN KEASLIAN TULISAN	v
HALAMAN MOTTO	vi
HALAMAN PERSEMBAHAN	vii
KATA PENGANTAR.....	viii
DAFTAR ISI	xi
DAFTAR GAMBAR.....	xii
DAFTAR TABEL	xiii
ABSTRAK.....	xiv
ABSTRACT	xv
الملخص	xvii
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	8
1.3 Tujuan Penelitian	8
1.4 Manfaat Penelitian	8
1.5 Batasan Masalah	9
BAB II STUDI PUSTAKA	10
2.1 Penelitian Terkait.....	10
2.2 Machine Learning	15
2.3 Profil Kabupaten Malang	18
2.4 Bencana	24
2.5 Banjir.....	26
2.6 Algoritma K-Means	31
2.7 Silhouette Coefficient.....	34
2.8 Elbow Method	37
BAB III METODOLOGI PENELITIAN	38

3.1 Desain Sistem	38
3.1.1 Seleksi Data	39
3.1.2 Praproses Data.....	40
3.1.3 K-Means Clustering	41
3.1.4 Evaluasi Hasil	48
BAB IV HASIL DAN PEMBAHASAN.....	52
4.1 Hasil	52
4.1.1 Akuisisi Data.....	52
4.1.2 Praproses Data.....	53
4.1.3 K-Means Clustering	55
4.1.4 Elbow Method.....	58
4.1.5 Hasil Pengujian	59
4.1.6 Pengujian Kinerja Sistem	64
4.2 Pembahasan	65
BAB V KESIMPULAN DAN SARAN	68
5.1 Kesimpulan.....	68
5.2 Saran.....	68
DAFTAR PUSTAKA	69

DAFTAR GAMBAR

Gambar 3.1 Desain Sistem Wilayah Rawan Banjir	38
Gambar 3.2 Desain Alur Algoritma K-Means	42
Gambar 4.1 Visualisasi Klaster Hasil K-Means	57
Gambar 4.2 Visualisasi 3D Klaster Hasil K-Means	58
Gambar 4.3 Hasil Pengujian <i>Elbow Method</i>	60
Gambar 4.4 Visualisasi K-Means Dengan Nilai $K=2$	61
Gambar 4.5 Visualisasi K-Means Dengan Nilai $K=4$	61
Gambar 4.6 Visualisasi K-Means Dengan Nilai $K=5$	62
Gambar 4.7 Visualisasi K-Means Dengan Nilai $K=3$	62

DAFTAR TABEL

Tabel 2.1 Tabel Penelitian Terkait.....	13
Tabel 2.2 Luas Daerah dan Persentase Terhadap Kabupaten Lamongan	19
Tabel 2.3 Interpretasi nilai <i>Silhouette Coefficient</i>	36
Tabel 3.1 Tabel Data Set (Cluster Kabupaten Lamongan)	39
Tabel 3.2 Cluster Yang Digunakan.....	43
Tabel 3.3 Tabel Centroid Awal	43
Tabel 3.4 Jarak Titik Data Dengan Centroid (Iterasi 1).....	46
Tabel 3.5 Hasil Iterasi 2	48
Tabel 3.6 Hasil Percobaan Nilai <i>Silhouette</i> untuk $k=2$	49
Tabel 3.7 Hasil Percobaan Nilai <i>Silhouette</i> untuk $k=3$	50
Tabel 4.1 Dataset Penelitian.....	52
Tabel 4.2 Hasil Clustering.....	63
Tabel 4.3 Hasil Nilai $a(i)$, $b(i)$, dan nilai <i>Silhouette</i> Masing-masing Data	64

ABSTRAK

Kusuma, Adi Candra Arya. 2024. **Implementasi Metode K-Means Dalam Pengelompokan Wilayah Rawan Banjir Di Kabupaten Lamongan**. Skripsi. Prodi Teknik Informatika, Fakultas sains dan Teknologi. Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Dr. Muhammad Faisal, M. T, (II) Dr. Fresy Nugroho, M. T.

Kata Kunci: Banjir, *K-Means*, *Clustering*.

Banjir merupakan bencana yang paling sering terjadi, dilihat dari intensitasnya pada suatu tempat yaitu sebesar 40% diantara bencana alam yang lain, oleh karena itu diperlukan metode pengelompokan yang menentukan suatu wilayah termasuk kedalam wilayah rawan banjir atau tidak. Penelitian ini membahas tentang pengklasteran wilayah rawan banjir di Kabupaten Lamongan. Jenis penelitian ini adalah penelitian kuantitatif dengan menggunakan metode *K-Means clustering* melalui tiga variabel antara lain curah hujan, kemiringan tanah, dan jumlah kejadian banjir. Data yang digunakan diolah melalui beberapa tahapan yaitu seleksi data dan praproses data. Setelah dilakukan praproses data, berikutnya dilakukan klasterisasi menggunakan metode *K-Means*. Dari hasil segmentasi data menggunakan *K-Means*, menghasilkan beberapa kesimpulan yaitu wilayah sering banjir, wilayah sedang banjir, dan wilayah sedikit banjir. Hasil data yang telah disegmentasi kemudian divalidasi menggunakan metode *Silhouette coefficient* dan menghasilkan nilai *silhouette* sebesar 0,6683. Hasil dari klasterisasi menggunakan *K-Means* menghasilkan 13 kecamatan pada *cluster* 1, 7 kecamatan pada *cluster* 2, dan 7 kecamatan pada *cluster* 3.

ABSTRACT

Kusuma, Adi Candra Arya. 2023 **Implementation of *K-Means* Method in Flood-Prone Area Clustering in Lamongan Regency**. Thesis. Department of Informatics Engineering, Faculty of Science and Technology. Maulana Malik Ibrahim State Islamic University Malang. Advisors: (I) Dr. Muhammad Faisal, M. T, (II) Dr. Fresy Nugroho, M. T.

Keyword: Flood, K-Means, Clustering.

Flood is the most frequently occurring disaster, with an intensity of 40% compared to other natural disasters in certain areas. Therefore, a clustering method is needed to determine whether an area falls into a flood-prone category. This study discusses the clustering of flood-prone areas in Lamongan Regency. This research is quantitative in nature and utilizes the K-Means clustering method based on three variables: rainfall, soil slope, and flood occurrences. The data underwent several stages of processing including data selection and preprocessing. After preprocessing, clustering was performed using the K-Means method. The results of data segmentation using K-Means led to several conclusions: areas prone to frequent flooding, areas prone to moderate flooding, and areas with minimal flooding. The segmented data was validated using the Silhouette coefficient method, yielding a silhouette value of 0.6683. The K-Means clustering resulted in 13 districts in cluster 1, 7 districts in cluster 2, and 7 districts in cluster 3.

الملخص

كوسوما، عادي تشاندرا أريا. 2024. تطبيق طريقة **K-Means** في تجميع المناطق عرضة للفيضانات في مقاطعة لامونغان. رسالة بكالوريوس. قسم هندسة الحاسوب، كلية العلوم والتكنولوجيا. جامعة مولانا مالك إبراهيم الإسلامية، مالانج. المشرفون : (I)الدكتور محمد فيصل، M.T.، (II)الدكتور فريسي نوغروهو، M.T.

الكلمات المفتاحية: الفيضانات، K-Means، التجميع

الفيضان هو الكارثة التي تحدث بشكل متكرر، حيث تبلغ كثافتها 40% مقارنة بالكوارث الطبيعية الأخرى في مناطق معينة. لذا، يُعتبر ضرورياً استخدام طريقة التجميع لتحديد ما إذا كانت منطقة ما عرضة للفيضانات أم لا. يتناول هذا البحث تجميع المناطق للتجميع عبر **K-Means** العرضة للفيضانات في مقاطعة لامونغان. يعد هذا البحث من نوع البحث الكمي باستخدام طريقة ثلاث متغيرات هي كميات الأمطار وميل التربة وعدد حالات الفيضانات. تم معالجة البيانات المستخدمة من خلال عدة مراحل بما من **K-Means**. في ذلك اختيار البيانات وتجهيزها مسبقاً. بعد التجهيز المسبق للبيانات، تم تنفيذ التجميع باستخدام طريقة ، توصل البحث إلى عدة استنتاجات وهي المناطق المعرضة للفيضانات المتكررة **K-Means** خلال نتائج تقسيم البيانات باستخدام والمناطق المعرضة للفيضانات المعتدلة والمناطق المعرضة للفيضانات القليلة. تم تحقيق التحقق من البيانات المقسمة باستخدام معامل إلى الحصول على 13 حياة **K-Means** السيوليت وأسفر ذلك عن قيمة سيوليت قدرها 0.6683. أدى التجميع باستخدام مدينة في العنقود الأول، 7 حياة مدينة في العنقود الثاني، و 7 حياة مدينة في العنقود الثالث

BAB I

PENDAHULUAN

1.1 Latar Belakang

Bencana alam merupakan suatu fenomena yang sering dihadapi oleh manusia saat ini. Menurut terminologi UNISDR pada tahun 2009, bencana diartikan sebagai gangguan yang bersifat serius terhadap keberlangsungan suatu masyarakat atau komunitas tertentu sehingga berdampak luas dan menimbulkan kerugian. Bencana alam dapat disebabkan oleh kejadian alam (natural disaster) maupun oleh ulah manusia (human-made disaster). Menurut United Nations International Strategi for Disaster Reduction bencana dapat dikelompokkan menjadi bahaya geologi (geological hazards), bahaya hidrometeorologi (hydrometeorological hazards), bahaya biologi (biological hazards), bahaya teknologi (technological hazards) dan penurunan kualitas lingkungan (environmental degradation) (UNISDR, 2009). Negara yang sering mengalami berbagai macam bencana adalah negara - negara di Asia seperti Indonesia, Jepang, Malaysia dan lain sebagainya.

Indonesia merupakan negara berpotensi terhadap bencana jika ditinjau dari aspek geografis, klimatologis dan demografis. Pertemuan antara dua samudera dan dua benua menyebabkan indonesia sangat berpotensi akan terjadinya bencana yang mengancam. Salah satu dari beberapa bencana yang berpotensi tinggi terjadi di Indonesia adalah bencana banjir (Soliha 2020). Indonesia yang merupakan negara kepulauan yang terletak pada pertemuan 4 (empat) lempeng tektonik, yaitu lempeng Benua Asia, Benua Australia, Samudera Hindia, dan Samudera Pasifik. Di

bagian selatan dan timur Indonesia, terdapat sabuk vulkanik (volcanic arc) yang memanjang dari Pulau Sumatera - Jawa - Nusa Tenggara - Sulawesi, yang sisinya berupa pegunungan vulkanik tua dan dataran rendah yang sebagian didominasi oleh rawa-rawa. Indonesia juga terletak pada titik pertemuan tiga lempeng tektonik (tectonic plate) yang saling bertabrakan, yaitu Lempeng Eurasia, Lempeng Indo Australia, dan Lempeng Pasifik. Indonesia menjadi negara yang memiliki rangkaian gunung api aktif (rangkaiannya Gunung Api Indonesia) dengan jumlah gunung api aktif sekitar 140 gunung. Kondisi tersebut menjadikan Indonesia sangat rawan bencana seperti letusan gunung berapi, gempa bumi, tsunami, banjir, tanah longsor, dan angin puting beliung.

Bencana alam di Indonesia seakan tidak mengenal musim. Banjir dan banjir bandang, tanah longsor, serta angin puting beliung terjadi di musim hujan. Pada musim kemarau, bencana kebakaran hutan dan lahan (Karhutla) datang mengancam. Sementara itu, letusan gunung berapi, gempa, dan tsunami sewaktu-waktu dapat terjadi. Ancaman bencana yang sedemikian rupa tentu menuntut adanya kewaspadaan serta kesiapsiagaan semua pihak, baik pemerintah maupun masyarakat. Proses dan akibat dari bencana alam jelas-jelas mengancam keselamatan masyarakat secara luas. Oleh karena itu, pencegahan, penanganan, serta penanggulangan bencana alam memerlukan kehadiran negara. Dengan kata lain, pemerintah harus berperan langsung dalam langkah-langkah mitigasi bencana karena bencana alam bukan merupakan urusan individu melainkan telah menyangkut keselamatan publik. Atas dasar itu, pemerintah telah menetapkan Undang-Undang Nomor 24 Tahun 2007 tentang Penanggulangan Bencana.

Berdasarkan data yang dihimpun oleh Pusat Data Informasi dan Komunikasi Kebencanaan BNPB, pada bulan februari 2023 tercatat telah terjadi 295 kejadian bencana yang terjadi di seluruh Indonesia. Musim penghujan yang terjadi pada bulan februari menjadikan bencana Hidrometeorologi banjir mendominasi dengan 125 kejadian, Cuaca ekstrim 96 kejadian, tanah longsor 53 kejadian dan abrasi dengan 4 kejadian, sedangkan bencana hidrometeorologi kering yang terjadi yaitu kebakaran hutan dan lahan sebanyak 12 kejadian. Selain korban jiwa, terdapat juga kerusakan pada rumah dan beberapa fasilitas umum (BNPB, 2023).

Kabupaten lamongan terletak pada koordinat $112^{\circ} 33'' - 112^{\circ} 34''$ BT dan $6^{\circ} 51'' - 7^{\circ} 23''$ LS jika ditinjau dari letak geografisnya. Kabupaten Lamongan menjadi salah satu dari beberapa Kabupaten di Jawa Timur yang berpotensi terjadinya bencana banjir (Arfah, 2023). Bencana banjir yang terjadi di wilayah pesisir Kabupaten Lamongan menjadi bencana rutinitas yang terjadi tiap tahunnya. Seperti yang dikutip dari situs berita online kompas.com pada bulan Februari 2023, menyebutkan bahwa kejadian bencana banjir di Kabupaten Lamongan menyebabkan 25 desa di 4 Kecamatan terendam dengan ketinggian genangan yang bervariasi. Bencana banjir tersebut terjadi akibat curah hujan tinggi sehingga menyebabkan naiknya muka air laut serta diikuti dengan meluapnya sungai-sungai di sekitar Kabupaten Lamongan khususnya sungai bengawan jero dan bengawan solo. Titik terparah terdapat di Kecamatan Kalitengah dimana terdapat 8 Desa yang terdampak. Akibat terjadinya bencana banjir tersebut ribuan rumah terendam dan fasilitas umum juga terdampak.

Menurut Niode banjir dapat dikategorikan berdasarkan mekanisme terjadinya dan berdasarkan posisi dari sumber banjir terhadap daerah yang digenangnya (Niode, 2016). Berdasarkan mekanisme terjadinya dapat dibedakan menjadi banjir biasa (regular) dan banjir tidak biasa (irregular) (Risanty, 2015). Banjir regular terjadi akibat jumlah limpasan yang sangat banyak sehingga melampaui kapasitas dari pembuangan air dan banjir irregular terjadi akibat tsunami, gelombang pasang, dan luapan air sungai atau rob.

Suatu teknik pengelompokan berdasarkan kedekatan (*similarity*) disebut clustering (Wathoriq & Subandi, 2020). Jika istilah "kelompok" mengacu pada sekelompok orang yang berbagi keadaan yang sama, pengelompokan berbeda dari kelompok. Pengelompokan tersebut didasarkan pada kesamaan karakteristik sampel yang ada, salah satunya dengan memanfaatkan rumus jarak Euclidean. Pengelompokan kelompok tidak harus sama, teknik text mining seperti clustering digunakan untuk membagi objek atau data menjadi beberapa kelompok (cluster) sedemikian rupa sehingga data yang terkait akan berada dalam satu cluster dan jarak antar cluster akan dibuat sejauh mungkin (Irianto, 2022). Karena kesederhanaan dan tingkat akurasi yang tinggi sehubungan dengan ukuran data, K Means Clustering adalah algoritma yang populer untuk masalah pengelompokan. K-Means merupakan salah satu metode clustering/pengelompokan non hirarki. Teknik pengelompokan datanya sederhana dan cepat. Ada banyak pendekatan untuk membuat cluster, diantaranya adalah membuat aturan yang mendikte keanggotaan dalam grup yang sama berdasarkan tingkat persamaan diantara anggota – anggotanya (Rosmini, 2018). Pendekatan lainnya adalah dengan

membuat sekumpulan fungsi yang mengukur beberapa properti dari pengelompokan tersebut sebagai fungsi dari beberapa parameter dari sebuah clustering.. Selain itu, algoritma ini tidak terpengaruh oleh urutan objek sehingga lebih efektif dan terukur saat memproses data dalam jumlah besar. DBI adalah salah satu teknik penilaian klaster pada pendekatan pengelompokan berbasis nilai kohesi dan pemisahan. Kohesi dalam sebuah clustering adalah total jarak data dari centroid cluster yang sedang dilacak. Untuk selanjutnya pemisahan ditentukan oleh jarak antar centroid cluster.

Dalam Al-Qur'an, apapun yang menimpa manusia baik kejadian positif maupun negatif, dalam peristiwa alam maupun sosial disebut dengan musibah. Begitupula dengan bencana banjir yang didalam Al-Qur'an dikatakan bahwa musibah berupa nikmat itu berasal dari Allah sedangkan musibah berupa bencana itu karena hasil perbuatan manusia sendiri seperti dalam surat An-Nisa' (4) ayat 79 "apa saja nikmat yang kamu peroleh adalah dari Allah, dan apa saja bencana yang menimpamu adalah dari (kesalahan) dirimu sendiri. Kami mengutusmu menjadi rasul kepada segenap manusia. Dan cukuplah Allah menjadi saksi"

Dalam Tafsir Ibnu Katsir menjelaskan bahwa Allah Subhanahu Wa Ta'ala. berfirman kepada Rasul-Nya, tetapi makna yang dimaksud ialah mencakup semua orang, sehingga firman berikut dapat dianggap sebagai jawaban, yaitu: "Apa saja nikmat yang kamu peroleh adalah dari Allah". Yakni dari kemurahan Allah, kasih sayang serta rahmat-Nya. "dan apa saja bencana yang menimpamu, maka dari (kesalahan) dirimu sendiri" Yaitu akibat perbuatanin的角度 sendiri. Perihalnya sama dengan makna yang terkandung di dalam ayat lain, yaitu firman-Nya:

وَمَا أَصَابَكُمْ مِنْ مُصِيبَةٍ فِيمَا كَسَبْتُمْ أَيْدِيكُمْ وَيَعْفُوا عَنْ كَثِيرٍ

“Dan apa saja musibah yang menimpa kamu adalah disebabkan oleh perbuatan tanganmu sendiri, dan Allah memaafkan sebagian besar (dari kesalahanmu)”. (Asy Syuura:30).

As-Saddi, Al-Hasan Al-Basri, Ibnu Juraij, dan Ibnu Zaid mengatakan sehubungan dengan firman-Nya: maka dari dirimu sendiri. Yaitu disebabkan dosamu sendiri. Qatadah mengatakan sehubungan dengan makna firman-Nya: maka dari dirimu sendiri. sebagai hukuman buatmu, hai anak Adam, karena dosamu sendiri. Qatadah mengatakan, telah diriwayatkan kepada kami bahwa Nabi Shalallaahu Alaihi Wassalaam. telah bersabda:

لَا يُصِيبُ رَجُلًا حَدَشُ عُوْدٍ وَلَا عَنْرَةُ قَدَمٍ، وَلَا اخْتِلَاجُ عِرْقٍ إِلَّا بِذَنْبٍ، وَمَا يَعْفُو اللَّهُ أَكْثَرَ

“Tidak sekali-kali seseorang terkena lecet (karena tertusuk) kayu, tidak pula kakinya tersandung, tidak pula uratnya terkilir, melainkan karena dosa(nya), tetapi yang dimaafkan oleh Allah jauh lebih banyak.”

Hadis mursal yang diriwayatkan oleh Qatadah ini telah diriwayatkan secara muttasil di dalam kitab sahih, yang bunyinya mengatakan: “Demi Tuhan yang jiwaku berada di dalam genggamannya, tiada suatu kesusahan pun yang menimpa orang mukmin, tiada suatu kesedihan pun, dan tiada suatu kelelahan pun, hingga duri yang menusuk (kaki)nya, melainkan Allah menghapuskan sebagian dari dosa-dosanya karena musibah itu.” Abu Saleh mengatakan sehubungan dengan firman-Nya: Dan apa saja bencana yang menimpamu, maka dari (kesalahan)mu sendiri. (An-Nisa: 79) Yakni karena dosamu sendiri, dan Akulah (kata Allah) yang menakdirkannya atas dirimu. Demikianlah menurut riwayat Ibnu Jarir. (Katsir, 2015)

Menurut Prasetyo Budi Agustinus dalam penelitiannya tentang pemetaan lokasi rawan dan risiko banjir di kabupaten Surakarta menyatakan bahwa pemetaan banjir merupakan usaha merepresentasikan data yang berupa angka atau tulisan tentang distribusi banjir ke dalam bentuk peta agar persebaran datanya dapat langsung diketahui dengan mudah dan cepat. Peta yang menggambarkan fenomena geografikal tidak hanya sekedar pengecilan suatu fenomena saja, tetapi jika perlu itu dibuat dan didesain dengan baik, maka akan menjadi alat bantu yang baik untuk kepentingan melaporkan, memperagakan, menganalisis dan secara umum untuk memahami suatu objek atau kenampakan di muka bumi. Peta menggunakan simbol dua dimensi untuk mencerminkan fenomena geografikal yang dilakukan secara sistematis dan memerlukan kecakapan untuk membuat dan membacanya. Peta merupakan teknik komunikasi yang tergolong dalam cara grafis dan untuk efisiensinya harus mempelajari atribut atau elemen-elemen dasarnya (Prasetyo, 2013).

Pemetaan daerah rawan banjir merupakan salah satu cara pengendalian banjir secara non-structural yang menjadi salah satu upaya dalam mengoptimalkan langkah antisipasi bencana. Penerapan dari metode K-Means menghasilkan tingkat akurasi dari nilai keanggotaan klaster yang cukup baik (Herlinda, 2021). Oleh karena itu, pengelompokan daerah rawan banjir ini dilakukan menggunakan metode K-Means karena dengan menggunakan metode ini dapat membantu mengidentifikasi wilayah rawan banjir. Dengan metode ini dapat membantu peneliti dalam melakukan identifikasi wilayah rawan banjir menggunakan parameter yang telah ditentukan dalam variabel masukan (input) yaitu curah hujan,

kemiringan tanah, jumlah kejadian banjir. Maka, dengan menggunakan metode K-Means nantinya variabel keluaran (output) adalah berupa wilayah yang rawan terdampak banjir.

Dari latar belakang diatas, penelitian ini akan mengimplementasikan K-Means untuk mengelompokkan wilayah-wilayah rawan banjir yang ada di kabupaten Lamongan.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan sebelumnya, masalah dalam penelitian ini adalah bagaimana implementasi metode K-Means dalam menentukan wilayah rawan banjir di Kabupaten Lamongan?

1.3 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah untuk mengetahui bagaimana implementasi metode K-Means dalam menentukan wilayah rawan banjir di Kabupaten Lamongan.

1.4 Manfaat Penelitian

Adapun beberapa manfaat penelitian yang didapatkan dari penelitian ini adalah sebagai berikut :

- a. Membantu pemerintah Kabupaten Lamongan dalam melakukan analisa dan antisipasi bencana di wilayah terdampak.
- b. Menambah wawasan dan pengetahuan peneliti dan pembaca tentang implementasi metode K-Means.

- c. Memberi informasi kepada masyarakat sekitar tentang wilayah yang rawan terdampak banjir dan menambah kesadaran masyarakat akan pentingnya menjaga kebersihan lingkungan dan tidak membuang sampah ditempat yang bukan semestinya.

1.5 Batasan Masalah

Dalam melakukan pengklatseran wilayah rawan banjir, terdapat beberapa batasan masalah dalam penelitian ini, antara lain :

- a. Sumber data yang dipakai bersumber dari penelitian terdahulu, BMKG dan badan pusat statistik Kabupaten Lamongan.
- b. Wilayah yang diteliti adalah kecamatan-kecamatan di Kabupaten Lamongan.
- c. Variabel yang digunakan sebagai parameter penentu adalah curah hujan, kemiringan tanah dan jumlah kejadian banjir.
- d. Penelitian ini hanya sampai menghasilkan data wilayah yang sudah di segmentasi.

BAB II

STUDI PUSTAKA

Dalam melakukan penelitian ini, peneliti telah mempelajari dan mengkaji beberapa penelitian terkait untuk mendukung teori dan hasil yang ingin dicapai peneliti. Penelitian yang diteliti merupakan penelitian terdahulu maupun penelitian yang telah dilakukan sebelumnya yang mempunyai relevansi dengan penelitian yang sedang dilakukan saat ini. Penelitian terdahulu dilakukan dengan mengumpulkan studi literatur yang relevan dan mencari yang berkaitan dengan penelitian yang dilakukan. Tujuan yang ingin dicapai dengan melakukan tinjauan terhadap penelitian terdahulu adalah agar penelitian ini mempunyai landasan pemahaman teoritis yang dapat dipertanggungjawabkan.

2.1 Penelitian Terkait

Penelitian terkait adalah suatu studi yang dilakukan untuk menggali informasi, mengumpulkan data, dan menganalisis topik atau pertanyaan tertentu yang terkait dengan pengetahuan atau permasalahan yang telah ada sebelumnya. Penelitian ini dilakukan dengan tujuan untuk memperluas pemahaman tentang topik yang telah diteliti sebelumnya atau untuk mengidentifikasi kesenjangan pengetahuan yang masih ada.

Menurut Nurdiawan, hasil analisa spasial dengan SIG pada penelitiannya yang berjudul Pemetaan Daerah Rawan Banjir Berbasis Sistem Informasi Geografis dalam Upaya Mengoptimalkan Langkah Antisipasi Bencana memprediksi bahwa ada 16 daerah di kabupaten Cirebon yang di prediksi menjadi daerah rawan banjir,

antara lain Waled, Pabedilan, Gebang, Lemahabang, Susukan Lebak, Astanajapura, Pangenan, Mundu, Talun, Tengahtani, Gunungjati, Kapetakan, Suranenggala, Klenganan, Panguraggan dan Susukan (Nurdiawan et al. 2018). Dari hasil analisis penyebab banjir di Kabupaten Cirebon diketahui bahwa saluran drainase, kemiringan lereng dan penggunaan lahan sangat berperan dalam terjadinya banjir yang menyebabkan daerah tersebut rawan terhadap banjir. Persamaan dengan penelitian ini adalah kesamaan dalam meneliti wilayah rawan banjir. Perbedaan dengan penelitian ini adalah memunculkan tingkat kerawanan bencana banjir.

Penelitian berikutnya dilakukan oleh Nanda pada tahun 2020, pada penelitiannya berhasil mengimplementasikan metode Fuzzy C-means untuk menentukan wilayah rawan nyamuk di Kota Malang menggunakan tiga parameter yaitu curah hujan, kelembaban suhu dan jumlah penderita yang menghasilkan wilayah sedikit nyamuk, wilayah sedang nyamuk dan wilayah banyak nyamuk pada setiap kelurahan di Kota Malang. Dengan hasil pengujian kinerja sistem menggunakan rule base memperoleh nilai hasil presentasi keberhasilan clustering wilayah rawan nyamuk sebesar 74% dan ketidakberhasilan 26% (Nanda, 2020). Kelebihan dari penelitian ini adalah menggunakan sistem informasi geografis sehingga mudah untuk dipahami dan user dapat mengetahui sebaran daerah rawan DBD di tiap kecamatan secara geografis. Kelemahan dari penelitian ini adalah peta geografis hanya mampu menampilkan satu marker map saja dan tidak bisa menampilkan sesuai hasil dari setiap cluster. Persamaan dengan penelitian ini adalah sama-sama meneliti tentang tingkat kerawanan suatu wilayah. Perbedaan

dengan penelitian ini adalah penelitian ini meneliti tentang tingkat kerawanan banjir.

Penelitian lainnya dilakukan oleh Henny pada tahun 2020 yang berjudul Analisis Tingkat Kerawanan Banjir Di Kabupaten Lamongan dan memberikan hasil berupa peta tingkat kerawanan banjir dengan dua batasan analisis wilayah yaitu wilayah Sub-DAS Bengawan Solo hilir dan Sub-DAS Kali Lamong. Kawasan sangat rawan banjir mengelompok di kecamatan Glagah, Karangbinangun, Kalitengah, Karanggeneng, Maduran, Laren, Sekaran, Babat, Sukodadi, Turi, Deket, Lamongan, dan Sarirejo. Faktor dominan yang menjadi penyebab banjir di Kabupaten Lamongan adalah kemiringan lereng dan penggunaan lahan (Pratiwi Eka Henny, 2020). Sebagian besar wilayah Kabupaten Lamongan memiliki kemiringan lereng 0-8% dengan presentasi 51,57% dari luas total kabupaten, dan 59,3 % penggunaan lahan berupa tambak serta sawah. Persamaan dengan penelitian ini adalah kesamaan dalam meneliti tingkat kerawanan banjir suatu wilayah. Perbedaan dengan penelitian ini adalah penelitian ini tidak spesifik pada wilayah sub-DAS Bengawan Solo dan sub-DAS Kali Lamong.

Penelitian selanjutnya dilakukan oleh Subandi pada tahun 2023 yang meneliti tentang daerah rawan banjir di kota Tangerang menggunakan data historis tentang bencana banjir di Kota Tangerang yang dikumpulkan dan dipersiapkan untuk analisis. Metode K-Means Clustering diterapkan untuk mengelompokkan wilayah berdasarkan tingkat dampak yang dialami akibat banjir. Pengelompokan ini dilakukan dengan mempertimbangkan parameter seperti jumlah kejadian, jumlah kerusakan, korban menderita, korban meninggal, korban menghilang, korban

mengungsi. Evaluasi dilakukan untuk mengukur kualitas pengelompokan dan mendapatkan wawasan yang berarti (Try & Subandi, 2023). Dari hasil analisis dan percobaan pada penelitian ini dapat disimpulkan bahwa analisis clustering menggunakan algoritma K-Means yang dirancang untuk mengelompokkan data banjir dari BPBD Kota Tangerang dapat mengidentifikasi daerah dengan tingkat kerentanan terbesar. Sistem ini mampu menampilkan hasil dari tiap-tiap proses perhitungan agar dapat mempermudah user untuk menganalisa data yang di proses menggunakan K-Means Clustering

Penelitian berikutnya dilakukan oleh Setiawan pada tahun 2020 yang meneliti tentang pemetaan clustering tingkat rawan banjir di Kota Bengkulu menggunakan fuzzy simple adaptive weighting menggunakan 4 parameter yaitu data spasial tutupan lahan, topografi, jarak tepi sungai terhadap pemukiman dan data non-spasial berupa curah hujan. Hasil dari penelitian ini adalah terdapat 6 kelurahan yang dapat memiliki prioritas kebijakan penanganan dan penataan wilayah DAS sungai Bengkulu. Sejak 2014, enam kelurahan tersebut terus mengalami peningkatan tingkat kerawanan banjir (Setiawan, 2020).

Tabel 2.1 Tabel Penelitian Terkait

No	Nama Peneliti	Judul Penelitian	Persamaan	Perbedaan
1	Nurdiawan	Pemetaan Daerah Rawan Banjir Berbasis Sistem Informasi Geografis dalam Upaya Mengoptimalkan	Meneliti wilayah rawan banjir	Memunculkan tingkat kerawanan banjir

		Langkah Antisipasi Bencana		
2	Nanda Ayu Istiqomah	Sistem Informasi Geografis Wilayah Rawan Nyamuk di Kota Malang Menggunakan Metode Fuzzy C-Means	Meneliti tingkat kerawanan suatu wilayah	Objek penelitian wilayah rawan banjir di Kabupaten Lamongan
3	Henny Eka Pratiwi	Analisis Tingkat Kerawanan Banjir Di Kabupaten Lamongan	Meneliti tingkat kerawanan	Menggunakan metode K-Means
4	Rosmini	Implementasi Metode K-Means Dalam Pemetaan Kelompok Mahasiswa Melalui Data Aktivitas Kuliah	Menggunakan Metode K-Means	Menambahkan Bahasa Pemrograman
5	Yudi Setiawan	Pemetaan Zonasi Rawan Banjir Dengan Analisis Index Rawan Banjir Menggunakan Metode Fuzzy Simple Adaptive Weighting	Pemetaan wilayah	Menggunakan metode K-Means

2.2 *Machine Learning*

Machine Learning diartikan sebagai aplikasi komputer atau algoritma matematika yang di kombinasikan dengan proses pembelajaran data sehingga menghasilkan prediksi di masa yang akan datang (Goldberg & Holland, 1988). Selain itu machine learning merupakan cabang ilmu dalam kecerdasan buatan yang dibangun dengan bahasa pemrograman yang memungkinkan komputer berpikir seperti manusia dan dapat meningkatkan pemahaman mereka melalui pengalaman secara otomatis (Untoro et al., 2022). Proses pembelajaran yang dimaksud terbagi menjadi dua tahap yaitu training dan testing (Huang et al., 2006). Machine Learning adalah salah satu komponen penting di bidang ilmu data yang berkembang. Dengan penggunaan metode statistik, algoritma dilatih untuk membentuk sebuah klasifikasi maupun prediksi, sehingga dapat memberikan insight dalam proyek penambangan data. Machine Learning dapat digunakan untuk salah satu alat yang membantu dalam membantu keputusan dari informasi yang dihasilkan (Khairani et al., 2022). Dalam artikel yang dimuat dalam website Toward Data Science, ada tujuh tahapan dalam Machine Learning, diantaranya : (Yufeng G, 2017)

1. Mengumpulkan Data

Tahap ini sangat penting karena kuantitas dan kualitas data yang dikumpulkan akan memengaruhi seberapa baik model prediksi yang dihasilkan.

2. Menyiapkan Data

Di tahap ini, data yang telah dikumpulkan akan disimpan dan disiapkan untuk pelatihan *Machine Learning*. Beberapa persiapan data yang biasa dilakukan

adalah *binarization*, *mean removal*, *scaling*, *normalization*, *label encoding* (Andreas Chandra, 2018). Tahap ini dapat disesuaikan dengan data dan kebutuhan untuk pelatihan model *machine learning*. Selanjutnya data dibagi menjadi data *training* yang digunakan sebagai referensi.

3. Memilih Model

Model yang dipilih akan disesuaikan dengan data yang digunakan dan hasil yang ingin didapatkan dari data yang kita kumpulkan.

4. *Training*

Pada langkah ini, kami akan menggunakan data kami untuk secara bertahap meningkatkan kemampuan model kami untuk memprediksi. Saat pertama kali memulai pelatihan, kami seperti menggambar garis acak melalui data. Kemudian saat setiap langkah pelatihan berlangsung, garis bergerak, selangkah demi selangkah, mendekati pemisahan anggur dan bir yang ideal.

5. Evaluasi

Setelah pelatihan selesai, saatnya untuk melihat apakah modelnya bagus, menggunakan Evaluasi. Di sinilah kumpulan data yang di sisihkan sebelumnya berperan. Evaluasi memungkinkan kita untuk menguji model kita terhadap data yang belum terlihat. Ini dimaksudkan untuk mewakili bagaimana model dapat bekerja di dunia nyata.

6. *Parameter Tuning*

Tahap ini dilakukan bila ingin melihat apakah bisa meningkatkan hasil pelatihan dengan mengatur parameter dengan nilai lain.

7. *Prediction*

Machine Learning menggunakan data untuk menjawab pertanyaan. Maka di tahap ini, kita dapat menjawab pertanyaan dari model yang telah dihasilkan.

Dalam penelitian ini penulis menerapkan semua tahap 1-5 karena untuk pengaplikasian algoritma *K-Means Clustering*, pengaturan parameter dilakukan di tahap *Training*. Kumpulan data yang besar atau *big data* dapat dianalisis dengan berbagai macam tipe (Nurhayati et al., 2019). *Machine Learning* dibagi menjadi 3 tipe, diantaranya adalah : (Algren et al., 2021)

1. *Supervised Learning*

Machine learning tipe ini digunakan untuk membuat model dengan satu atau lebih variabel independen dan variabel dependen yang sesuai. Didalam *Machine learning*, variabel independen disebut fitur dan variabel dependen disebut label. Dengan data input yang cukup, algoritma supervised learning dapat menghasilkan model dari hubungan antara fitur dan label.

2. *Unsupervised Learning*

Unsupervised learning adalah salah satu jenis algoritma pembelajaran mesin yang digunakan untuk mengambil kesimpulan dari kumpulan data yang terdiri dari *input* tanpa respons yang diberi label (Pradono Iswara, Nurhayati, et al., 2019). Sebaliknya, algoritma ini bertugas menyortir titik data ke dalam kelompok, membersihkan data, atau menemukan struktur dalam kumpulan data.

3. *Reinforcement Learning*

Didalam *Machine learning* tipe ini tidak ada dataset *training*. Algoritma ini didesain untuk mencoba beberapa aksi atau jalan dari sebuah sistem penghargaan, sasaran atau batasan yang memberikan umpan balik tentang kualitas pilihannya.

Dalam penelitian ini algoritma *K-Means Clustering* yang digunakan termasuk ke dalam tipe *Unsupervised Learning* karena algoritma ini tidak memerlukan label pada data inputnya.

2.3 Profil Kabupaten Lamongan

2.3.1 Kondisi Geografi

Kabupaten Lamongan terletak di 6° sampai dengan 7° lintang selatan dan diantara 112° garis bujur. Luas wilayah dari kabupaten lamongan kurang lebih 1812,8 km² atau kurang lebih 3,78% dari luas wilayah propinsi Jawa Timur (Munasyaroh, 2015). Kabupaten Lamongan terletak di pesisir, sehingga memiliki panjang garis pantai sepanjang 47 km dengan wilayah perairan laut seluas 902,4 km². Daratan kabupaten Lamongan dibelah oleh sungai bengawan solo yang membentang dari barat ke timur. Jika dilihat dari kemiringan tanahnya, wilayah kabupaten Lamongan merupakan wilayah yang relatif datar, karena hampir 72,5% lahannya adalah datar dengan tingkat kemiringan 0-2%.

Menurut RPJMD Kabupaten Lamongan pada tahun 2022, Kabupaten Lamongan memiliki luas wilayah ±1.752,21 km² atau setara dengan 175.221 Ha atau ±3.67% dari luas wilayah Provinsi Jawa Timur. Mempunyai panjang garis pantai sepanjang 47 km dan wilayah perairan laut seluas 902,4 km², apabila

dihitung 12 mil dari permukaan laut. Kabupaten Lamongan terdiri dari 27 Kecamatan dengan rincian sebanyak 462 Desa dan 12 Kelurahan. Berikut rincian data terkait dengan luas daerah dan persentase terhadap luas Kabupaten Lamongan pada Tahun 2020:

Tabel 2.2 Luas Daerah dan Presentase terhadap Luas Kabupaten Lamongan

No	Kecamatan	Ibukota Kecamatan	Luas (ha)	Persentase terhadap Luas Kabupaten
1	Sukorame	Sukorame	3.929	2,24
2	Bluluk	Bluluk	5.800	3,31
3	Ngimbang	Sendangrejo	10.006	5,71
4	Sambeng	Ardirejo	15.078	8,61
5	Mantup	Mantup	9.123	5,21
6	Kembangbahu	Kembangbahu	6.825	3,90
7	Sugio	Sugio	8.974	5,12
8	Kedungpring	Kedungpring	8.573	4,89
9	Modo	Mojorejo	8.170	4,66
10	Babat	Bedahan	6.459	3,69
11	Pucuk	Pucuk	4.538	2,59
12	Sukodadi	Sukodadi	4.748	2,71
13	Lamongan	Lamongan	4.094	2,34
14	Tikung	Bakalanpule	5.542	2,16
15	Sarirejo	Dermolemahabang	4.887	2,79
16	Deket	Deketwetan	4.395	2,51
17	Glagah	Glagah	5.002	2,85
18	Karangbinangun	Sambopinggir	4.726	2,70
19	Turi	Sukoanyar	5.250	3,00
20	Kalitengah	Dibee	3.695	2,11
21	Karanggeneng	Karanggeneng	4.074	2,33
22	Sekaran	Bulutengger	5.405	3,08
23	Maduran	Maduran	3.278	1,87
24	Laren	Gampangsejati	9.411	5,37
25	Selokuro	Payaman	9.487	5,41
26	Paciran	Paciran	5.681	3,24
27	Brondong	Brondong	8.072	4,61
Jumlah			175.221	100,00

Secara astronomis, Kabupaten Lamongan terletak pada koordinat $6^{\circ}51'54''$ - $7^{\circ}23'6''$ Lintang Selatan dan berada di antara $112^{\circ}4'41''$ - $112^{\circ}33'12''$ Bujur Timur. Selanjutnya jika dilihat berdasarkan wilayah administrasi, Kabupaten Lamongan memiliki batas-batas wilayah sebagai berikut:

- a. Sebelah Utara, berbatasan dengan Laut Jawa
- b. Sebelah Barat, berbatasan dengan Kabupaten Bojonegoro dan Tuban
- c. Sebelah Selatan, berbatasan dengan Kabupaten Jombang dan Mojokerto
- d. Sebelah Timur, Berbatasan dengan Kabupaten Gresik

2.3.2 Kondisi Kawasan

Menurut RPJMD Kabupaten Lamongan tahun 2021-2026, Wilayah daratan Kabupaten Lamongan secara garis besar dibedakan menjadi 3 bagian, yaitu :

- a. Bagian tengah selatan merupakan daerah daratan rendah yang terdiri dari kecamatan Kedungpring, Babat, Sukodadi, Pucuk, Lamongan, Deket, Tikung, Sugio, Maduran, Sarirejo dan Kembangbahu.
- b. Bagian selatan dan utara merupakan pegunungan kapur berbatu dengan tingkat kesuburan sedang. Kawasan ini terdiri dari kecamatan Mantup, Sambeng, Ngimbang, Bluluk, Sukorame, Modo, Brondong, Paciran dan Solokuro.
- c. Bagian tengah utara merupakan daerah bonorowo yang rawan banjir. Kawasan ini terdiri dari kecamatan Sekaran, Laren, Karanggenng, Kalitengah, Turi, Karangbinangun dan Glagah. Kawasan ini terdampak banjir hampir di setiap musim hujan tiba.

2.3.3 Kondisi Topografi

Menurut RPJMD Kabupaten Lamongan pada tahun 2022, topografi/ketinggian adalah kondisi ketinggian suatu wilayah dari atas permukaan laut yang dinyatakan dengan mdpl (meter di atas permukaan laut). Kondisi topografi Kabupaten Lamongan berkisar antara 0- 186 meter di atas permukaan laut. Ketinggian tertinggi berada di Kecamatan Ngimbang yang terletak 186 meter di atas permukaan laut. Secara administrasi Kabupaten Lamongan terdiri dari 27 kecamatan yang terbagi menjadi 3 karakteristik daratan berdasarkan aliran sungai bengawan solo yaitu:

- a. Bagian tengah selatan merupakan daratan rendah yang relatif agak subur yang membentang dari Kecamatan Kedungpring, Babat, Sukodadi, Pucuk, Lamongan, Deket, Tikung, Sugio, Maduran, Sarirejo dan Kembangbahu
- b. Bagian utara dan selatan yang merupakan pegunungan kapur berbatu-batu dengan kesuburan sedang meliputi Kecamatan Mantup, Sambeng, Ngimbang, Bluluk, Sukorame, Modo, Brondong, Paciran, dan Solokuro.
- c. Bagian tengah utara yang merupakan daerah rawan banjir meliputi Kecamatan Sekaran, Laren, Karanggeneng, Kalitengah, Turi, Karangbinangun, Glagah.

2.3.4 Kondisi Hidrologi

Secara umum keberadaan air di Kabupaten Lamongan didominasi oleh air permukaan, dimana pada saat musim penghujan dijumpai air dalam jumlah yang melimpah hingga mengakibatkan bencana banjir pada sebagian wilayah. Namun sebaliknya pada saat musim kemarau, di sebagian besar wilayah Kabupaten Lamongan relatif berkurang. Kabupaten Lamongan dilewati oleh 3 sungai besar,

yaitu Sungai Bengawan Solo sepanjang \pm 68 Km dengan debit rata – rata 531,61 m³/bulan (debit maksimum 1.758,46 m³ dan debit minimum 19,58 m³) yang bermata air di Waduk Gajah Mungkur (Kabupaten Wonogiri, Jawa Tengah), Kali Blawi sepanjang \pm 27 Km dan Kali Lamong sepanjang \pm 65 Km yang bermata air di Kabupaten Lamongan. Waduk di Kabupaten Lamongan berjumlah 33 Waduk dengan luas keseluruhan 2.326 Ha dan volume 51.093.192 m³.

Keberadaan waduk di Kabupaten Lamongan dapat mengairi lahan sawah seluas 20.960 Ha utamanya lahan sawah berpengairan teknis dan $\frac{1}{2}$ teknis, sehingga keberadaan waduk sangat bermanfaat bagi masyarakat khususnya untuk meningkatkan produksi pertanian. Waduk yang paling besar di Kabupaten Lamongan adalah waduk Gondang yang berada di Kecamatan Sugio dengan luas 660 Ha dan kapasitas volume sebesar 23.712.500 m³. Di Kabupaten Lamongan juga terdapat rawa sebagai penampungan air. Jumlah rawa di Kabupaten Lamongan sebanyak 11 rawa dengan luas keseluruhan 6.878 Ha dan kapasitas volumenya sebesar 59.743.500 m³. Rawa juga sangat bermanfaat bagi masyarakat khususnya untuk kegiatan pertanian. Rawa yang ada di Kabupaten Lamongan dapat mengairi lahan sawah berpengairan sederhana seluas 9.934 Ha. Rawa terluas di Kabupaten Lamongan adalah waduk Jabung dengan luas 4.968 Ha dan volume 37.660.000 m³.

2.3.5 Kondisi Klimatologi

Kondisi klimatologi Kondisi klimatologi Kabupaten Lamongan dipengaruhi oleh kondisi suhu udara, kelembaban udara, kecepatan angin, tekanan udara, dan jumlah curah hujan serta kondisi penyinaran matahari. suhu udara maksimum di Kabupaten Lamongan pada tahun 2020 berada di antara 31,100C-33,600C, yang

mana tertinggi terjadi di bulan Maret dan suhu minimum berada pada 22,40oC-25,100C. Kemudian untuk kondisi kelembaban udara maksimum pada angka 92,00%-99,00% dan tertinggi terjadi pada bulan Maret serta kondisi minimum pada angka 60,00%- 69,00% yang terendah terjadi di bulan Oktober.

Selanjutnya terkait dengan kondisi kecepatan angin dan tekanan udara di Kabupaten Lamongan tahun 2020, kecepatan angin maksimum berkisar pada 5,66-12,35m/det dan tertinggi pada bulan Februari serta kecepatan angin minimum pada angka 0,00- 1,03m/det. Sedangkan untuk kondisi tekanan udara, maksimum antara 1010,80-1014,10mb yang tertinggi terjadi di bulan Juni dan kondisi minimum antara 1004,00-1006,20mb yang terendah terjadi pada bulan Desember.

Berdasarkan data dari BPS Kabupaten Lamongan pada tahun 2021, curah hujan di Kabupaten Lamongan pada tahun 2020 tertinggi terjadi di bulan Januari sebesar 589,90mm dan terendah di bulan September yaitu 2,00mm. Kemudian untuk jumlah hari hujan terbanyak pada bulan Desember yaitu 27 hari, dan paling sedikit pada bulan September sebanyak 5 hari. Kondisi penyinaran matahari tertinggi sebesar 8,97% di bulan Agustus, dan terendah sebesar 3,07% pada bulan Desember.

2.4 Bencana

Bencana merupakan sebuah keniscayaan yang mengiringi kehidupan umat manusia. Bencana alam tidak dapat dihindarkan, tetapi setidaknya dapat diupayakan agar risiko serta dampak yang ditimbulkan tidak menimbulkan banyak kerugian, baik korban jiwa maupun kerugian materiil dan non-materiil. Oleh karena

menyangkut kepentingan dan keselamatan masyarakat luas, kehadiran negara dalam tindakan penanggulangan mutlak diperlukan. Berikut adalah penjelasan tentang apa itu bencana dan jenis-jenis bencana :

2.4.1 Definisi Umum Bencana

Menurut UU nomor 24 Tahun 2007, bencana alam memiliki arti suatu peristiwa baik disebabkan oleh faktor manusia, faktor alam, maupun non-alam yang dapat mengancam serta mengganggu kehidupan masyarakat sehingga dapat menimbulkan kerugian berupa korban jiwa, harta benda, maupun lingkungan. Bencana.

Bencana didefinisikan sebagai suatu penyimpangan yang terjadi di lingkungan hidup yang terjadi secara tiba-tiba dan menyebabkan dampak terhadap manusia berupa kematian, dan gangguan kesehatan serta menyebabkan dampak terhadap struktur sosial dan kebutuhan masyarakat berupa bangunan gedung, infrastruktur pelayanan publik, makanan dan pelayanan sosial (Widyastuti 2014).

BNPB merinci 12 (dua belas) jenis bencana yang kerap kali terjadi di Indonesia, yaitu gempa bumi, tsunami, letusan gunung api, gerakan tanah (tanah longsor), banjir, banjir bandang, kekeringan, cuaca ekstrim (puting beliung), gelombang ekstrim dan abrasi, kebakaran hutan dan lahan, epidemi dan wabah penyakit, serta kegagalan teknologi (BNPB, 2014).

Berdasarkan data yang dihimpun oleh Pusat Data Informasi dan Komunikasi Kebencanaan (BNPB), pada bulan februari 2023 tercatat telah terjadi 295 kejadian bencana yang terjadi di seluruh Indonesia. Musim penghujan yang terjadi pada

bulan februari menjadikan bencana Hidrometeorologi banjir mendominasi dengan 125 kejadian, Cuaca ekstrim 96 kejadian, tanah longsor 53 kejadian dan abrasi dengan 4 kejadian. Sedangkan bencana Hidrometeorologi kering yang terjadi yaitu kebakaran hutan dan lahan sebanyak 12 kejadian. Selain korban jiwa, terdapat juga kerusakan pada rumah dan beberapa fasilitas umum. Pada bulan februari 2023 provinsi Jawa Tengah menjadi provinsi dengan kejadian bencana terbanyak yaitu 58 kejadian, disusul provinsi Jawa Barat dengan 53 kejadian, Jawa Timur dengan 38 kejadian, Sulawesi Selatan dengan 20 kejadian. Sementara itu, 4 provinsi dengan kejadian bencana paling sedikit antara lain provinsi Maluku Utara, Riau, Kalimantan Utara dan Jambi dengan masing-masing 1 kejadian di setiap provinsi (BNPB, 2023).

2.4.2 Jenis-Jenis Bencana

Berdasarkan beberapa penjelasan tentang definisi bencana diatas, maka secara sederhana menurut Undang-Undang No.24 Tahun 2007 menyebutkan bahwa bencana dapat dikelompokkan menjadi tiga macam bencana yaitu bencana alam, bencana non-alam, dan bencana sosial. Jenis-jenis bencana tersebut antara lain :

- a. Bencana alam, merupakan suatu peristiwa yang disebabkan oleh faktor alam seperti banjir, tanah longsor, gempa bumi dan lain sebagainya. Bencana banjir dapat terjadi karena meluapnya air hujan sehingga debit air meningkat dan drainase tidak dapat menampung lagi sehingga air meluap, bencana seperti diatas tidak dapat diprediksi secara pasti.
- b. Bencana non-alam, merupakan suatu peristiwa yang terjadi akibat faktor non-alam seperti kegagalan teknologi, epidemic dan wabah penyakit.

- c. Bencana sosial, merupakan bencana yang diakibatkan oleh peristiwa antar manusia seperti kerusuhan dan konflik antar kelompok masyarakat.

2.5 Banjir

Banjir dapat berupa genangan pada lahan yang biasanya kering seperti pada lahan pertanian, permukiman, pusat kota. Banjir juga dapat terjadi karena debit/volume air yang mengalir pada sungai atau saluran drainase melebihi kapasitas pengalirannya. Berikut adalah penjelasan tentang apa itu banjir beserta faktor-faktor yang mengakibatkan banjir :

2.5.1 Definisi Umum Banjir

Menurut UU No.24 Tahun 2007 Banjir merupakan suatu peristiwa yang terjadi dengan ciri-ciri terendamnya suatu daerah yang diakibatkan oleh volume air yang mengalami peningkatan. Banjir dapat disebabkan oleh beberapa faktor seperti faktor perubahan iklim, peningkatan frekuensi dan intensitas curah hujan yang tinggi atau karena akibat banjir kiriman dari daerah lain yang lebih tinggi (Soliha 2020). Penyebab banjir di Indonesia umumnya terjadi akibat penebangan hutan tanpa adanya reboisasi, pembuangan sampah di aliran sungai, serta pembangunan rumah di bantaran sungai (Prasetyo dan Hariyanto, 2017). Banjir merupakan bencana alam yang paling sering terjadi, baik dilihat dari intensitasnya pada suatu tempat maupun jumlah lokasi kejadian dalam setahun yaitu sekitar 40% di antara bencana alam yang lain (Hasan 2015).

Di Indonesia banjir sudah lama terjadi. Di Jakarta, misalnya, banjir sudah terjadi sejak 1959, ketika jumlah penduduk masih relatif sedikit. Menurut (Matt

McGrath, 2021) dalam laman *Global Flood Database*, Indonesia tidak termasuk dalam negara-negara yang mengalami peningkatan bencana banjir pada 2030. Namun, dalam laman tersebut, beberapa wilayah di Indonesia tercatat menjadi wilayah yang terdampak banjir parah akibat intensitas hujan yang tinggi dalam 20 tahun terakhir.

Kepala BMKG Dwikorita menyatakan bahwa pada tahun 2023 semua pihak harus bersiap untuk menghadapi terjangan bencana hidrometeorologi akibat tingginya curah hujan tahunan 2023 yang diperkirakan melebihi rata-ratanya atau melebihi batas normalnya di sebagian wilayah Indonesia. Berdasarkan *Climate Outlook 2023* yang diterbitkan BMKG, beberapa wilayah yang diprediksi berpotensi mendapatkan curah hujan tahunan yang cukup tinggi, yaitu 2500 mm/tahun, terjadi di wilayah Sumatra, Jawa Barat, Jawa Tengah, Sebagian besar Kalimantan, sebagian Sulawesi Barat, Sulawesi Selatan dan sebagian besar Papua. Sedangkan daerah yang diprediksi mengalami hujan tahunan di atas normal adalah sebagian kecil Jambi bagian selatan, sebagian kecil Jawa Barat bagian utara, sebagian kecil Jawa Timur bagian timur, sebagian kecil Kalimantan Timur bagian selatan, sebagian kecil Bali bagian utara, sebagian NTB dan sebagian kecil Sulawesi Tengah bagian timur (BMKG, 2023).

Menurut Yode Arliando, Indra Kanedi dan Rahma Wayan Lestari dalam penelitiannya yang berjudul “Sistem informasi geografis (SIG) daerah rawan banjir di kota Bengkulu menggunakan arcview” menyatakan bahwa banjir adalah genangan air yang mengalir deras dengan ketinggian melebihi tingkat normal. Pada saat banjir, air akan menggenangi sebagian besar daratan yang biasanya tidak

tergenangi air. Ketika musim hujan tiba, sebagian wilayah ada yang mengalami banjir. Curah hujan yang tinggi membuat beberapa wilayah tergenang air. Banjir yang sangat berbahaya adalah banjir dahsyat yang terjadi secara tiba-tiba dan bersifat menghayutkan (Rahma dan Indra, 2016).

2.5.2 Faktor-Faktor Penyebab Banjir

Menurut Hasan dalam penelitiannya menyebutkan bahwa banjir dapat dipengaruhi oleh beberapa faktor antara lain : (Hasan 2015)

a. Intensitas Curah Hujan Tahunan

Daerah yang memiliki tebal hujan tinggi maka daerah tersebut akan lebih berpengaruh terhadap kejadian banjir. Oleh karena itu, semakin tinggi tebal curah hujan, maka tingkat kerawanan semakin tinggi.

b. Kemiringan Lereng

Kemiringan lereng merupakan perbandingan antara selisih ketinggian dengan jarak datar pada dua tempat yang dinyatakan dalam persen. Semakin tinggi kemiringan lahan, semakin tinggi pula kecepatan air yang diteruskan oleh lahan tersebut ke tempat yang lebih rendah. Oleh karena itu, semakin besar derajat kemiringan lahan, maka tingkat kerawanan banjir semakin kecil.

c. Penggunaan Lahan

Penggunaan lahan berkaitan dengan kegiatan manusia dalam memanfaatkan lahan untuk tujuan tertentu. Penggunaan lahan oleh manusia seperti untuk pemukiman, sawah, irigasi, lahan industri dan lain sebagainya. Lahan yang banyak ditanami oleh vegetasi, maka air hujan akan banyak diinfiltrasi dan

lebih banyak waktu yang akan ditempuh sebelum sampai ke sungai, sehingga kemungkinan banjir lebih kecil daripada daerah yang tidak ditanami vegetasi.

d. **Tekstur Tanah**

Tanah yang memiliki tekstur halus memiliki peluang terjadinya banjir yang tinggi. Sebaliknya, tekstur yang kasar memiliki peluang kejadian banjir yang rendah. Hal ini dikarenakan semakin halus tekstur tanah, semakin sulit air pada permukaan tanah yang berasal dari hujan maupun luapan sungai untuk meresap ke dalam tanah, sehingga terjadi penggenangan. Oleh karena itu, semakin halus tekstur suatu tanah, tingkat kerawanan banjir semakin tinggi.

2.5.3 Dampak-Dampak Banjir

Bencana banjir yang terjadi memiliki dampak yang dapat dirasakan secara langsung maupun tidak langsung (Soliha 2020). Dampak yang dirasa secara langsung lebih mudah diprediksi. Dampak yang dialami oleh penduduk perkotaan yang cenderung didominasi oleh pemukiman akan berbeda dengan dampak pada daerah pedesaan yang cenderung didominasi oleh lahan produktif. Banjir adalah genangan air yang mengalir deras dengan ketinggian melebihi tingkat normal (Rahma dan Indra, 2016). Banjir yang terjadi di suatu wilayah dapat mengakibatkan kerusakan pada beberapa bangunan hingga menimbulkan korban jiwa serta berdampak terhadap lumpuhnya kegiatan pengoperasian transportasi umum seperti bus, truk, kereta api dan bahkan penerbangan pesawat (Rosyidie 2013).

Menurut BNPB pada tahun 2013 kerugian yang dialami akibat bencana banjir menduduki urutan kedua setelah gempa bumi atau tsunami (BNPB, 2013). Oleh karena itu, banjir didefinisikan sebagai bencana yang relatif menimbulkan kerugian

yang tidak sedikit. Selain kerugian fisik, ada juga kerugian non-fisik yang dialami oleh masyarakat yang terdampak bencana banjir, misalnya terganggunya kegiatan belajar mengajar, terhambatnya arus transportasi, meningkatnya harga kebutuhan pokok, serta menyebabkan banyaknya korban jiwa. Selain itu banjir juga mengakibatkan kerugian secara materi serta menimbulkan terganggunya pertumbuhan suatu wilayah (Risanty, Arisanty, and Alviawati 2015).

Dampak yang ditimbulkan akibat dari bencana banjir antara lain rusaknya lahan permukiman dan perdagangan, kerusakan pada lahan pertanian, kerusakan jalur transportasi, sarana komunikasi, serta sistem resapan air (Yunida, Kumalawati, and Arisanty 2017).

2.6 Algoritma K-Means

K-means adalah metode yang mengalokasikan nilai *cluster* (k) secara acak (Ananda, 2023). Untuk sementara, nilainya berperan sebagai *centroid*, atau pusat *cluster*. Kemudian, dengan menggunakan metode tersebut, tentukan jarak antara setiap potongan data yang ada dan setiap *centroid* hingga nilai *centroid* tetap konstan. K-Means merupakan algoritma dalam *data mining* yang biasa digunakan dalam mengelompokkan atau *clustering* data (Witten et al., 2005). Selain itu, Algoritma K-Means adalah metode untuk mengklasifikasikan n objek berdasarkan atributnya ke dalam k partisi berdasarkan kriteria yang akan ditentukan dari data yang luas dan beragam (Buslim et al., 2020). Banyak sekali pendekatan yang dilakukan untuk membuat suatu *cluster*, diantaranya seperti membuat aturan yang mendikte keanggotaan dalam grup sama berdasarkan tingkat persamaan antara

anggotanya atau dengan membuat kumpulan fungsi yang mengukur beberapa properti dari pengelompokan tersebut sebagai parameter sebuah *clustering*. *K-Means* merupakan algoritma dengan metode *Unsupervised Learning* karena menarik kesimpulan dari dataset dengan *input data labeled response* (Pradono Iswara, et al., 2019). Tahapan K-Means adalah sebagai berikut : (Rohmah et al., 2021)

1. Menentukan jumlah *cluster* yang digunakan

Penentuan jumlah *cluster* ini harus tidak lebih banyak dari jumlah kriteria yang ada.

2. Menentukan titik pusat *cluster* (*centroid*)

Jika akan dibuat 3 *cluster*, penentuan titik pusat *cluster* atau *centroid* ini diambil dari nilai minimal, rata-rata dan maksimal sedangkan bila ingin membuat 2 kluster penentuan titik pusat *cluster* atau *centroid* hanya diambil dari nilai minimal dan maksimal pada setiap kriteria.

3. Menghitung jarak titik data objek dengan titik data pusat (*centroid*)

Pada tahap ini dapat dilakukan menggunakan rumus *Euclidean Distance* yaitu :

Keterangan :

$$d(x, y) = \|x - y\|^2 = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

2.1

d = determinan (*Euclidean Distance*)

x = titik pusat cluster

$y = \text{data}$

$n = \text{jumlah data}$

$i = \text{data ke- (atribut data)}$

4. Mengelompokkan data objek untuk menentukan anggota *cluster* berdasarkan jarak minimum

Setelah menghitung jarak, diambil nilai minimum yang diberikan nilai 1, 2 dan 3, dimana nilai 1 untuk data yang ditempatkan pada *cluster* dan nilai 0 untuk data yang ditempatkan ke *cluster* yang lain hingga membentuk suatu matriks.

5. Mengulangi tahap 2 hingga 4 sampai tidak ada pergerakan *centroid* lagi.

Pengulangan akan terus dilakukan hingga nilai pada tiap *cluster* tidak berubah tempat.

Semua tahapan tersebut sesuai dengan pengaplikasian K-Means clustering pada penelitian ini.

Dalam *Supervised learning*, label dari data sudah diketahui sehingga evaluasi dapat dilakukan dengan menghitung tingkat kebenaran dengan membandingkan nilai-nilai yang diprediksi dengan label-label tersebut. Namun, dalam *Unsupervised Learning*, label-label tidak diketahui, yang membuat sulit untuk mengevaluasi tingkat kebenarannya karena tidak ada acuan yang jelas.

Meskipun demikian, secara konsisten, algoritma *clustering* yang baik memiliki variasi yang kecil dalam kelompok (titik data dalam kelompok memiliki

kesamaan satu sama lain) dan variasi antar kelompok yang besar. Dalam artikel yang ditulis oleh Kay Jan Wong ada dua jenis matriks evaluasi untuk pengelompokan, antar lain : (Kay Jan Wong, 2022)

1. Metrik Ekstrinsik : metrik-metrik ini membutuhkan label-label acuan yang mungkin tidak tersedia dalam praktiknya.
2. Metrik Insintrik : metrik -metrik ini tidak memerlukan label-label acuan (dapat diterapkan pada semua hasil pembelajaran tanpa pengawasan)

Metrik evaluasi eksintrik membutuhkan label-label acuan sebagai pembanding untuk mengevaluasi kualitas pengelompokan. Namun, dalam kebanyakan kasus, label-label acuan ini tidak tersedia atau tidak dapat diperoleh dengan mudah. Oleh karena itu, dalam evaluasi *Unsupervised Learning*, metrik-metrik insintrik digunakan sebagai alternatif. Metrik-metrik insintrik ini melibatkan pengukuran kualitas pengelompokan berdasarkan struktur internal data yang diamati tanpa memerlukan label acuan. Contoh metrik intrinsik yang umum digunakan adalah *Silhouette Coefficient*.

2.7 *Silhouette Coefficient*

Metode ini merupakan metode evaluasi *clustering* yang menggabungkan metode *cohesion* dan *separation*. *Cohesion* diukur dengan menghitung seluruh objek yang terdapat dalam sebuah *cluster* dan *separation* diukur dengan menghitung jarak rata-rata setiap objek dalam sebuah *cluster* dengan *cluster* terdekatnya (Solmin Paembonan, 2021). Jarak antar data dihitung dengan menggunakan rumus *Euclidean Distance*. Untuk menyediakan informasi tentang

kualitas hasil *clustering* pada proses *clustering*, dapat dihitung *silhouette* dari masing-masing *cluster* bahkan keseluruhan *cluster* dari hasil kerja satu algoritma *clustering* (Wira, 2019). Untuk melihat kualitas hasil pengelompokan data, maka perlu dilakukan uji homogenitas (Rahmatina, 2021). Pengujian dilakukan setelah mencapai kovergensi 0 dimana hasil pengelompokan terakhir sama dengan pengelompokan sebelumnya. Dengan kata lain, tidak ada data yang berpindah *cluster*. Langkah dalam menghitung *Silhouette coefficient* dimulai dengan mencari jarak rata-rata data ke-*i* dengan semua data di klaster yang sama, disini diasumsikan bahwa data ke-*i* berada di *cluster* A. Rumus dari $a(i)$ ditulis dalam persamaan (2.2).

$$a(i) = \frac{1}{|A| - 1} \sum_{j \in A, j \neq i} d(i, j)$$

2.2

Dimana :

A = banyaknya data di *cluster* A

Selanjutnya menghitung nilai $b(i)$ yang merupakan nilai minimum dari jarak rata-rata data ke-*i* dengan semua data di *cluster* berbeda. Diasumsikan bahwa *cluster* berbeda selain A adalah *cluster* C. Maka perhitungan jarak rata-rata data ke-*i* dengan semua data di *cluster* C ditulis sebagai berikut.

$$d(i, C) = \frac{1}{|C|} \sum_{j \in C} d(i, j)$$

2.3

Dimana :

C = banyaknya data di *cluster* C

Setelah menghitung $d(i,C)$ untuk semua *cluster* $C \neq A$, selanjutnya memilih nilai jarak paling minimum sebagai nilai $b(i)$.

$$b(i) = \min_{C \neq A} d(i,j)$$

2.4

Jika *cluster* B memiliki nilai jarak minimum, maka $d(i,B) = b(i)$ yang disebut sebagai tetangga dari data ke- i dan merupakan *cluster* terbaik kedua untuk data ke- i setelah *cluster* A . Setelah $a(i)$ dan $b(i)$ diketahui, maka proses terakhir adalah menghitung *silhouette coefficient*.

$$s(i) = \frac{b(i) - a(i)}{\max \{a(i) - b(i)\}}$$

2.5

Nilai $s(i)$ berada diantara -1 dan 1, dimana setiap nilai diinterpretasi sebagai berikut :

$s(i) \approx 1$ = data ke- i digolongkan dengan baik (dalam A)

$s(i) \approx 0$ = data ke- i berada ditengah antara dua *cluster* (A dan B)

$s(i) \approx -1$ = data ke- i digolongkan dengan lemah (dekat ke *cluster* B daripada A)

penafsiran nilai *Silhouette Coefficient* ditunjukkan pada tabel berikut :

Tabel 2.3 Interpretasi Nilai *Silhouette Coefficient*

<i>Silhouette Coefficient</i>	Interpretasi
0.71 – 1.00	Struktur kuat
0.51 – 0.70	Struktur sedang
0.26 – 0.50	Struktur lemah
< 0.26	Tidak terstruktur

Nilai *silhouette* untuk keseluruhan data dengan jumlah *cluster* k , dapat didefinisikan sebagai $sil(k)$ yaitu rata-rata *silhouette value* untuk semua *cluster* yang dihitung dengan persamaan

$$sil(c) = sil(k) \frac{1}{|k|} \sum_{i=1}^k sil(c_i)$$

2.6

Dimana :

$Sil(k)$: nilai *silhouette* semua *cluster*

$|k|$: banyaknya *cluster* k

$Sil(c_i)$: rata-rata nilai *silhouette*

2.8 *Elbow Method*

Metode *Elbow* merupakan suatu metode yang digunakan untuk menghasilkan informasi dalam menentukan jumlah *cluster* terbaik dengan cara melihat presentase hasil perbandingan antara jumlah *cluster* yang akan membentuk siku pada suatu titik (Rahman 2017). Selain dari presentasenya, untuk menentukan jumlah *cluster*

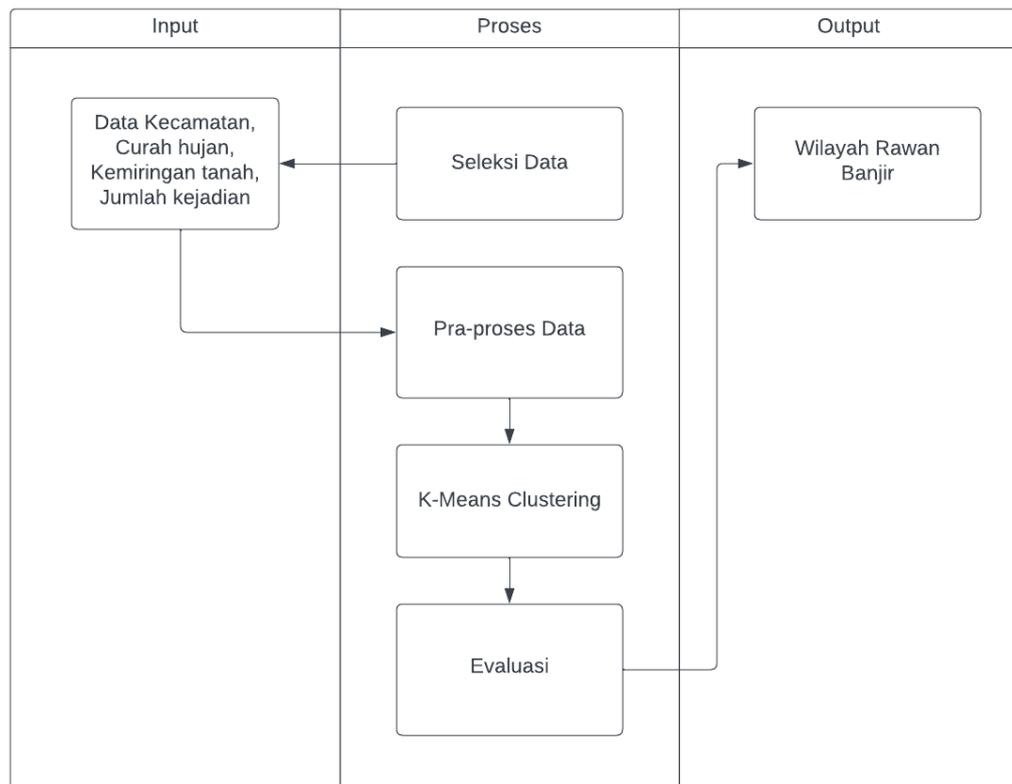
yang optimal dengan metode ini juga dapat dilihat dari hasil grafik yang dihasilkan dimana grafik akan membentuk suatu sudut siku ataupun mendekati sudut siku (Atif, 2023). Hasil presentase yang berbeda dari setiap nilai cluster dapat ditunjukkan dengan menggunakan grafik sebagai sumber informasinya. Dimana jika nilai *cluster* pertama dengan nilai cluster kedua memberikan sudut dalam grafik atau nilainya mengalami penurunan paling besar maka nilai *cluster* tersebut adalah yang terbaik (Maria, 2023).

BAB III

METODOLOGI PENELITIAN

3.1 Desain Sistem

Desain sistem, merupakan implementasi sistem yang akan dirancang. Pentingnya terdapat desain sistem yaitu untuk mengetahui alur dari sebuah sistem yang akan dioperasikan. Data-data yang akan diolah adalah data yang digunakan untuk menentukan wilayah mana yang rawan akan banjir. Data tersebut kemudian akan diolah dengan menggunakan metode *K-Means Clustering* hingga didapatkan hasil outputan wilayah rawan banjir di kabupaten Lamongan.



Gambar 3.1. Desain Sistem Wilayah Rawan Banjir

3.1.1 Seleksi Data

Pada tahap ini penulis melakukan *selection* untuk mengumpulkan atau mencari dataset yang pas untuk dijadikan objek dari penelitian ini, penulis memakai sumber yang *free open source* yaitu website Badan Pusat Statistik dan BMKG Kabupaten lamongan. Data yang digunakan pada penelitian ini adalah kumpulan data selama 3 tahun terhitung mulai tahun 2020 hingga 2022 yang berjumlah 27 baris dengan 4 variabel yaitu Kecamatan, Curah hujan, kemiringan tanah, dan jumlah kejadian. Proses seleksi data dalam website open source sampai menemukan waktu yang pas ini dilakukan selama 1 bulan.

Data yang digunakan dalam penelitian adalah data setiap tahun dari variabel kecamatan, curah hujan, kemiringan tanah, dan jumlah kejadian banjir yang ada di kabupaten Lamongan. Berikut adalah data dari parameter yang akan digunakan dalam pengujian dan dataset.

Tabel 3.1. Tabel Data set (Cluster Kabupaten Lamongan)

No	Kecamatan	Curah Hujan	Kemiringan Tanah	Jumlah Kejadian
1	Sukorame	884	10	35
2	Bluluk	902	15	7
3	Ngimbang	804	16	5
4	Sambeng	793	15	6
5	Mantup	707	13	4
6	Kembangbahu	674	13	5
7	Sugio	688	11	5
8	Kedungpring	655	11	16
9	Modo	700	12	11
10	Babat	719	8	17
11	Pucuk	716	8	6
12	Sukodadi	693	5	13
13	Lamongan	679	6	14

14	Tikung	711	4	23
15	Sarirejo	711	8	6
16	Deket	709	7	15
17	Glagah	750	5	32
18	Karangbinangun	630	5	36
19	Turi	664	6	29
20	Kalitengah	654	6	38
21	Karanggeneng	771	8	27
22	Sekaran	667	6	7
23	Maduran	615	5	11
24	Laren	686	5	21
25	Solokuro	693	4	6
26	Paciran	596	4	9
27	Brondong	685	6	8

Pada tabel 3.1 data yang digunakan untuk dataset adalah jumlah data keseluruhan kecamatan 27 data. Adapun sumber data berasal dari BMKG dan Badan Pusat Statistik Kabupaten Lamongan, pada tahun 2020 - 2022.

3.1.2 Praproses Data

Tahap sebelum pemrosesan data dikenal dengan praproses atau *preprocessing data*. *Preprocessing* data terbagi menjadi tiga langkah yaitu *Integration* untuk menggabungkan data dari beberapa sumber, *Selection* untuk memilih atribut data yang akan digunakan dan yang terakhir adalah *Cleaning* yang bertujuan untuk menghilangkan *noise* dan data yang tidak relevan (Rante Rerung, 2018). *Preprocessing* ini dilakukan dengan waktu yang cukup singkat kurang dari tiga hari agar data yang akan digunakan sesuai dengan kebutuhan. pada penelitian ini, tahapan *preprocessing* data yang dilakukan adalah sebagai berikut :

1. Integrasi Data

Integrasi data dilakukan pada data yang didapat dari dua sumber yang berbeda. Seperti pada penelitian ini, data diperoleh dari dua sumber berbeda yaitu Jumlah Kejadian, dan kemiringan tanah dari Badan Pusat Statistik. Sedangkan curah hujan dari website BMKG Kabupaten Lamongan.

2. Seleksi Atribut

Dari data yang diambil dari Badan Pusat Statistik dan BMKG Kabupaten Lamongan, dipilih atribut yang sesuai untuk kebutuhan *clustering*. Adapun atribut yang tidak digunakan dalam perhitungan adalah nama kecamatan. Sehingga atribut yang digunakan adalah curah hujan, kemiringan tanah, dan jumlah kejadian.

3. Normalisasi Data

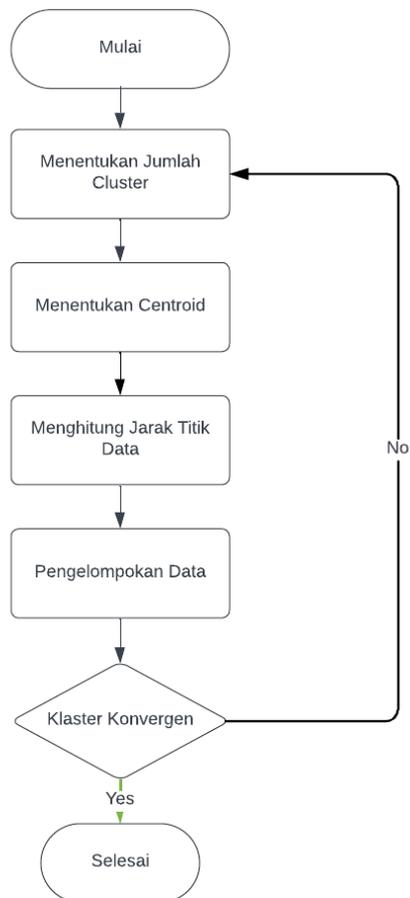
Normalisasi data dilakukan pada data numerik agar skala data yang digunakan sama, dengan normalisasi atau *scalling* setiap variabel akan memiliki rentang nilai yang serupa sehingga setiap variabel dapat memiliki kontribusi yang sama dalam proses *clustering*. Pada tahap ini digunakan teknik *standard scalling* yang mengubah skala data menjadi memiliki mean 0 dan standard deviation 1, sehingga data memiliki rentang serupa dan dapat dibandingkan secara langsung.

3.1.3 *K-Means Clustering*

Dalam tahap ini, *Clustering* dilakukan dengan menggunakan data input dari variabel yang telah ditentukan sebelumnya. Proses *clustering* dilakukan dengan

menggunakan metode *K-Means*, untuk menentukan jumlah *cluster* terbaik yang akan dipakai, digunakan metode *Euclidean Distance*.

Setelah dilakukan analisis data, selanjutnya menerapkan metode pada sistem, yaitu metode *K-Means*. Berikut adalah Langkah-langkah penerapan metode *K-Means* :



Gambar 3.2 Desain alur algoritma K-means

Metode *K-Means* dimulai dari menentukan jumlah *cluster* yang akan digunakan, Menentukan nilai *centroid* (*K*). Jika sudah menentukan *centroid* awal,

selanjutnya menghitung nilai rata-rata data yang ada pada masing-masing *cluster*, kemudian mengelompokkan data berdasarkan kedekatan terhadap titik pusat *cluster*. Berikut adalah langkah-langkah algoritma hasil *clustering* dengan metode *K-Means* :

1. Menentukan jumlah *cluster* yang akan digunakan

Pada contoh kasus ini jumlah *cluster* yang akan digunakan adalah 2

Tabel 3.2. *Cluster* Yang Digunakan

Cluster	Keterangan
1	Sering Banjir
2	Sedikit Banjir

2. Menentukan titik pusat *cluster* (*centroid*).

Pusat awal *cluster* didapatkan secara random. Untuk penentuan *centroid* awal pada contoh kasus ini adalah data dengan nilai terbesar dan terkecil. Data yang dipilih menjadi pusat *cluster* (*centroid*) awal adalah data ke-1 dan ke-5.

Centroid 1

Ki1 = 884

Ki2 = 10

Ki3 = 35

Centroid 2

Ki1 = 707

Ki2 = 13

Ki3 = 4

Tabel 3.3. Tabel *Centroid* Awal

Data ke-	Ci1	Ci2	Ci3
1	884	10	35
2	902	15	7

3	804	16	5
4	793	15	6
5	707	13	4
6	674	13	6

3. Menghitung jarak titik data berdasarkan titik *centroid* terdekat.

Untuk menentukan jarak terpendek antara setiap titik data dan setiap *centroid*, perhitungan jarak antara setiap titik data masukan dan setiap *centroid* dilakukan menggunakan algoritma jarak *Euclidean*.

$$d(x, y) = \|x - y\|^2 = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

3.1

Keterangan :

d = determinan (*Euclidian Distance*)

x = titik pusat *cluster*

y = data

n = jumlah data

i = data ke- (atribut data)

Rumus *Euclidean Distance*

$$\text{Data} = ((C_{i1} - K_{i1})^2 + (C_{i2} - K_{i2})^2 + (C_{i3} - K_{i3})^2)$$

Hitung *Euclidean distance* dari semua data pada *Cluster 1* iterasi 1

$$\text{Data ke-1} = \sqrt{((884 - 884)^2 + (10 - 10)^2 + (35 - 35)^2)}$$

$$= 0$$

$$\begin{aligned} \text{Data ke-2} &= \sqrt{((902 - 884)^2 + (15 - 10)^2 + (7 - 35)^2)} \\ &= 33,66007 \end{aligned}$$

$$\begin{aligned} \text{Data ke-3} &= \sqrt{((804 - 884)^2 + (16 - 10)^2 + (5 - 35)^2)} \\ &= 85,65045 \end{aligned}$$

$$\begin{aligned} \text{Data ke-4} &= \sqrt{((793 - 884)^2 + (15 - 10)^2 + (6 - 35)^2)} \\ &= 95,63995 \end{aligned}$$

$$\begin{aligned} \text{Data ke-5} &= \sqrt{((707 - 884)^2 + (13 - 10)^2 + (4 - 35)^2)} \\ &= 179,7192 \end{aligned}$$

$$\begin{aligned} \text{Data ke-6} &= \sqrt{((674 - 884)^2 + (13 - 10)^2 + (6 - 35)^2)} \\ &= 212,1532 \end{aligned}$$

Hitung *Euclidean distance* dari semua data pada *Cluster 2* iterasi 1

$$\begin{aligned} \text{Data ke-1} &= \sqrt{((884 - 293)^2 + (10 - 13)^2 + (35 - 4)^2)} \\ &= 179,7192 \end{aligned}$$

$$\begin{aligned} \text{Data ke-2} &= \sqrt{((902 - 293)^2 + (15 - 13)^2 + (35 - 4)^2)} \\ &= 195,0333 \end{aligned}$$

$$\begin{aligned} \text{Data ke-3} &= \sqrt{((804 - 293)^2 + (16 - 13)^2 + (5 - 4)^2)} \\ &= 97,05153 \end{aligned}$$

$$\begin{aligned} \text{Data ke-4} &= \sqrt{((793 - 293)^2 + (15 - 13)^2 + (6 - 4)^2)} \\ &= 86,0465 \end{aligned}$$

$$\begin{aligned} \text{Data ke-5} &= \sqrt{((707 - 293)^2 + (13 - 13)^2 + (4 - 4)^2)} \\ &= 0 \end{aligned}$$

$$\begin{aligned} \text{Data ke-6} &= \sqrt{((674 - 293)^2 + (13 - 13)^2 + (6 - 4)^2)} \\ &= 33,01515 \end{aligned}$$

Tabel 3.4. Jarak Titik Data Dengan *Centroid* Iterasi 1

K1	K2
0	179,7192
33,66007	195,0333
85,65045	97,05153
95,63995	86,0465
179,7192	0
212,1532	33,01515

4. Mengelompokkan data objek untuk menentukan anggota *cluster* berdasarkan jarak minimum

Untuk menentukan data tergabung dalam kelompok *cluster* mana, maka diambil nilai minimum data dan diberi nilai 1 untuk cluster 1, dan 2 untuk cluster 2

Iterasi ke-1

Data ke-1 = 1

Data ke-2 = 1

Data ke-3 = 1

Data ke-4 = 2

Data ke-5 = 2

Data ke-6 = 2

Anggota C1 = {data ke-1, data ke-2, data ke-3}

Anggota C2 = {data ke-4, data ke-5, data ke-6}

5. Menghitung nilai *centroid* iterasi selanjutnya

Untuk mencari nilai *centroid* baru dilakukan dengan menghitung nilai rata-rata tiap variabel pada masing-masing *cluster* yang telah ditentukan

$$C_k = \left(\frac{1}{n_k} \right) \sum d_1$$

3.2

Dimana n_k adalah jumlah data dalam *cluster* k dan d_1 adalah data dalam *cluster*

k. Sehingga didapatkan titik pusat *cluster* yang baru yaitu :

Cluster 1 iterasi ke-1

$$C_{i1} \Rightarrow (884 + 902 + 804) / 6 = 863$$

$$C_{i2} \Rightarrow (10 + 15 + 16) / 6 = 13,7$$

$$C_{i3} \Rightarrow (35 + 7 + 5) / 6 = 15,7$$

Cluster 2 iterasi ke-1

$$C_{i1} \Rightarrow (793 + 707 + 674) / 6 = 725$$

$$C_{i2} \Rightarrow (15 + 13 + 13) / 6 = 13,7$$

$$C_{i3} \Rightarrow (6 + 4 + 5) / 6 = 5$$

6. Mengulangi tahap (3.3) hingga (3.5) sampai tidak ditemukan adanya perubahan pada nilai titik data setiap *cluster*.

Menghitung jarak pusat *cluster* menggunakan *euclidean distance* pada titik pusat *cluster* yang baru. Kemudian membandingkan hasil tersebut dengan hasil iteasi sebelumnya. Jika hasil posisi *cluster* pada iterasi ke-2 sama dengan posisi

iterasi pertama, maka proses dihentikan, namun jika tidak maka proses dilanjutkan ke iterasi ke-3.

Iterasi ke-2

Tabel 3.5 Hasil Iterasi 2

K1	K2
28,53653	162,1745
39,64846	177,3496
60,32965	79,36764
71,00704	68,35366
156,7695	17,7075
189,6347	50,67105

Anggota C1 = {data ke-1, data ke-2, data ke-3}

Anggota C2 = {data ke-4, data ke-5, data ke-6}

Dikarenakan hasil posisi *cluster* pada iterasi ke-2 sama dengan iterasi sebelumnya, maka proses iterasi dihentikan. Dan diperoleh hasil bahwa data ke-1 dan data ke-2 termasuk ke dalam kategori wilayah dengan tingkat kerawanan sering. Sedangkan data ke-3, ke-4, ke-5 dan ke-6 termasuk ke dalam kategori wilayah dengan tingkat kerawanan sedikit.

3.1.4 Evaluasi Hasil

Berbeda dengan *supervised learning* yang memiliki “*ground truth*” untuk mengevaluasi kinerja model, analisis *clustering* tidak memiliki metrik evaluasi

yang solid yang dapat digunakan untuk mengevaluasi hasil dari berbagai algoritma pengelompokan. Karena algoritma k-means yang umum digunakan dalam *clustering* membutuhkan jumlah *cluster* (k) sebagai input, namun tidak ada jawaban yang pasti tentang jumlah *cluster* yang ideal dalam suatu masalah. Dalam metodologi prediksi *cluster*, kita dapat mengevaluasi seberapa baik kinerja model berdasarkan jumlah *cluster* (nilai k) yang ditentukan sehingga dalam penelitian ini, evaluasi akan dilakukan pengujian menggunakan *Elbow Method*, pengujian ini dipilih karena Penerapan *metode elbow* dalam menentukan jumlah *cluster* terbaik pada proses *clustering* K-Medoid menunjukkan nilai DBI yang lebih tinggi dengan nilai DBI sebesar 1,10 (Dewa Ayu, 2019). Pengujian dilakukan dengan penentuan beberapa nilai k dengan memilih nilai *Elbow* yang terbaik dari nilai k sebagai input jumlah *cluster*. Di dalam penelitian ini, tahap ini akan masuk ke dalam implementasi *K-Means Clustering*.

Untuk evaluasi hasil *clustering*, penelitian ini menggunakan metode *silhouette coefficient*. Berikut adalah rancangan nilai *silhouette* dengan percobaan $k=2$ dan $k=3$. Data yang digunakan adalah data yang sama dengan rancangan *K-means clustering* sebelumnya:

Percobaan $k=2$

Tabel 3.6 Hasil Percobaan Nilai *silhouette* Untuk $k=2$

a(i)	d(i,1)	d(i,2)	b(i)	S(i)
59,6553	28,53653	162,1745	28,5365	-0,5216
65,8428	39,64846	177,3496	39,6485	-0,3978
91,8380	60,32965	79,36764	60,3296	-0,3431
102,5338	71,00704	68,35366	68,3537	-0,3334
59,5308	156,7695	17,7075	17,7075	-0,7025

76,0181	189,6347	50,67105	50,6711	-0,3334
---------	----------	----------	---------	---------

Dari tabel diatas diperoleh jumlah nilai rata-rata *silhouette* dengan menggunakan persamaan 2.2 yaitu :

Rata-rata Silhouette = 1/jumlah data x jumlah nilai *silhouette* seluruh data

$$= 1/6 \times -2,6319$$

$$= -0,7520$$

Percobaan k=3

Tabel 3.7 Hasil Percobaan Nilai *silhouette* Untuk k=3

a(i)	d(i,1)	d(i,2)	d(i,3)	b(i)	S(i)
33,66007	28,53653	90,61319	195,912	28,5365	-0,1522
33,66007	39,64846	103,5121	211,5242	39,6485	0,1510
11,09054	60,32965	5,545268	113,5407	5,5453	-0,5000
11,09054	71,00704	5,545268	102,5305	5,5453	-0,5000
33,01515	156,7695	91,54644	16,50757	16,5076	-0,5000
33,01515	189,6347	124,5261	16,50757	16,5076	-0,5000

Dari tabel diatas diperoleh jumlah nilai rata-rata *silhouette* dengan menggunakan persamaan 2.2 yaitu :

Rata-rata Silhouette = 1/jumlah data x jumlah nilai *silhouette* seluruh data

$$= 1/6 \times -2,0012$$

$$= -0,5718$$

Dari kedua percobaan diatas dapat dilihat bahwa pada percobaan k=2 memperoleh nilai *silhouette* sebesar -0,7520 sedangkan pada percobaan k=3 memperoleh nilai *silhouette* sebesar -0,5718. Sehingga dapat disimpulkan bahwa

nilai $k=3$ adalah nilai k yg lebih baik daripada nilai $k=2$ dengan nilai silhouette - 0,5718.

BAB IV

HASIL DAN PEMBAHASAN

4.1 Hasil

Pada sub-bab ini dilakukan implementasi sistem menggunakan metode K-Means. Pada tahap ini terdiri dari beberapa tahapan yaitu Akuisisi data, Praproses data, *Clustering*, Penentuan nilai K terbaik menggunakan *elbow method* dan visualisasi hasil *clustering*.

4.1.1 Akuisisi Data

Penelitian ini dilakukan untuk mendapatkan data wilayah yang sesuai untuk dijadikan input dalam pemetaan wilayah rawan banjir yang diharapkan dapat membantu masyarakat setempat untuk lebih waspada dan siaga akan datangnya bencana banjir. Langkah pertama yang dilakukan penulis adalah akuisisi data wilayah dari website BPS dan BMKG Kabupaten Lamongan yang sesuai sebagai simulasi untuk dilakukan segmentasi wilayah. Data yang digunakan dalam pengujian adalah data setiap tahun dari 3 variabel yaitu Curah Hujan, Kemiringan Tanah, dan Jumlah kejadian banjir yang ada di Kabupaten Lamongan. Berikut adalah data dari parameter yang akan digunakan dalam pengujian :

Tabel 4.1 Dataset Penelitian

No	Kecamatan	Curah Hujan	Kemiringan Tanah	Jumlah Kejadian
1	Sukorame	884	10	35
2	Bluluk	902	15	7
3	Ngimbang	804	16	5
4	Sambeng	793	15	6
5	Mantup	707	13	4

6	Kembangbahu	674	13	5
7	Sugio	688	11	5
8	Kedungpring	655	11	16
9	Modo	700	12	11
10	Babat	719	8	17
11	Pucuk	716	8	6
12	Sukodadi	693	5	13
13	Lamongan	679	6	14
14	Tikung	711	4	23
15	Sarirejo	711	8	6
16	Deket	709	7	15
17	Glagah	750	5	32
18	Karangbinangun	630	5	36
19	Turi	664	6	29
20	Kalitengah	654	6	38
21	Karanggeneng	771	8	27
22	Sekaran	667	6	7
23	Maduran	615	5	11
24	Laren	686	5	21
25	Solokuro	693	4	6
26	Paciran	596	4	9
27	Brondong	685	6	8

Pada tabel 4.1 Data yang digunakan untuk data pengujian adalah data seluruh kecamatan yang berjumlah 27. Adapun sumber data berasal dari Badan Pusat Statistik Kabupaten Lamongan dan BMKG yang diambil pada tahun 2020 - 2022.

4.1.2 Praproses Data

Setelah dataset selesai dikumpulkan, langkah pertama yang penulis lakukan adalah melakukan praproses data. Praproses data ini dilakukan untuk mempersiapkan data yang telah diambil dari berbagai sumber agar siap digunakan dalam proses *data mining*. Praproses data merupakan tahap kritis dalam penelitian ini, yang melibatkan serangkaian langkah untuk membersihkan, mengubah format dan mempersiapkan data agar sesuai dengan kebutuhan penelitian. Dalam

penelitian ini, penulis memanfaatkan *vscode* sebagai lingkungan pengkodean untuk melaksanakan berbagai tugas dan analisis data. Tahap pertama yang penulis lakukan dalam tahap pengkodean adalah meng-*import library* serta data yang akan digunakan. Berikut kodenya :

```
import pandas as pd
import matplotlib.pyplot as plt
import plotly as py
import plotly.graph_objs as go
import warnings

warnings.filterwarnings('ignore')

dfbjr = pd.read_csv('dataskripsi.csv', sep=';', engine='python')
```

Terkait pewarnaan dalam kode ini dan kode yang akan ditampilkan selanjutnya merupakan *default* dari editor kode *vscode*.

Dalam langkah selanjutnya dilakukan pemilihan variabel yang dianggap memiliki relevansi penting terhadap proses *clustering* dan hasil yang diinginkan.

Variabel yang digunakan adalah :

- Curah hujan, memiliki nilai numerik curah hujan dalam satuan mm
- Kemiringan, berisi nilai derajat kemiringan suatu wilayah
- Jumlah kejadian, memiliki informasi tentang banyaknya suatu kejadian

Proses pemilihan ketiga variabel ini dilakukan dengan menggunakan kode berikut :

```
x_train = dfbjr[['Curah hujan', 'Kemiringan', 'Jumlah
Kejadian']].values
```

Sehingga setelah dilakukannya tahap praproses data ini, data yang semula memiliki 4 variabel berkurang menjadi 3 variabel yang valid digunakan.

Selanjutnya adalah tahap normalisasi data, pada tahap ini digunakan metode *MinMaxScaler* yang merupakan salah satu library dari modul *sklearn.preprocessing*. metode ini digunakan untuk menskalakan dan menerjemahkan setiap fitur sehingga berada dalam kisaran tertentu pada set pelatihan, seperti antara 0 dan 1. Untuk mengaplikasikan *MinMaxScaler*, berikut langkah-langkah yang akan dijalankan :

1. *Import library* *MinMaxScaler* dari modul *sklearn.preprocessing*
2. Inisialisasi objek *MinMaxScaler*.
3. Terapkan metode *fit_transform()* pada atribut yang ingin diubah nilainya.
4. Simpan hasil transformasi ke dalam variabel baru.

Berikut adalah contoh pengaplikasian *MinMaxScaler* pada dataset :

```
from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler()
x_train = scaler.fit_transform(x_train)
```

4.1.3 *K-Means Clustering*

Pada implementasi algoritma ini, langkah pertama yang penulis lakukan adalah meng-*import library* *KMeans* dari modul *sklearn.cluster*, kemudian membuat inisialisasi sebagai berikut :

- Atribut *kmean* untuk menyimpan jumlah *cluster* yang diinginkan.

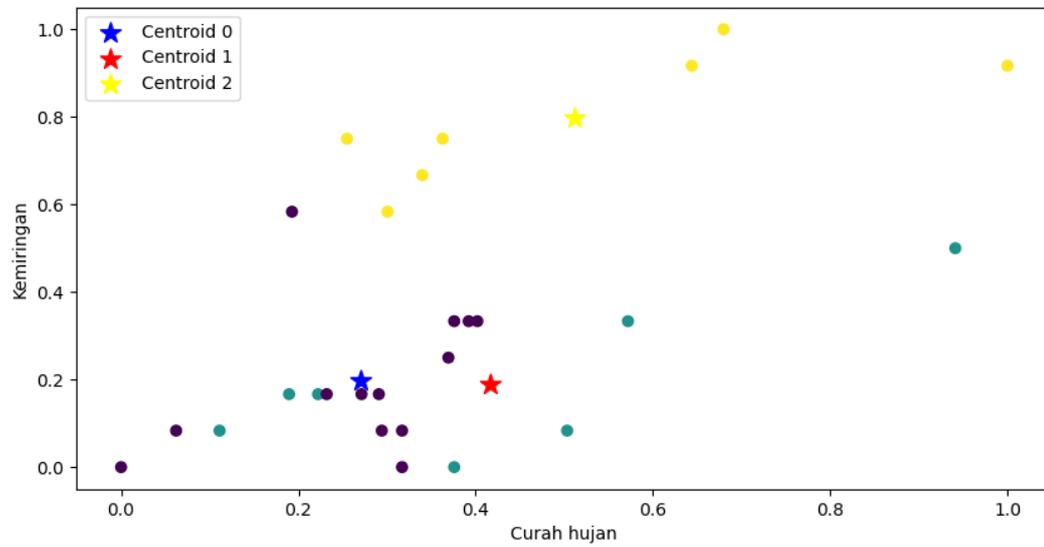
- Metode *fit_predict* digunakan untuk melatih model K-Means pada data input. Inisialisasi *centroid* dilakukan secara acak, kemudian dilakukan iterasi untuk mengupdate *centroid* dan label *cluster* hingga konvergen dan memprediksi label *cluster* untuk data input

Berikut adalah contoh pengaplikasian *library* Kmeans pada data :

```
from sklearn.cluster import KMeans
kmean = KMeans(n_clusters=3)
kmean
y_cluster = kmean.fit_predict(x_train)
dfbjr['Cluster'] = y_cluster
```

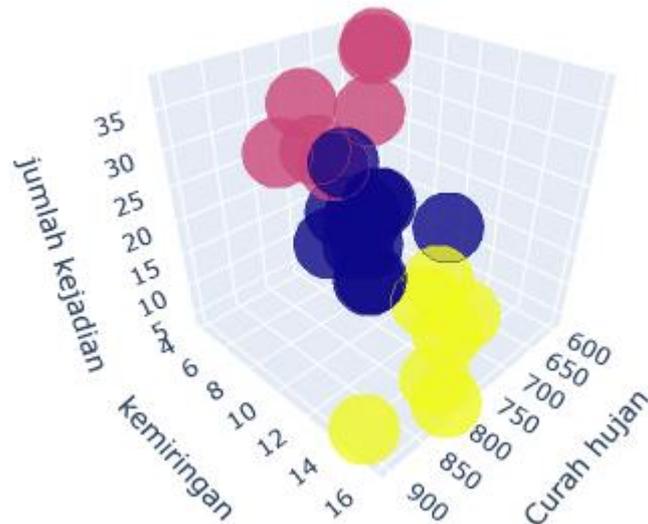
Proses klasterisasi dimulai dengan menginisialisasi jumlah *cluster* yang diinginkan. Kemudian, data yang akan diklasterisasi dimasukkan ke dalam model K-Means dan dilakukan iterasi untuk memperbarui posisi *centroid* dan label kluster hingga mencapai konvergensi. Setelah pelatihan selesai model K-Means dapat digunakan untuk memprediksi label *cluster* dari data baru.

Terakhir, dilakukan visualisasi hasil klasterisasi dengan menggunakan *plot*. Setiap *cluster* ditandai dengan warna yang berbeda, dan titik-titik data yang termasuk dalam *cluster* yang sama digambarkan pada plot. Legenda ditambahkan untuk membedakan *cluster* yang berbeda. Berikut hasil visualisasi persebaran data:



Gambar 4.1 Visualisasi Klaster Hasil K-Means

Gambar 4.1 merupakan hasil visualisasi tiga kelompok *cluster* yang diberi label *centroid* dengan keterangan *centroid 0* adalah *cluster 1*, *centroid 1* adalah *cluster 2*, dan *centroid 2* adalah *cluster 3*.



Gambar 4.2 Visualisasi 3D Kluster Hasil K-Means

Gambar 4.2 merupakan hasil visualisasi tiga kelompok *cluster* dalam bentuk 3 dimensi menggunakan *library go* dan *py* dari modul *plotly*. Proses *clustering* juga dilakukan untuk nilai $k = 2-5$ dengan kode yang sama untuk membandingkan hasilnya dengan visualisasi *cluster* $k = 3$, hasil visualisasinya akan ditampilkan pada sub-bab hasil analisis.

4.1.4 *Elbow Method*

Setelah tahap *clustering* dilakukan untuk mengelompokkan dataset menjadi beberapa *cluster*. Dilakukan visualisasi *Elbow Method* untuk mengetahui nilai k yang optimal digunakan dalam *clustering*, berikut adalah kode penggunaan *Elbow Method* :

```
inertias = []
k_range = range(1,10)
```

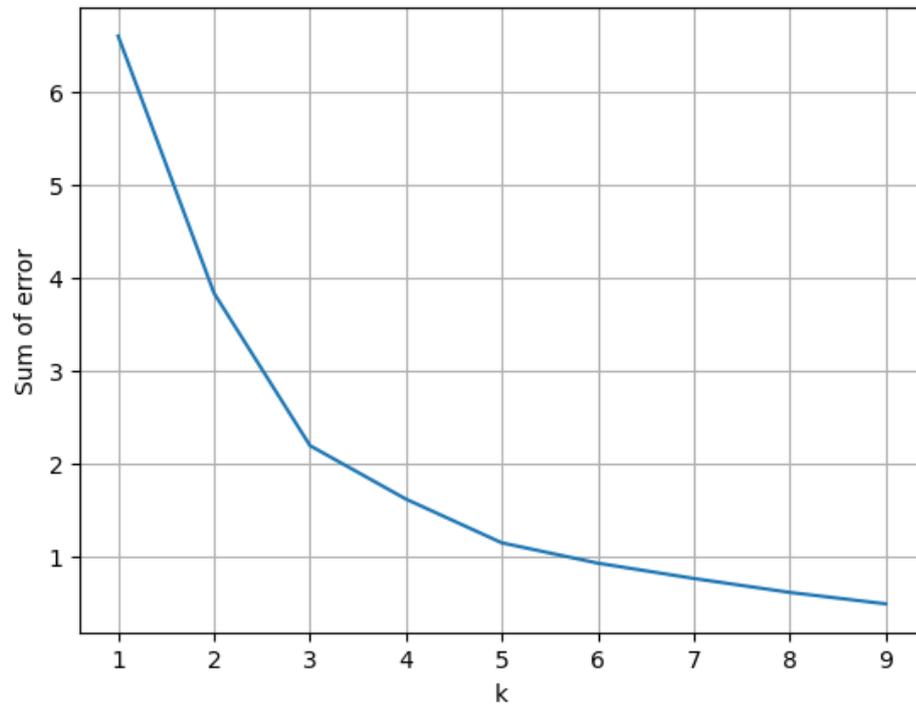
```
for k in k_range:
    km = KMeans(n_clusters=k).fit(x_train)
    inertias.append(km.inertia_)

plt.xlabel('k')
plt.ylabel('Sum of error')
plt.plot(k_range,inertias)
plt.grid()
```

Dari kode diatas menghasilkan barplot dengan nilai *elbow* skor dari setiap *cluster* dengan rentang 1-9. Output dari kode tersebut menyatakan bahwa nilai k yang paling baik adalah nilai k dengan grafik yang paling membentuk siku (*elbow*), yaitu 3 dengan nilai koefisien sebesar 2,1. Selanjutnya adalah mengulangi langkah seperti sebelumnya dengan kode yang sama dan hanya merubah nilai *n_cluster* menjadi nilai yang optimal yaitu k=3.

4.1.5 Hasil Pengujian

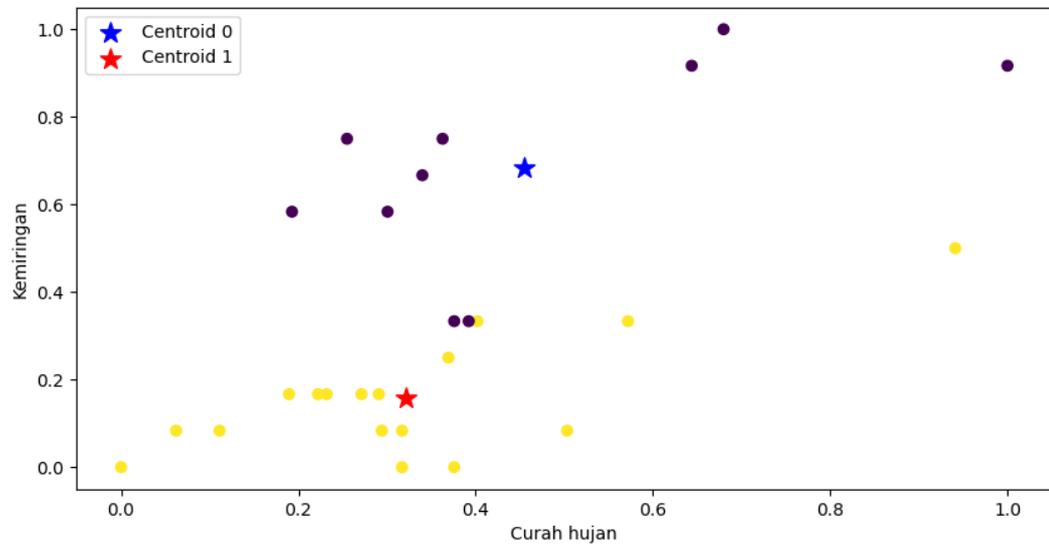
Pengujian yang dilakukan dalam penelitian ini dilakukan secara intrinsik karena data yang digunakan sebelumnya belum memiliki label. Pengujian dilakukan dengan menghitung nilai uji coba k dari rentang 1-9 menggunakan *elbow method*, berikut merupakan bar plot visualisasi perbandingan *elbow method* dari setiap *cluster* :



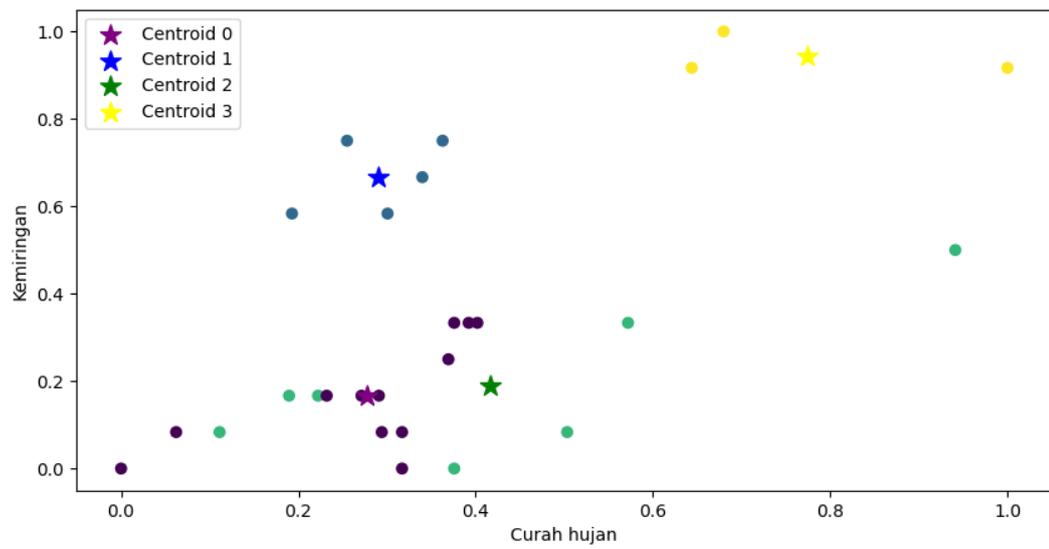
Gambar 4.3 Hasil Pengujian Elbow Method

Berdasarkan gambar 4.3 dapat dilihat nilai $k = 3$ memiliki nilai grafik yang paling membentuk siku (*elbow*) dibandingkan dengan nilai k yang lain yaitu sebesar 2,1. Sehingga untuk pengujian nilai k menghasilkan $k = 3$ merupakan *cluster* yang optimal.

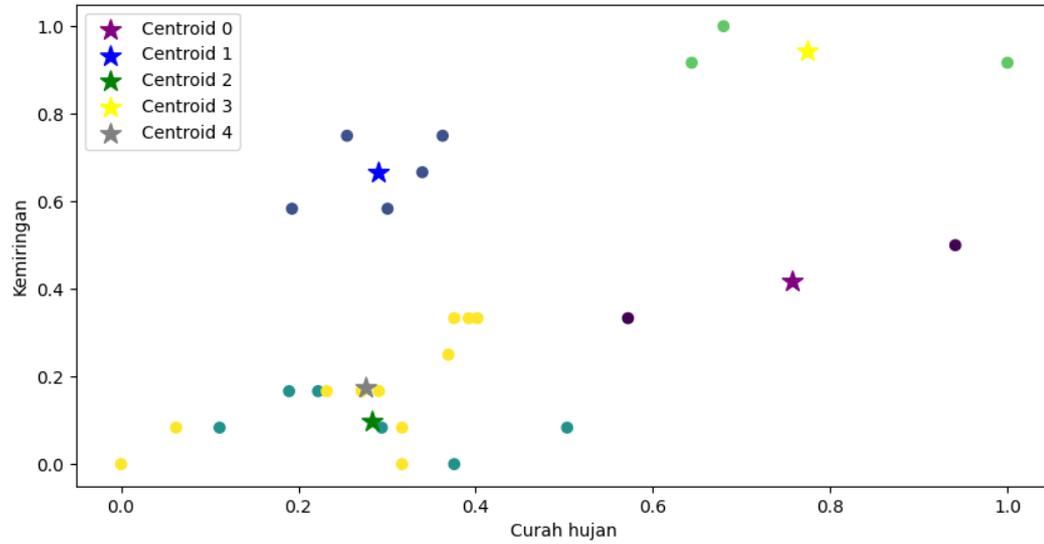
Hasil yang didapatkan dari proses *clustering* dengan Algoritma K-Means adalah sebagai berikut :



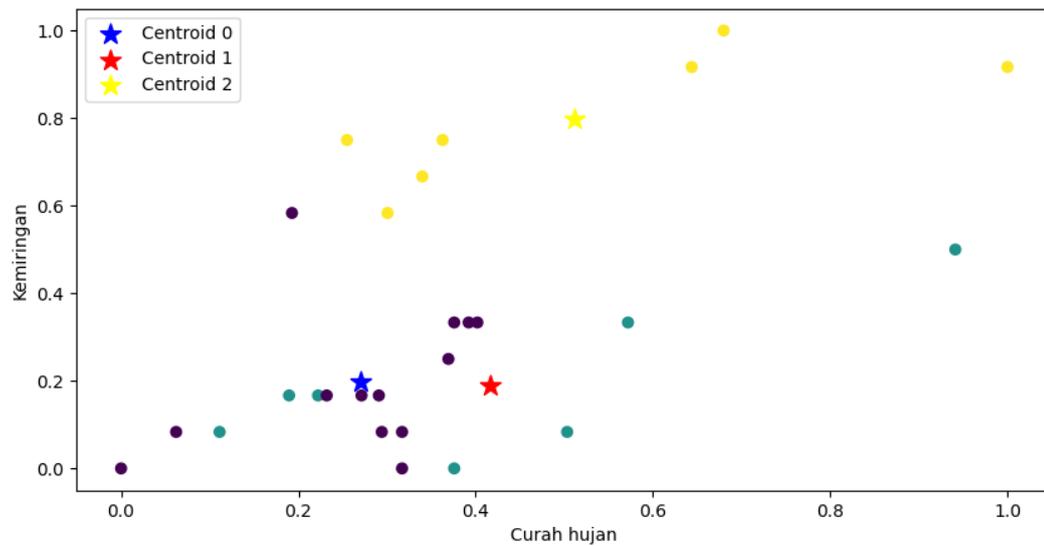
Gambar 4.4 Visualisasi K-Means Dengan Nilai K=2



Gambar 4.5 Visualisasi K-Means Dengan Nilai K=4



Gambar 4.6 Visualisasi K-Means Dengan Nilai k=5



Gambar 4.7 Visualisasi K-Means Dengan Nilai K=3

Dari gambar 4.4 sampai 4.7 dapat dilihat perbandingan visualisasi data dengan nilai k yang berbeda. Dari keempat gambar tersebut dapat dilihat bahwa persebaran dengan nilai k=3 memiliki data yang paling terpisah dengan baik.

Sehingga didapatkan hasil *clustering* wilayah rawan banjir menggunakan K-means adalah sebagai berikut:

Tabel 4.2 Hasil Klustering

No	Kecamatan	Curah Hujan	Kemiringan Tanah	Jumlah Kejadian	Cluster
1	Sukorame	884	10	35	2
2	Bluluk	902	15	7	3
3	Ngimbang	804	16	5	3
4	Sambeng	793	15	6	3
5	Mantup	707	13	4	3
6	Kembangbahu	674	13	5	3
7	Sugio	688	11	5	3
8	Kedungpring	655	11	16	1
9	Modo	700	12	11	3
10	Babat	719	8	17	1
11	Pucuk	716	8	6	1
12	Sukodadi	693	5	13	1
13	Lamongan	679	6	14	1
14	Tikung	711	4	23	2
15	Sarirejo	711	8	6	1
16	Deket	709	7	15	1
17	Glagah	750	5	32	2
18	Karangbinangun	630	5	36	2
19	Turi	664	6	29	2
20	Kalitengah	654	6	38	2
21	Karanggeneng	771	8	27	2
22	Sekaran	667	6	7	1
23	Maduran	615	5	11	1
24	Laren	686	5	21	1
25	Solokuro	693	4	6	1
26	Paciran	596	4	9	1
27	Brondong	685	6	8	1

Pada tabel 4.2 adalah data hasil K-Means klustering menggunakan sistem. Dari tabel diatas dapat dilihat bahwa terdapat wilayah dengan *cluster* 1 sebanyak 13 kecamatan, *cluster* 2 sebanyak 7 kecamatan, dan *cluster* 3 sebanyak 7 kecamatan. Dengan catatan, *cluster* 1 adalah wilayah sedang, *cluster* 2 adalah wilayah sering dan *cluster* 3 adalah wilayah sedikit.

4.1.6 Pengujian Kinerja Sistem

Validasi hasil *clustering* wilayah rawan banjir dilakukan dengan menggunakan metode *silhouette coefficient*. Pengujian dilakukan terhadap 27 data yang didapatkan dari hasil klasterisasi. Tahap pertama yaitu mencari nilai $a(i)$ menggunakan persamaan 2.2, nilai $a(i)$ diperoleh dari perhitungan jarak data ke data yang lain dalam satu *cluster*. Tahap selanjutnya adalah mencari nilai $b(i)$ dengan menggunakan persamaan 2.3 dan 2.4. Berikut hasil perhitungan nilai *silhouette* pada hasil *clustering* wilayah rawan banjir di Kabupaten Lamongan :

Tabel 4.3 Hasil Nilai $a(i)$, $b(i)$, dan nilai *Silhouette* Masing-masing Data

No	Kecamatan	$a(i)$	$b(i)$	Nilai <i>Silhouette</i>
1	Sukorame	82,0627	184,1728	0,5544
2	Bluluk	93,3482	201,0196	0,5356
3	Ngimbang	58,8066	103,7614	0,4333
4	Sambeng	61,7360	92,7682	0,3345
5	Mantup	10,2814	70,7949	0,8548
6	Kembangbahu	13,8633	41,5948	0,6667
7	Sugio	10,6010	52,8998	0,7996
8	Kedungpring	20,9626	47,5011	0,5587
9	Modo	8,8613	62,5299	0,8583
10	Babat	10,4745	80,0473	0,8691
11	Pucuk	10,2182	78,5766	0,8700
12	Sukodadi	8,9729	55,2774	0,8377
13	Lamongan	11,2545	42,1903	0,7332
14	Tikung	10,7853	72,1103	0,8504
15	Sarirejo	9,4698	73,7856	0,8717
16	Deket	8,7395	70,3942	0,8758
17	Glagah	21,3710	84,5780	0,7473
18	Karangbinangun	21,5411	76,1097	0,7170
19	Turi	22,6468	42,8607	0,4716
20	Kalitengah	22,8248	55,5391	0,5890
21	Karanggeneng	79,4180	71,5786	-0,0987

22	Sekaran	25,2606	35,8021	0,2944
23	Maduran	23,6176	86,8212	0,7280
24	Laren	11,0465	47,9639	0,7697
25	Solokuro	9,7945	56,9374	0,8280
26	Paciran	31,2040	105,8140	0,7051
27	Brondong	10,3358	49,0198	0,7892

Pada tabel 4.3 menunjukkan hasil perhitungan nilai *silhouette coefficient* masing-masing data. Dari 27 data yang di uji, dapat dijumlahkan bahwa nilai *silhouette* seluruh data adalah 18,0444. Berikut adalah perhitungan rata-rata dari nilai *silhouette* menggunakan persamaan 2.6 :

$$S_{\text{rata-rata}} = 1 / \text{jumlah data} \times \text{jumlah nilai } \textit{silhouette} \text{ seluruh data}$$

$$= 1/27 \times 18,0444 = 0,6683$$

Dari hasil perhitungan diatas, diketahui bahwa hasil validasi *clustering* menggunakan *silhouette coefficient* memperoleh hasil sebesar 0,6683.

4.2 Pembahasan

Berdasarkan hasil uji coba yang telah dilakukan dengan menggunakan metode K-Means dalam implementasi wilayah rawan banjir di Kabupaten Lamongan. Maka dapat dijelaskan bahwa metode K-Means dapat digunakan untuk menentukan wilayah rawan banjir di Kabupaten Lamongan. Adapun hasil akhir dari perhitungan menunjukkan bahwa di Kabupaten Lamongan terdapat total 7 kecamatan dengan cluster 2, 13 kecamatan dengan cluster 1 dan 7 kecamatan dengan cluster 3.

Pada tampilan gambar 4.3 menunjukkan bahwa nilai k paling baik menurut perhitungan *Elbow method* adalah $k=3$. Pada gambar nomor 4.4 hingga 4.6 adalah tampilan visualisasi hasil K-Means dengan nilai k yang berbeda. Sedangkan pada gambar 4.7 adalah visualisasi hasil K-Means dengan nilai k terbaik yaitu $k=3$.

Pengujian kinerja sistem ini dihitung menggunakan metode *silhouette coefficient*. Hasil dari perhitungan kinerja sistem menggunakan *silhouette coefficient* memperoleh hasil sebesar 0,6683. Dari hasil tersebut dapat diartikan bahwa struktur yang dihasilkan adalah sedang dan dapat digunakan untuk *clustering* wilayah rawan banjir.

Dikarenakan adanya masalah terkait wilayah yang sering terjadi bencana banjir, oleh karena itu dilakukannya penelitian ini untuk memberi informasi kepada masyarakat Kabupaten Lamongan tentang wilayah mana saja yang rawan terjadi bencana banjir. Adapun hasil dari penelitian ini adalah berupa tiga jenis wilayah yaitu wilayah sering banjir, wilayah sedikit banjir dan wilayah banjir sedang. Hasil tersebut akan sangat membantu masyarakat dalam mengetahui wilayah mana saja yang menjadi wilayah yang sering banjir dan tidak sehingga membuat masyarakat menjadi lebih waspada akan bencana banjir.

Adapun integrasi islam, bahwa penjelasan pada semua umat islam tentang berbuat baik dan memberi manfaat kepada sesama, kejadian alam atau geografis diantaranya adalah gambaran tentang lapisan bumi dan warna-warni alam. Pada peristiwa alam sendiri Allah Subhanahu Wa Ta'ala telah menunjukkan beberapa

peristiwa yaitu : turunnya hujan, pergantian siang malam dan pergantian cuaca pada setiap musim.

Seperti yang telah dijelaskan dalam surat QS. Ali ‘Imran 3: ayat 110 tentang berbuat baik atau memberi manfaat pada setiap orang, sebagai berikut :

الْمُنْكَرِ عَنِ وَتَنْهَوْنَ بِالْمَعْرُوفِ تَأْمُرُونَ لِلنَّاسِ أُخْرِجَتْ أُمَّةٌ خَيْرَ كُنْتُمْ

الْمُؤْمِنُونَ مِنْهُمْ ۖ لَهُمْ خَيْرًا لَكَانَ الْكِتَابِ أَهْلُ آمَنَ وَلَوْ ۖ بِاللَّهِ وَتُؤْمِنُونَ

الْفَاسِقُونَ وَأَكْثَرُهُمْ

”Kamu adalah umat yang terbaik yang dilahirkan untuk manusia , menyuruh kepada yang ma`ruf dan mencegah dari yang munkar, dan beriman kepada Allah. Sekiranya ahli kitab beriman, tentulah itu lebih baik bagi mereka, diantara mereka ada yang beriman, dan kebanyakan mereka adalah orang-orang yang fasik.” (QS. Ali ‘Imran 3: Ayat 110).

Ayat ini menjelaskan bahwa umat islam adalah sebaik-baik umat karena mereka senantiasa memerintahkan kepada yang ma`ruf, mencegah perbuatan mungkar dan beriman kepada Allah Subhanahu Wa Ta’ala. Karena itu ayat ini mengandung dorongan kepada umat islam agar terus memelihara sifat-sifat utama tersebut dan agar mereka tetap mempunyai semangat yang tinggi untuk melakukan amar makruf dan nahi mungkar berlandaskan iman yang kokoh kepada Allah Subhanahu Wa Ta’ala. (Hayati, 2017).

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan penelitian yang dilakukan, penulis dapat menyimpulkan bahwa penelitian ini berhasil mengimplementasikan metode K-Means untuk menentukan wilayah rawan banjir di Kabupaten Lamongan menggunakan tiga parameter yaitu curah hujan, kemiringan tanah dan jumlah kejadian banjir yang menghasilkan wilayah sedikit banjir, sedang banjir, dan sering banjir pada setiap kecamatan di Kabupaten Lamongan. Dengan hasil pengujian kinerja sistem menggunakan *silhouette coefficient* memperoleh nilai sebesar 0,66831. Elbow Method sebagai penentu banyaknya klaster yang akan dibentuk, menghasilkan visualisasi hasil persebaran data klaster yang terpisah dengan baik.

5.2 Saran

Berdasarkan hasil penelitian ini, penulis merekomendasikan untuk penelitian selanjutnya agar penelitian ini dikembangkan lagi menggunakan metode prediksi yang berbeda, sehingga didapatkan perbandingan hasil dari setiap pengujian berbagai metode yang lain. Selain itu diharapkan implementasi K-Means juga dapat dibangun berbasis web, agar lebih memudahkan akses pengguna dalam mencari informasi.

DAFTAR PUSTAKA

- Badan Meteorologi Klimatologi Geofisika (2022). Climate Outlook 2023 BMKG: Waspada Terjangan Bencana Hidrometeorologi Basah dan Kering. Buletin BMKG.
- Shihab, M.Q. (2002). Tafsir Al-Misbah jilid 02. Jakarta : Lentera Hati
- Badan Nasional Penanggulangan Bencana (2023). Info Bencana BNPB februari 2023
- S. Yudi, Putri P. Endina, W. Andang, S. Etis. 2020. Pemetaan Zonasi Rawan Banjir Dengan Analisis Indeks Rawan Banjir Menggunakan Metode Fuzzy *Simple Adaptive Weighting*. *Jurnal Pseudocode*, 4(1): 78-87.
- UNISDR. (2009). Terminologi pengurangan risiko bencana.
- Istiqomah, N.A. (2020). “DI KOTA MALANG MENGGUNAKAN METODE FUZZY C-MEANS SKRIPSI Oleh : NANDA AYU ISTIQOMAH.” 2020.
- Hasan, M. Fuad. 2015. “Analisis Tingkat Kerawanan Banjir Di Kabupaten Lamongan.” *Swara Bhumi* 3(3): 239–47.
<https://jurnalmahasiswa.unesa.ac.id/index.php/swara-bhumi/article/view/13123>.
- Nurdiawan, Odi, Harumi Putri, Program Studi, and Teknologi Informasi. 2018. “Pemetaan Daerah Rawan Banjir Berbasis Sistem Informasi Geografis Dalam Upaya Mengoptimalkan Langkah Antisipasi Bencana.” *Jurnal Infotech* 4(2): 1–9.
- Pratiwi Eka Henny. 2020. “Analisis Tingkat Kerawanan Banjir Di Kabupaten Lamongan.” *Swara Bhumi* 3(3): 1–9.
- Risanty, Juwita, Deasy Arisanty, and Eva Alviawati. 2015. “Kerentanan Banjir Di Kecamatan Martapura Barat Kabupaten Banjar.” *Jurnal Pendidikan Geografi* 2(5): 24–43.
- Rosyidie, Arief. 2013. “Banjir: Fakta Dan Dampaknya, Serta Pengaruh Dari Perubahan Guna Lahan.” *Journal of Regional and City Planning* 24(3): 241.
- Soliha, N. 2020. “Analisis Indeks Ancaman Dan Kerentanan Wilayah Pesisir Terhadap Banjir Di Kabupaten Lamongan.”

<http://digilib.uinsby.ac.id/id/eprint/43219>.

- Widyastuti, Mita. 2014. "Kajian Dan Ruang Lingkup Manajemen Bencana." *Jurnal Madani* 2.
- Yunida, Reni, Rosalina Kumalawati, and Deasy Arisanty. 2017. "Dampak Bencana Banjir Terhadap Kondisi Sosial Ekonomi Masyarakat Di Kecamatan Batu Benawa Kabupaten Hulu Sungai Tengah, Kalimantan Selatan." *Jurnal Pendidikan Geografi* 4(4): 42–52.
<http://ppjp.unlam.ac.id/journal/index.php/jpg>.
- Arfah, H. (2023). 6230 Rumah di Lamongan Masih Terendam Banjir Akibat Luapan Bengawan Njero dan Bengawan Solo. Lamongan : www.kompas.com
- Munasyaroh. (2015). Profil Kabupaten Lamongan Jawa Timur. *Munasya.com*.
<https://munasya.com/kabupaten-lamongan-jawa-timur/>
- Ibnukatsironline. (2015). "Tafsir Surat An-Nisa ayat 77-79". diakses dari <http://www.ibnukatsironline.com/2015/05/tafsir-surat-nisa-ayat-77-79.html>
- Lineker Cary, Kusumastuti N, Irawan Beni. (2013). "Clustering Lulusan Mahasiswa Matematika Fmipa Untan Pontianak Menggunakan Algoritma Fuzzy C-Means." *Buletin Ilmiah Mat. Stat. Dan Terapannya*. 2(1): 21-26.
- Matt McGrath. (2021). Bencana banjir: 'Lebih dari 179 juta jiwa diperkirakan terdampak banjir pada 2030'. BBC News Indonesia. Diakses dari <https://www.bbc.com/indonesia/majalah-58109947>
- Herlinda Vera, Darwis Dedy, Dartono. (2021). "Analisis Clustering Untuk Recredesialing Fasilitas Kesehatan Menggunakan Metode Fuzzy C-Means." *Jurnal Teknologi dan Sistem Informasi (JTSI)* Vol. 2, No. 2, Juni 2021, 94 - 99
- Buslim, N., Pradono Iswara. R. & Agustian, F. (2020). THE MODELING OF "MUSHTAHIQ" DATA USING K-MEANS CLUSTERING ALGORITHM AND BIG DATA ANALYSIS (CASE STUDY:LAZ). *Jurnal Teknik Informatika*, 13(2).
- Cahaya Dewi Dewa Ayu I., Pramita Dewa Ayu K. (2019). Analisis Perbandingan Metode Elbow dan Silhouette Pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali. *Jurnal Matrix*, vol 9, No 3.
- Rohmah, A., Sembiring, F., Erfina, A., Program,), Informasi, S. S., Nusaputra, U.,

- Raya, J., Cisaat, C., 21, N., Kaler, C., Cisaat, K., Sukabumi, K. & Barat, J. (2021). *IMPLEMENTASI ALGORITMA K-MEANS CLUSTERING ANALYSIS UNTUK MENENTUKAN HAMBATAN PEMBELAJARAN DARING (STUDI KASUS: SMK YASPIM GEGERBITUNG)*.
<https://www.alfasoleh.com/2019/11/k-mean-clustering-contoh>
- Patrick Trusto Jati Wibowo. (2021). Apa Itu Kaggle?. *Warta Ekonomi*.
<https://wartaekonomi.co.id/read379561/apa-itu-kaggle>
- Percovic, L. (2015). "Introduction to Computing Using Python: An Application Development Focus." Wiley.
<https://ubcc.smartway.co/shop/product/introduction-to-computing-using-python-an-application-development-focus-e-book-85970>
- Nurhayati, Busman & Pradono Iswara, R. (2019). "Pengembangan Algoritma Unsupervised Learning Technique Pada Big Data Analysis di Media Sosial Sebagai Media Promosi Online Bagi Masyarakat". *Jurnal Teknik Informatika*, 12(1).
- Nur Ariyanto Ramdhan & Devi Adi Nufriana. (2019). "Rancang Bangun Dan Implementasi Sistem Informasi Skripsi Online Berbasis Web". *Jurnal Ilmiah Intech : Information Technology Journal of UMUS*, 1. <https://media.neliti.com/media/publications/293457-rancang-bangun-dan-implementasi-sistem-i-7b8cb07b.pdf>
- Witten, Frank & Eibe. (2005). *Data Mining: Practical Machine Learning Tools and Techniques, Second Edition*.
- Wong, K. J. (2022). *7 Evaluation Metrics for Clustering Algorithms | by Kay Jan Wong | Towards Data Science*. Towards Data Science.
<https://towardsdatascience.com/7-evaluation-metrics-for-clustering-algorithms-bdc537ff54d2#2071>
- Andreas Chandra. (2018). Beberapa Cara untuk Preprocessing Data dalam Machine Learning. *Warung Pintar, Medium*. <https://medium.com/warung-pintar/beberapa-cara-untuk-preprocessing-data-dalam-machine-learning-13cef4294614>
- Yufeng G. (2017). *The 7 Steps of Machine Learning*. Towards Data Science.

<https://towardsdatascience.com/the-7-steps-of-machine-learning-2877d7e5548e>

Khairani, D. (2015). *STUDI KASUS PENGUKURAN SISTEM INFORMASI MENGGUNAKAN FUNCTION POINT (FP)*. 8(2). <http://spmb.uinjkt.ac.id>

Khairani, D., Siregar, M. A. R., Masruroh, S. U. & Azizah, M. N. (2022). Selection of Home Wifi Internet: Machine Learning Implementation With Decision Tree C4.5 Algorithm Method. *JURNAL TEKNIK INFORMATIKA*, 15(2), 178 – 187. <http://doi.org/10.15408/jti.v15i2.27741>

Paembonan, S., Abduh, H. (2021) Penerapan Silhouette Coefficient Untuk Evaluasi Clustering Obat. *Jurnal ilmiah ilmu-ilmu teknik*. Vol. 6, No 2. <https://ojs.unanda.ac.id/index.php/jiit/index>

A. T. Rahman, Wiranto, A. Rini. (2017). Coal Trade Data Clustering Using K-Means(Case Study Pt. Global Bangkit Utama). *Jurnal Teknologi dan Informasi*. Vol. 6, No. 1, pp. 24-31, 2017, doi: <https://doi.org/10.20961/itsmart.v6i1.11296>.

Sebatubun, M., Maria, Prayitna A. (2023). Implementasi Metode Elbow dan K-Means Clustering Untuk Mengetahui Kapasitas Solar Panel. *Jurnal Informasi Interaktif* vol. 8 No. 2, 45 – 50.

Atif, S. (2023). *Implementasi Algoritma K-Means Dengan Menggunakan Metode Elbow dan Silhouette Pada Klasterisasi Tingkat Gizi Balita Di Puskesmas Margoyoso Tahun 2020*. 2023.