

***CLUSTERING POLA HUJAN MENGGUNAKAN PRINCIPAL COMPONENT ANALYSIS  
BERBASIS DATA KLIMATOLOGI***

**SKRIPSI**

**Oleh :**  
**FINA MASLAHATUL FIRHAH**  
**NIM. 220605110009**



**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2025**

***CLUSTERING POLA HUJAN MENGGUNAKAN PRINCIPAL COMPONENT  
ANALYSIS BERBASIS DATA KLIMATOLOGI***

**SKRIPSI**

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

Oleh :  
**FINA MASLAHATUL FIRHAH**  
**NIM. 220605110009**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2025**

## HALAMAN PERSETUJUAN

### HALAMAN PERSETUJUAN

#### *CLUSTERING POLA HUJAN MENGGUNAKAN PRINCIPAL COMPONENT ANALYSIS BERBASIS DATA KLIMATOLOGI*

#### SKRIPSI

Oleh:  
**FINA MASLIHATUL FIRHAH**  
NIM. 220605110009

Telah Diperiksa dan Disetujui untuk Diuji:  
Tanggal: 16 Desember 2025

Pembimbing I,

  
**Dr. Agung Teguh Wibowo Almais, M.T.**  
NIP. 19860301 202321 1 016

Pembimbing II,

  
**Roro Inda Melani, M.T., M.Sc.**  
NIP. 19780925 200501 2 008

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



iii

iii

## HALAMAN PENGESAHAN

### HALAMAN PENGESAHAN

*CLUSTERING POLA HUJAN MENGGUNAKAN PRINCIPAL  
COMPONENT ANALYSIS BERBASIS DATA KLIMATOLOGI*

### SKRIPSI

Oleh:  
**FINA MASLAHATUL FIRHAH**  
NIM. 220605110009

Telah Dipertahankan di Depan Dewan Pengaji Skripsi  
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan  
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)  
Tanggal: 16 Desember 2025

#### Susunan Dewan Pengaji

Ketua Pengaji	: Prof. Dr. Suhartono, S.Si., M.Kom. NIP. 19680519 200312 1 001
Anggota Pengaji I	: Prof. Dr. Muhammad Faisal, M.T. NIP. 19740510 200501 1 007
Anggota Pengaji II	: Dr. Agung Teguh Wibowo Almais, M.T. NIP. 19860301 202321 1 016
Anggota Pengaji III	: Roro Inda Melani, M.T., M.Sc. NIP. 19780925 200501 2 008

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



## PERNYATAAN KEASLIAN TULISAN

### PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Fina Maslahatul Firhah  
NIM : 220605110009  
Fakultas / Program Studi : Sains dan Teknologi / Teknik Informatika  
Judul Skripsi : *Clustering Pola Hujan Menggunakan Principal Component Analysis Berbasis Data Klimatologi*

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, 24 Desember 2025  
Yang membuat pernyataan,



Y2H4802JX42653448

Fina Maslahatul Firhah  
NIM.220605110009

## **MOTTO**

*“Hidup terus bergerak dalam batas waktu yang tersedia, selagi langkah masih dapat dipercepat, selesaikan lebih awal agar ruang kegagalan dapat diminimalkan”*

## KATA PENGANTAR

*Assalamu'alaikum Warahmatullahi Wabarakatuh.*

Syukur Alhamdulillah senantiasa penulis panjatkan kepada Allah SWT karena atas rahmat dan hidayah-Nya, sehingga penulis dapat tuntas menulis skripsi yang berjudul “Clustering Pola Hujan Menggunakan Principal Component Analysis Berbasis Data Klimatologi” dengan baik dan lancar. Penulisan skripsi ini tidak terlepas dari bimbingan, doa, dan dukungan dari berbagai pihak. Oleh sebab itu, penulis menyampaikan terimakasih kepada :

1. Prof. Dr. Ilfi Nur Diana, M.Si., CAHRM., CRMP., selaku Rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Dr. Agus Mulyono, M.Kes., selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Supriyono, M.Kom., selaku Ketua Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Dr. Agung Teguh Wibowo Almais, M.T., selaku dosen pembimbing I yang dengan penuh kesabaran dan ketulusan hati memberikan bimbingan, memberi semangat, arahan, masukan, serta membantu penulis dalam mengerjakan dan menyelesaikan skripsi ini.
5. Roro Inda Melani, M.T, M.Sc selaku dosen pembimbing II penulis yang senantiasa memberikan bimbingan dan masukan untuk penulis dalam mengerjakan dan menyelesaikan skripsi ini.
6. Prof. Dr. Suhartono S.Si, M.Kom selaku penguji I dan Prof.Dr. Muhammad Faisal. M.T selaku penguji II yang telah berkenan menguji, memberikan

ilmu, kritik, saran, serta memberikan masukan yang membangun sehingga penulis dapat menyelesaikan skripsi ini dengan baik.

7. Ajib Hanani, M.T., selaku dosen wali yang senantiasa membantu serta turut memberikan ilmu dan saran untuk penulis selama menjalani masa studi pada Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
8. Segenap dosen, laboran, dan jajaran staff Program Studi Teknik Informatika yang telah memberikan ilmu, pengetahuan, dan dukungan selama penulis menjalani studi.
9. Untuk kedua orang tua adik tercinta saya Bapak Muhamad Yazid, M.Pd dan Ibu Nor Faizah, M.Ag serta Muhamad jirjis teguh prasetyo untuk segala dukungan dan doa yang diberikan selama penulis menyelesaikan studi ini.
10. Untuk Teman terdekat, Diah Ayu Rahma yang sudah menemani serta memberikan segala bantuan dari awal perkuliahan hingga saat ini serta saling mendukung dan menguatkan penulis untuk terus semangat dan berjuang bersama.
11. Rekan seperjuangan “Kost Bismillah Bandung”, yang beranggotakan Maulany, Faiza dan Via Terima kasih atas segala dukungan dan motivasi yang diberikan serta kenangan dan suka duka yang diberikan selama masa kuliah .
12. Rekan bimbingan dan magang yang beranggotakan Fadhillah,Faiza,Arum,Intan dan Renata yang saling mendukung dan mendorong penulis untuk terus semangat dan berjuang bersama .

13. Untuk Sahabat Terdekat penulis Alfi,Afi dan Sa'di terima kasih atas segala bantuan, dukungan, dan perhatian yang diberikan tanpa ragu selama kuliah.
14. Pihak BMKG Karangkatees yang nemberikan fasilitas untuk penelitian serta membantu, berbagi ide, dan bertukar pendapat selama proses penelitian berlangsung.
15. Teman-teman KKM 189 “ABIPRAYA LAKSANA” yang telah menjadi keluarga, sahabat, sekaligus rekan selama berkegiatan dan mengabdi di desa
16. Keluarga besar “Infinity” Teknik Informatika Angkatan 2022 yang saling membantu, mengingatkan, berbagi ilmu, dan berbagi pengalaman yang tak terlupakan bersama dari awal studi hingga akhir.
17. Diri sendiri yang telah berjuang, bertahan, dan berdiri tegap hingga saat ini menyelesaikan apa yang telah dimulai dengan baik

Malang,24 Desember 2025

Fina Maslahatul Firhah

## DAFTAR ISI

<b>HALAMAN PERSETUJUAN .....</b>	<b>III</b>
<b>HALAMAN PENGESAHAN.....</b>	<b>IV</b>
<b>PERNYATAAN KEASLIAN TULISAN .....</b>	<b>V</b>
<b>MOTTO .....</b>	<b>VI</b>
<b>KATA PENGANTAR.....</b>	<b>VII</b>
<b>DAFTAR ISI .....</b>	<b>X</b>
<b>DAFTAR GAMBAR .....</b>	<b>XII</b>
<b>DAFTAR TABEL .....</b>	<b>XIII</b>
<b>ABSTRAK .....</b>	<b>XIV</b>
<b>ABSTRACT .....</b>	<b>XV</b>
<b>مستخلص البحث.....</b>	<b>XVI</b>
<b>BAB 1 PENDAHULUAN .....</b>	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	6
1.3 Batasan Masalah .....	6
1.4 Tujuan Penelitian .....	7
1.5 Manfaat Penelitian .....	7
<b>BAB II STUDI PUSTAKA .....</b>	<b>9</b>
2.1 Penelitian Terkait .....	9
2.2 <i>Principal Componen Analysis-clustering</i> .....	14
2.2.1 Normalisasi Data .....	16
2.2.2 Menentukan rasio varians dan nilai eigen .....	17
2.2.3 Menentukan komponen utama.....	18
2.2.4 Membuat grafik 2D .....	19
2.3 Data Klimatologi.....	21
2.4 Silhouette Score .....	22
2.5 Validasi Eksternal .....	23
2.6 Konsep Pembentukan Tiga Cluster dalam Analisis Clustering.....	24
<b>BAB III METODE PENELITIAN .....</b>	<b>27</b>
3.1 Data Preparation.....	27
3.2 Alur Penelitian .....	28
3.3 Desain System.....	29
3.3.1 Input.....	30
3.3.2 Pre Processing .....	30
3.3.3 Main Process .....	31
3.3.4. Output .....	35
2.3 Skenario Pengujian .....	36
2.3.1 Silhouette Score.....	36
2.3.2 Validasi eksternal data .....	37
2.4 Pseudocode Clustering.....	38
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>	<b>40</b>
4.1 Preprocessing Data.....	40
4.4.1 Teknik Pengumpulan Data .....	40

4.4.2 Listwise Deletion .....	40
4.2 Normalisasi Data.....	41
4.2.1 StandarScaler .....	42
4.3 Component Anaisys .....	43
4.3.1 Menetukan Nilai eigen dan Varian Rasio.....	44
4.3.2 Menentukan komponen utama .....	46
4.4. Visualisasi Grafik .....	53
4.4.1 Visualisasi 2 dimensi.....	53
4.4.2 Pengelompokan hasil visualisas dengan cluster terdekat.....	55
4.5 Model Jumlah Cluster .....	58
4.5.1 Model 3 Cluster.....	58
4.5.2 Model 4 Cluster.....	60
4.5.3 Model 5 Cluster.....	63
4.5 Evaluasi Model.....	65
4.6 Integrasi Islam.....	67
4.7 Hasil .....	69
<b>BAB V KESIMPULAN .....</b>	<b>76</b>
5.1 Kesimpulan .....	76
5.2 Saran.....	77

**DAFTAR PUSTAKA**  
**KESIMPULAN**

## DAFTAR GAMBAR

Gambar 3.2 Alur Penelitian .....	28
Gambar 4.1 Grafik 2 dimensi hasil PCA.....	54
Gambar 4.2 Persebaran data grafik 2 dimensi.....	56
Gambar 4.3 Visualisasi Cluster PCA.....	59
Gambar 4.5 Visualisasi Grafik 2D 4 Cluster .....	62
Gambar 4.6 Visualisasi 2D 5 Cluster .....	64

## DAFTAR TABEL

Tabel 2.1 Penelitian terdahulu .....	10
Tabel 3.1 Data Klimatologi .....	27
Tabel 3.4 Pseudocode algoritma Clustering .....	38
Tabel 4.1 Hasil Normalisasi .....	43
Tabel 4.2 hasil Eigenvalue dan Variance Ratio .....	45
Tabel 4.3.hasil faktor loading variabel .....	46
Tabel 4.4 Hasil Komponen Utama .....	49
Tabel 4.5. Rentang nilai clustering .....	59
Tabel 4.5 Rentang nilai 4 cluster .....	61
Tabel 4.6 Rentang nilai 5 cluster .....	63
Tabel 4.7 Hasil Uji validasi .....	65

## ABSTRAK

Firhah, Fina Maslahatul. 2025. **Clustering Pola Hujan Menggunakan Principal Component Analysis Berbasis Data Klimatologi.** Skripsi. Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Dr. Agung Teguh Wibowo Almais, M.T (II) Roro Inda Melani M.T, M.Sc

Kata kunci: Klimatologi, Curah Hujan, PCA, Clustering.

Perubahan pola curah hujan merupakan aspek penting dalam kajian klimatologi karena berpengaruh langsung terhadap ketersediaan sumber daya air, aktivitas pertanian, serta potensi terjadinya bencana hidrometeorologi. Data klimatologi memiliki kompleksitas tinggi karena melibatkan berbagai variabel atmosfer, sehingga diperlukan metode analisis yang mampu mereduksi dimensi data tanpa menghilangkan informasi utama. Penelitian ini bertujuan untuk mengidentifikasi dan mengelompokkan pola curah hujan menggunakan pendekatan Principal Component Analysis (PCA) dan clustering. Data yang digunakan berupa data klimatologi harian tahun 2023 yang diperoleh dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) Stasiun Geofisika Kelas III Malang. Tahap pre-processing meliputi penghapusan data hilang menggunakan metode listwise deletion serta normalisasi data dengan StandardScaler. PCA diterapkan untuk mengekstraksi komponen utama, di mana komponen utama pertama (PC1) mampu menjelaskan 33,89% variasi total data dan didominasi oleh variabel penyinaran matahari serta durasi penyinaran. Skor PC1 selanjutnya digunakan sebagai dasar dalam proses clustering. Pengujian clustering dilakukan dengan membentuk 3, 4, dan 5 klaster, kemudian dievaluasi menggunakan Silhouette Score. Hasil evaluasi menunjukkan bahwa model 3 klaster menghasilkan nilai Silhouette Score tertinggi sebesar 0,5193, diikuti oleh model 4 klaster sebesar 0,4964 dan model 5 klaster sebesar 0,4015. Hal ini menunjukkan bahwa pembentukan tiga klaster memberikan kualitas pemisahan terbaik. Secara keseluruhan, hasil clustering merepresentasikan variasi pola curah hujan dengan karakteristik klimatologi yang berbeda. Penelitian ini membuktikan bahwa integrasi PCA dan clustering efektif dalam menyederhanakan data klimatologi yang kompleks serta mengidentifikasi pola curah hujan secara sistematis.

## ABSTRACT

Firhah, Fina Maslahatul. 2025. *Clustering Pola Hujan Menggunakan Principal Component Analysis Berbasis Data Klimatologi*. Undergraduate Thesis. Departement of informatics Engineering, Faculty of Science and Technology, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Advisors: (I) Dr. Agung Teguh Wibowo Almais, M.T., (II) Roro Inda Melani M.T, M.Sc

Changes in rainfall patterns are an important aspect of climatological studies because they directly affect water resource availability, agricultural activities, and the potential occurrence of hydrometeorological disasters. Climatological data are highly complex due to the involvement of various atmospheric variables, thus requiring analytical methods capable of reducing data dimensionality without losing essential information. This study aims to identify and cluster rainfall patterns using a Principal Component Analysis (PCA) and clustering approach. The data used consist of daily climatological data for the year 2023 obtained from the Meteorology, Climatology, and Geophysics Agency (BMKG), Class III Geophysics Station of Malang. The preprocessing stage includes the removal of missing data using the listwise deletion method and data normalization using StandardScaler. PCA is applied to extract the principal components, where the first principal component (PC1) explains 33.89% of the total data variance and is dominated by solar radiation and sunshine duration variables. The PC1 scores are then used as the basis for the clustering process. Clustering experiments are conducted by forming 3, 4, and 5 clusters and evaluated using the Silhouette Score. The evaluation results indicate that the 3-cluster model achieves the highest Silhouette Score of 0.5193, followed by the 4-cluster model at 0.4964 and the 5-cluster model at 0.4015. These results demonstrate that the three-cluster configuration provides the best cluster separation quality. Overall, the clustering results represent variations in rainfall patterns with distinct climatological characteristics. This study demonstrates that the integration of PCA and clustering is effective in simplifying complex climatological data and systematically identifying rainfall pattern segmentation.

**Key words:** *Climatology, Rainfall, PCA, Clustering.*

## مستخلص البحث

فرحة، فينا مصلحة. ٢٠٢٥. عنقدة أنماط هطول الأمطار باستخدام تحليل المكونات الرئيسية المعتمد على البيانات المناخية. رسالة جامعية. قسم تقنية المعلومات، كلية العلوم والتكنولوجيا، جامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانغ. المشرف الأول: د. أخونغ تيغوه وبيوو ألايس، ماجستير. المشرف الثاني رورو إندا ميلاني، ماجستير تقني، ماجستير.

تُعدّ التغيرات في أنماط هطول الأمطار من الجوانب المهمة في الدراسات المناخية لما لها من تأثير مباشر على توافر الموارد المائية، والأنشطة الزراعية، واحتمالية حدوث الكوارث الميدرومولوجية. وتميز البيانات المناخية بدرجة عالية من التعقيد نتيجة لتدخل العديد من المتغيرات الجوية، مما يستلزم استخدام أساليب تحليلية قادرة على تقليل أبعاد البيانات دون فقدان المعلومات الأساسية. وتحدف هذه الدراسة إلى تحديد وجمعية أنماط هطول الأمطار باستخدام منهجية تحليل المكونات الرئيسية (PCA) وتقنيات التجميع (Clustering). تعتمد البيانات المستخدمة في هذه الدراسة على بيانات مناخية يومية لعام ٢٠٢٣ صادرة عن وكالة الأرصاد الجوية وعلم المناخ والجيوفيزاء (BMKG) بمحطة الجيوفيزاء من الفئة الثالثة في مالانغ. وتشمل مرحلة المعالجة المسبقية حذف القيم المفقودة باستخدام أسلوب الخذف الكلي (Listwise Deletion) وتطبيع البيانات باستخدام (StandardScaler). وقد طُبق تحليل المكونات الرئيسية لاستخلاص المكونات الأساسية، حيث تمكن المكون الرئيسي الأول (PC1) من تفسير نسبة ٣٣,٨٩٪ من إجمالي تباين البيانات، وكان متأثراً بشكلٍ رئيسي بمتغير الإشعاع الشمسي ومدة سطوع الشمس. واستُخدمت درجات (PC1) كأساس عملية التجميع. أُجريت عملية التجميع بتكوين ٣ و ٤ و ٥ عناقيد، ثم جرى تقييم النتائج باستخدام معامل السيلويت (Silhouette Score) وأظهرت نتائج التقييم أن غودج الثلاثة عناقيد حق أعلى قيمة لمعامل السيلويت بلغت ٠,٥١٩٣، يليه غودج الأربعة عناقيد بقيمة ٠,٤٩٦٤، ثم غودج الخمسة عناقيد بقيمة ٠,٤٠١٥، مما يدل على أن تقسيم البيانات إلى ثلاثة عناقيد يوفر أفضل جودة في فصل العناقيد. وبشكل عام، تعكس نتائج التجميع تنوع أنماط هطول الأمطار ذات الخصائص المناخية المختلفة. وتؤكد هذه الدراسة أن دمج تحليل المكونات الرئيسية مع تقنيات التجميع يُعدّ فعالاً في تبسيط البيانات المناخية المعقدة وتحديد أنماط هطول الأمطار بصورة منهجية.

الكلمات المفتاحية: علم المناخ، معدل المطر، تحليل المكونات الرئيسية(PCA) ، التجميع(Clustering).

# **BAB 1**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Klimatologi merupakan ilmu yang mempelajari pola cuaca dan iklim dalam jangka panjang dengan memanfaatkan data observasi seperti suhu, curah hujan, kelembapan, dan angin (Arisandi & Ruhiat, 2021). Data klimatologi merupakan kumpulan informasi dari berbagai unsur yang sudah tertata sistematis dalam kurun waktu tertentu (Sarker, 2022). Data klimatologi sangat penting karena mampu memberikan gambaran mengenai kondisi iklim dalam skala harian, bulanan, maupun tahunan (M. Audi Ikhsan & Samsudin, 2024)

Salah satu unsur klimatologi yang paling penting adalah curah hujan, karena variabel ini secara langsung berhubungan dengan siklus hidrologi dan ketersediaan sumber daya air di suatu wilayah (Zapata-Sierra et al., 2022). Curah hujan merupakan jumlah air yang jatuh ke permukaan bumi dalam bentuk tetesan air atau salju yang memiliki satuan milimeter (mm) pada luasan dan periode waktu tertentu (Suhadi et al., 2023). Untuk memprediksi Variabilitas pola hujan sangat sulit karena banyak faktor seperti perubahan iklim jangka panjang seperti perbedaan topografi wilayah (Jihan Bastini et al., 2024). Kompleksitas faktor-faktor tersebut menyebabkan distribusi curah hujan bervariasi secara spasial dan temporal, sehingga menimbulkan tantangan dalam memahami serta mengantisipasi dampaknya(Yuningsih et al., 2024).

Seperti pada penelitian Pratama et al, yang meneliti pola sebaran spasial curah hujan di Provinsi Jawa Tengah dengan memanfaatkan metode Poligon Thiessen

berbasis Sistem Informasi Geografis (SIG). Hasil penelitian menunjukkan bahwa wilayah dengan topografi perbukitan dan dataran tinggi, seperti Banjarnegara dan Wonosobo, cenderung memiliki curah hujan yang sangat tinggi, mencapai lebih dari 3200 mm per tahun (Pratama & Darmawan, 2025). Penelitian tersebut membuktikan bahwa kondisi topografi memiliki peranan penting dalam memengaruhi distribusi curah hujan di suatu wilayah. Pemetaan sebaran curah hujan yang akurat dapat menjadi dasar penting dalam upaya mitigasi bencana, khususnya banjir yang kerap terjadi di daerah rawan.

Pada penelitiannya Khasanah et al, yang melakukan analisis variabilitas curah hujan di Daerah Aliran Sungai (DAS) Kali Magok melalui pendekatan analisis trend Mann-Kendall. Hasil penelitian menunjukkan bahwa curah hujan di wilayah tersebut mengalami perubahan yang cukup besar dari waktu ke waktu, sehingga kondisi hidrologinya sangat dinamis (Khasanah & Gardjito, 2024). Perubahan ini menjadi faktor penting dalam perencanaan dan pengelolaan sumber daya air, baik untuk kebutuhan rumah tangga, pertanian, maupun upaya mengurangi risiko bencana hidrometeorologi (Khasanah & Gardjito, 2024). Penelitian ini memberikan pemahaman yang lebih jelas tentang karakteristik hujan di tingkat lokal untuk pengelolaan sumber daya air serta menegaskan pentingnya penggunaan data klimatologi sebagai dasar dalam menyusun strategi pengelolaan sumber daya air yang berkelanjutan.

Penelitian tentang pengenalan curah hujan juga yang dilakukan oleh Setiyowati et al, memanfaatkan metode *Empirical Orthogonal Function* (EOF) untuk mengkaji pola curah hujan di Pulau Jawa. Pendekatan ini terbukti efektif

dalam mengekstraksi pola curah hujan yang kompleks(Setiyowati & Ariska, 2024). Meski demikian, penelitian tersebut masih terbatas pada identifikasi pola dominan tanpa dilengkapi segmentasi lebih detail, sehingga menyisakan ruang untuk kajian lanjutan dalam upaya memahami distribusi hujan yang lebih spesifik.

Dari Penelitian tersebut kebutuhan informasi iklim sangat penting dengan pergeseran pola hujan di daerah dataran tinggi yang memiliki curah hujan yang tinggi sehingga mudah untuk terkena banjir (Santoso et al., 2022). Kondisi ini juga berpotensi mengakibatkan kerusakan ekosistem yang tinggi. Oleh karena itu, segmentasi curah hujan menjadi sangat penting karena informasi tersebut dapat dimanfaatkan untuk merancang strategi dalam antisipasi kegiatan saat curah hujan ekstrem dengan baik sehingga bisa untuk mengurangi jumlah kerugian yang disebabkan oleh curah hujan ekstrem(Benny Hartanto et al., 2022).

Dengan adanya pengetahuan tersebut, masyarakat diharapkan lebih siap menghadapi potensi perubahan lingkungan, serta mampu menyesuaikan aktivitas sehari-hari dengan kondisi iklim yang semakin dinamis. Seperti yang dijelaskan dalam firman Allah surah An-Nur ayat 43

الَّمْ نَرَ أَنَّ اللَّهَ يُرْجِي سَخَابَنَمَ يُؤْلِفُ بَيْنَهُمْ يَجْعَلُهُ رُكَاماً فَتَرَى الْوَدْقَ يَخْرُجُ مِنْ خَلْلَةٍ وَيُنَزَّلُ مِنَ السَّمَاءِ  
مِنْ جِبَالِنِيَّةِ مِنْ بَرَدٍ فَيُصِيبُ بِهِ مَنْ يَشَاءُ وَيَصْرُفُهُ عَنْ مَنْ يَشَاءُ يَكَادُ سَنَّا بَرْقَهُ يَدْهُبُ بِالْأَبْصَارِ (٤٣)

*Tidakkah engkau melihat bahwa sesungguhnya Allah mengarahkan awan secara perlahan, kemudian mengumpulkannya, lalu menjadikannya bertumpuk-tumpuk. Maka, engkau melihat hujan keluar dari celah-celahnya. Dia (juga) menurunkan (butiran-butiran) es dari langit, (yaitu) dari (gumpalan-gumpalan awan seperti) gunung-gunung. Maka, Dia menimpakannya (butiran-butiran es itu) kepada siapa yang Dia kehendaki dan memalingkannya dari siapa yang Dia kehendaki. Kilauan kilatnya hampir-hampir menghilangkan penglihatan(43)*

Dalam tafsir tahlili ayat menjelaskan bahwa didalam al-quran Allah Swt. mengarahkan perhatian Nabi Muhammad Saw. dan umat manusia agar merenungkan bagaimana awan digerakkan oleh kekuasaan-Nya dari satu tempat ke tempat lain, kemudian dikumpulkan sehingga membentuk gumpalan besar yang menyerupai gunung-gunung di angkasa. Dari gumpalan awan tersebut kemudian turunlah hujan, terkadang disertai butiran es. Fenomena ini lebih jelas terlihat ketika manusia berada di ketinggian, misalnya dari dalam pesawat, di mana awan-awan yang bergerak itu tampak menjulang seperti gunung dan air hujan terlihat jelas jatuh ke bumi. Turunnya hujan yang deras dapat menjadi rahmat ketika ia menyuburkan kembali tanah yang kering, menjadikan sawah ladang berproduksi, serta memberikan hasil panen yang menyejahterakan manusia. Namun, hujan juga bisa menjadi bencana apabila turun secara berlebihan, menyebabkan banjir, merendam lahan pertanian, hingga merusak pemukiman. Hal ini menunjukkan bahwa hujan memiliki dua sisi, sebagai nikmat dan sekaligus ujian, sesuai dengan kehendak Allah Swt (Al-Qur'an, 2025).

Dari tafsir ayat tersebut dijelaskan bahwa hujan bisa membawa manfaat, tetapi juga bisa menimbulkan kerugian. Karena itulah, segmentasi pola hujan menjadi penting dilakukan agar kita bisa memanfaatkan hujan sebagai rahmat sekaligus mencegah dampak buruk yang mungkin ditimbulkannya

Clustering pola hujan membutuhkan metode yang mampu menangani data yang kompleks(Gustoro et al., 2022). Data klimatologi, khususnya curah hujan, biasanya memiliki banyak variabel yang saling berkaitan dan dominan sehingga sulit untuk diinterpretasikan secara langsung. Terdapat teknik untuk mengetahui

parameter yang dominan pada suatu data yaitu *Principal Component Analysis (PCA)* (Parding et al., 2023). *PCA* merupakan teknik statistika multivariat yang berfungsi untuk mereduksi dimensi data dengan tetap mempertahankan variasi utama(Haqq et al., 2021). Melalui proses ini, data curah hujan yang awalnya sangat kompleks dapat diringkas menjadi beberapa komponen utama yang lebih sederhana, sehingga pola dominan dapat terlihat dengan lebih jelas(Hayati et al., 2025). Dalam penelitian Almais et al. (2023), dijelaskan bahwa proses *clustering* terhadap suatu data dapat dilakukan dengan memanfaatkan hasil dari Principal Component Analysis (PCA), sehingga terbentuk pendekatan yang dikenal sebagai PCA–Clustering. Metode ini menggunakan nilai komponen utama hasil PCA untuk membentuk kelompok data berdasarkan kesamaan karakteristiknya. Dalam penelitian ini, pendekatan tersebut dikembangkan lebih lanjut dengan mengintegrasikan PCA dan Clustering agar proses pengelompokan data curah hujan dapat dilakukan secara lebih akurat dan terstruktur. PCA digunakan untuk mengekstraksi serta mereduksi dimensi data agar pola dominan lebih mudah diidentifikasi, kemudian hasil proyeksi PCA dianalisis untuk membentuk *cluster* berdasarkan tingkat kemiripan antar data. Kombinasi metode ini diharapkan dapat membantu dalam segmentasi jenis pola curah hujan, sehingga analisis variasi pola hujan dapat dilakukan secara lebih sistematis dan efisien. Melalui hasil Clustering tersebut, wilayah dengan karakteristik curah hujan yang berbeda dapat dikenali dengan lebih jelas, yang selanjutnya dapat dimanfaatkan dalam analisis iklim, perencanaan pertanaman, dan pengelolaan sumber daya air yang berbasis data klimatologi.

Adapun, kontribusi utama dari penelitian ini adalah menyediakan pendekatan metode analisis data modern melalui penerapan PCA dan Clustering sebagai solusi yang relevan untuk menyederhanakan sekaligus mengelompokkan data curah hujan. Kombinasi kedua metode ini memungkinkan proses analisis terhadap data klimatologi yang kompleks menjadi lebih efisien, karena PCA berfungsi dalam mereduksi dimensi dan mengekstraksi komponen utama data dan Clustering memanfaatkan hasil proyeksi tersebut untuk membentuk kelompok berdasarkan tingkat kemiripan pola curah hujan. Dengan pendekatan ini, penelitian diharapkan mampu memberikan kontribusi nyata dalam memahami variabilitas pola hujan, serta mendukung upaya mitigasi risiko bencana, perencanaan pertanaman, dan pengelolaan sumber daya air yang lebih efektif dan berkelanjutan.

## **1.2 Rumusan Masalah**

1. Bagaimana menerapkan metode Principal Component Analysis (PCA) untuk mereduksi dimensi dan mengekstraksi fitur utama dari data klimatologi yang kompleks sebagai dasar proses clustering pola curah hujan?
2. Bagaimana mengukur hasil Clustering pola curah hujan yang dibentuk berdasarkan komponen utama hasil Principal Component Analysis (PCA)?

## **1.3 Batasan Masalah**

1. Data yang digunakan data klimatologi wilayah BMKG Karangkates dan Karangploso Wilayah Kabupaten Malang

2. Data Pola Curah hujan menggunakan data klimatologi dengan variabel curah hujan, Kelembapan, Penyinaran Matahari, Tekanan, Kecepatan angin, Arah angin, Durasi Penyinaran.

#### **1.4 Tujuan Penelitian**

1. Menerapkan metode Principal Component Analysis (PCA) untuk mereduksi dimensi dan mengekstraksi fitur utama dari data klimatologi yang kompleks sebagai dasar dalam proses clustering jenis pola curah hujan.
2. Mengevaluasi tingkat keakuratan hasil Clustering pola curah hujan yang diperoleh dari penerapan metode PCA.

#### **1.5 Manfaat Penelitian**

1. Memberikan pemahaman ilmiah tentang variasi dan pola curah hujan di Malang untuk penanganan kerugian curah hujan ekstrem.
2. hasil penelitian dapat dikembangkan menjadi klasifikasi curah hujan agar lebih akurat untuk mendukung prediksi cuaca dan peringatan dini.

## **BAB II**

### **STUDI PUSTAKA**

#### **2.1 Penelitian Terkait**

Penelitian-penelitian sebelumnya memiliki peran penting sebagai dasar perbandingan dan acuan dalam studi ini. Kajian terhadap penelitian terdahulu bertujuan untuk memastikan bahwa penelitian yang dilakukan memiliki perbedaan yang jelas sehingga tidak dianggap sama dengan penelitian sebelumnya. Oleh karena itu, pada subbab ini disajikan uraian mengenai hasil-hasil penelitian terdahulu.

Penelitian sebelumnya dilakukan oleh Almais et al., Menggunakan Principal Component Analysis (PCA) untuk menganalisis 216 data kerusakan sektor pascabencana dengan lima kriteria. Data direduksi menjadi dua komponen utama (PC1 dan PC2), di mana PC1 menjelaskan (85.17%) variansi dengan nilai eigen (4.28) sedangkan PC2 hanya menjelaskan (9. 36%) variansi dengan nilai eigen (0.47). Hasil reduksi data divisualisasikan dalam grafik dua dimensi dan menjadi dasar dalam melakukan clustering dengan memanfaatkan nilai distribusi dari grafik 2 dimensi sehingga terbentuk tiga kelompok data yang mewakili kategori kerusakan ringan, sedang, dan berat berdasarkan penetuan kategori dari supervisor (Almais et al., 2023). Pengujian menunjukkan bahwa PC1 memberikan distribusi data yang lebih baik dibandingkan PC2, sehingga hasil PCA dapat dimanfaatkan tidak hanya untuk reduksi dimensi tetapi juga sebagai dasar untuk melakukan metode pelabelan data *unsupervised* sebelum dilakukan analisis lebih lanjut.

Penelitian sebelumnya juga dilakukan oleh Liang et al., Menganalisis akumulasi bongkahan es di depan pintu air yang memengaruhi efisiensi transportasi air dan keamanan saluran pada musim es. Model gabungan Principal Component Analysis dan Support Vector Machine (PCA-SVM) digunakan untuk mengklasifikasikan bongkahan es, dengan dua komponen utama yang menjelaskan (71.76%) dan (15.64%) variansi sebagai input model(Liang et al., 2022). Validasi menggunakan *5-fold cross-validation* menunjukkan bahwa kernel Gaussian RBF memberikan kinerja terbaik. Hasil evaluasi memperlihatkan bahwa model PCA-SVM memiliki akurasi, nilai AUC, dan F1-score lebih tinggi dibandingkan Bernoulli Naive Bayes dan KNN.

Tabel 2.1 Penelitian terdahulu

NO	Refrensi	Topik	Metode	Subjek
1	(Musfiroh et al., 2023)	Prediksi curah hujan harian	Kombinasi Principal Component Analysis (PCA) dan Long Short-Term Memory (LSTM)	Data Klimatologi dengan variabel Tekanan Udara, Evaporasi, Temperatur Udara, Kelembapan, Lama penyinaran Matahari, Curah hujan pada tahun 2017 sampai 2022 dari BMKG Lawu Utara
2	(Hayati et al., 2025)	Prediksi curah hujan untuk mitigasi bencana dan pengelolaan iklim	Tweedie Mixture dan PCA	Data curah hujan dari 12 titik pos hujan di Provinsi Lampung dengan variabel curah hujan,

				suhu udara, tekanan udara, kelembapan, dan kecepatan angin data dikumpulkan dari BMKG Stasiun Klimatologi Lampung pada tahun 2024
3	(Hediyati & Suartana, 2021)	Reduksi dimensi pada clustering data produksi pertanian di Kabupaten Bojonegoro	Principal Component Analysis (PCA) dan Clustering (K-Means)	Data produksi pertanian dengan variabel tahun produksi, luas tanam, Jumlah produktivitas, Produktivitas, dari BPS Kabupaten Bojonegoro tahun 2017–2020
4	(Suryaman et al., 2021)	Klasifikasi kondisi cuaca	VGG-16 dan PCA dan K-Nearest Neighbor (KNN)	Data citra cuaca dengan variabel kondisi cuaca berupa berawan, berkabut, hujan, cerah, dan matahari terbit pada tahun 2021 yang berasal dari dataset Kaggle Universitas Telkom
5	(Sunarmi et al., 2022)	Analisis unsur cuaca yang memengaruhi pertanian bawang merah di Kabupaten Nganjuk	Principal Component Analysis (PCA)	Data unsur cuaca dengan variabel suhu udara, kelembapan, curah hujan, arah angin, kecepatan angin,

				penyinaran matahari Kabupaten Nganjuk tahun 2019:
6	(Almais et al., 2024)	Identifikasi tingkat kerusakan bangunan pasca bencana alam	Grey Level Co-occurrence Matrix (GLCM) + Principal Component Analysis (PCA)	Data penelitian ini adalah citra bangunan pasca-bencana alam dengan variabel kontras, dissimilarity, homogeneity, energssssy, dan correlation. Data diperoleh dari situs open data Kaggle serta hasil pencarian citra bangunan 2023
7	(Almais et al., 2023)	Analisis kerusakan sektor pascabencana dan pengelompokan tingkat kerusakan bangunan	Principal Component Analysis (PCA) + Clustering	Data kerusakan sektor pasca-bencana dengan variabel kondisi bangunan, struktur bangunan, fisik bangunan, fungsi bangunan, dan kondisi pendukung lainnya. Data berasal dari hasil analisis data Badan Nasional Penanggulangan Bencana (BNPB) yaitu data hasil survei kerusakan pasca-bencana yang terjadi di

				Provinsi Jawa Timur pada tahun 2021, diperoleh dari
8	(Dewi & Pakereng, 2023)	Pengelompokan tingkat pendidikan di Kabupaten Semarang	Principal Component Analysis (PCA) + K-Mseans Clustering	Data Survei Sosial Ekonomi Nasional (SUSENAS) dengan variabel pendidikan tertinggi, status dalam keluarga, jenis kelamin, Umur 2021 Maret (KOR) wilayah Kabupaten Semarang
9	Penelitian saat ini	Segmentasi Pola curah hujan	Principal Component Analysis (PCA) + Clustering	Data Klimatologi 2023 dengan menggunakan 7 variabel dan pengujian silhuette score

Penelitian oleh (Aidoo et al., 2021) memanfaatkan Principal Component Analysis (PCA) dan Geographically Weighted PCA (GWPCA) untuk mengkarakterisasi kandungan logam berat dalam tanah di Kumasi, Ghana. PCA digunakan untuk mengidentifikasi komponen utama yang menjelaskan (86%) variasi total logam berat, dengan Fe dan Zn sebagai kontributor dominan. Namun, karena distribusi logam berat bersifat heterogen secara spasial, GWPCA diterapkan untuk menangkap keragaman spasial tersebut. Hasil penelitian ini membuktikan bahwa PCA tidak hanya berguna untuk reduksi dimensi, tetapi juga dapat dikombinasikan dengan metode lain untuk meningkatkan akurasi analisis dan mendukung pengambilan keputusan berbasis data.

Berdasarkan hasil penelitian terdahulu, Principal Component Analysis (PCA) – Clustering terbukti efektif untuk melakukan ekstraksi fitur, reduksi dimensi, visualisasi data, dan pembentukan klaster pada berbagai bidang, termasuk klimatologi. Temuan-temuan tersebut menjadi dasar bahwa PCA – Clustering mampu menyederhanakan data yang kompleks sekaligus mempertahankan informasi utama yang dibutuhkan untuk analisis lebih lanjut. Oleh karena itu, penelitian ini menerapkan PCA untuk menganalisis dan mengelompokkan pola curah hujan.

## ***2.2 Principal Componen Analysis-clustering***

Principal Component Analysis (PCA) merupakan metode statistik multivariat yang banyak digunakan dalam analisis data berdimensi tinggi, khususnya ketika terdapat hubungan atau korelasi yang kuat antarvariabel. PCA bertujuan untuk menyederhanakan struktur data dengan cara mentransformasikan variabel asli ke dalam ruang baru yang memiliki dimensi lebih rendah tanpa menghilangkan informasi penting yang terkandung di dalam data tersebut. Metode ini sangat berguna dalam mengatasi permasalahan *curse of dimensionality*, di mana peningkatan jumlah variabel akan meningkatkan kompleksitas analisis dan interpretasi data(Jolliffe, 1986).

PCA mengubah sekumpulan variabel asli yang saling berkorelasi menjadi sejumlah variabel baru yang disebut komponen utama (principal components). Komponen utama ini bersifat tidak saling berkorelasi (orthogonal) dan disusun berdasarkan urutan kemampuan dalam menjelaskan variabilitas data. Komponen utama pertama menjelaskan variansi terbesar dalam data, sedangkan komponen

utama berikutnya menjelaskan sisa variansi dengan syarat tidak memiliki korelasi dengan komponen sebelumnya ( (Ritonga & Muhandhis, 2021).

PCA bekerja dengan pendekatan matematis yang berfokus pada struktur variansi dan kovariansi data. Variansi digunakan sebagai ukuran seberapa besar penyebaran data dari nilai rata-ratanya, sedangkan kovariansi menggambarkan hubungan linier antarvariabel(Mubarok et al., 2025). Dengan memanfaatkan informasi variansi dan kovariansi ini, PCA mampu mengidentifikasi arah baru dalam ruang data yang mengandung informasi paling dominan. Arah tersebut kemudian digunakan sebagai dasar pembentukan komponen utama

Keunggulan utama PCA terletak pada kemampuannya untuk mengurangi redundansi informasi yang disebabkan oleh korelasi antarvariabel. Dalam banyak kasus, variabel-variabel yang digunakan dalam suatu penelitian tidak sepenuhnya independen, sehingga sebagian informasi yang dikandung bersifat berulang. PCA mengatasi permasalahan ini dengan mengompres informasi yang berulang ke dalam sejumlah kecil komponen utama yang lebih ringkas namun tetap representatif(Arlandy et al., 2025).

PCA juga berperan penting dalam meningkatkan efisiensi komputasi dan stabilitas analisis statistik. Dengan berkurangnya jumlah dimensi data, proses analisis lanjutan seperti klasifikasi, clustering, maupun pemodelan statistik dapat dilakukan dengan lebih cepat dan akurat. PCA juga membantu mengurangi pengaruh noise pada data, karena komponen utama dengan variansi kecil yang cenderung merepresentasikan gangguan atau fluktuasi acak dapat diabaikan(Sartika et al., 2022).

Dalam bidang klimatologi, PCA telah banyak digunakan untuk menganalisis data iklim yang kompleks dan berdimensi tinggi, seperti curah hujan, suhu udara, kelembapan, tekanan udara, dan kecepatan angin. Variabel-variabel tersebut umumnya memiliki hubungan yang erat satu sama lain, baik secara temporal maupun spasial (Akbar et al., 2022). PCA memungkinkan peneliti untuk mengekstraksi pola iklim dominan yang mencerminkan karakteristik utama suatu wilayah, seperti pola musim hujan, variabilitas curah hujan tahunan, atau pengaruh fenomena iklim tertentu.

Berikut adalah penjelasan lebih rinci tentang proses PCA beserta langkah-langkah utama dan rumusnya (Almais et al., 2024):

### 2.2.1 Normalisasi Data

Tahap awal dalam implementasi PCA adalah normalisasi data untuk menyamakan skala antarvariabel sehingga tidak ada variabel yang memiliki nilai lebih tinggi dari variabel lain.

Salah satu metode normalisasi yang umum digunakan adalah *StandardScaler*, yaitu metode yang mentransformasikan data agar setiap variabel memiliki nilai rata-rata nol dan standar deviasi satu. Transformasi ini dirumuskan sebagai berikut:

$$X_{Standart} = \frac{X - \text{mean}(x)}{X_{(\text{standart deviasi})}} \quad (2.1)$$

Standar Deviasi :

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n} \quad (2.2)$$

n = jumlah sempel

$x_i$  = jumlah X ke i

X = simbol standar deviasi

Proses normalisasi ini bertujuan untuk memastikan bahwa seluruh variabel berkontribusi secara seimbang dalam pembentukan komponen utama. Dengan demikian, struktur variansi yang dianalisis oleh PCA benar-benar mencerminkan pola hubungan antarvariabel, bukan perbedaan skala semata.

### 2.2.2 Menentukan rasio varians dan nilai eigen

Setelah data dinormalisasi, tahap berikutnya adalah membentuk matriks kovarians. Matriks kovarians menggambarkan hubungan linier antarvariabel dan menjadi dasar utama dalam pembentukan PCA. Matriks kovarians dihitung menggunakan persamaan:

$$C = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (2.3)$$

C = matriks kovarians

N = jumlah sempel

X = rata rata vektor data

$x_i$  = Vektor data pada titik i

Dari matriks kovarians ini kemudian dihitung nilai eigen (eigenvalues) dan vektor eigen (eigenvectors). Nilai eigen menunjukkan besarnya variansi yang dapat dijelaskan oleh masing-masing komponen utama, sedangkan vektor eigen menunjukkan arah komponen utama dalam ruang data.

Sedangkan untuk menghitung nilai eigen, bisa menggunakan rumus persamaan berikut:

$$\det(C - \lambda I) = 0 \quad (2.4)$$

Keterangan:

- a) C = matriks kovarians
- b)  $\lambda$  = nilai eigen yang dicari
- c) I = matriks identitas dengan dimensi yang sama dengan C
- d)  $\det()$  = operator determinan

Nilai eigen yang besar menandakan bahwa komponen utama tersebut mampu menjelaskan variasi data yang signifikan. Sebaliknya, nilai eigen yang kecil menunjukkan bahwa komponen tersebut hanya menjelaskan variasi yang relatif kecil dan dapat diabaikan tanpa kehilangan informasi penting.

Untuk mengetahui kontribusi masing-masing komponen utama terhadap total variasi data, digunakan rasio varians (explained variance ratio) yang dirumuskan sebagai:

$$\text{variance ratio} = \frac{\lambda_i}{\sum \lambda^j} \quad (2.5)$$

$\lambda_i$ = nilai eigen dari komponen utama ke i  
 $\lambda^j$ = jumlah total dari semua nilai eigen  
 Rasio varians digunakan sebagai

dasar dalam menentukan jumlah komponen utama yang dipertahankan..

### 2.2.3 Menentukan komponen utama

Setelah vektor eigen diurutkan berdasarkan nilai eigen terbesar Komponen utama dibentuk dari kombinasi linier variabel asli berdasarkan vektor eigen. Komponen dengan nilai eigen tertinggi dipilih karena mampu memuat variasi informasi paling besar dari data. Jumlah komponen utama yang digunakan ditentukan berdasarkan persentase variasi total yang ingin dipertahankan dengan rumus

$$\text{variance ratio} = \frac{\lambda_i}{\sum \lambda^j} \quad (2.6)$$

$\lambda_i$ = nilai eigen dari komponen utama ke i  
 $\lambda^j$ = jumlah total dari semua nilai eigen

#### 2.2.4 Membuat grafik 2D

Data yang telah ditentukan komponen utamanya kemudian divisualisasikan dalam bentuk grafik. Visualisasi 3 dimensi digunakan untuk membentuk rentang nilai, yang hasilnya dapat ditampilkan kembali dalam grafik 2 dimensi.

Clustering merupakan teknik analisis data yang bertujuan untuk mengelompokkan objek ke dalam beberapa cluster berdasarkan tingkat kemiripan karakteristiknya. Prinsip dasar clustering adalah memaksimalkan kesamaan data dalam satu cluster (intra-cluster similarity) dan meminimalkan kesamaan antar cluster yang berbeda (inter-cluster dissimilarity) (Jain et al., 1999). Selain itu, clustering termasuk ke dalam metode unsupervised learning karena proses pengelompokan dilakukan tanpa menggunakan label kelas sebelumnya (Han et al., 2012)..

Clustering termasuk dalam kategori unsupervised learning, yaitu metode pembelajaran mesin yang tidak memerlukan informasi label atau kelas sebelumnya. Hal ini membedakan clustering dari metode klasifikasi yang bersifat supervised. Karena tidak bergantung pada label awal, clustering sangat efektif digunakan untuk mengeksplorasi struktur alami dan pola tersembunyi dalam data, khususnya pada data berskala besar dan kompleks (Wijayanto & Fathoni, 2021).

Dalam bidang klimatologi, clustering digunakan untuk mengelompokkan wilayah atau periode waktu berdasarkan kesamaan karakteristik iklim, seperti curah hujan, suhu, kelembapan, dan variabel atmosfer lainnya. Data klimatologi umumnya bersifat multivariat dan memiliki hubungan yang kompleks

antarvariabel, sehingga analisis langsung sering kali sulit dilakukan (Handoko et al., 2020).

Clustering memungkinkan pengelompokan pola iklim yang serupa ke dalam satu cluster, sehingga memudahkan identifikasi zona iklim, pola musim, serta variasi curah hujan antarwilayah. Hasil clustering dapat digunakan untuk memahami karakteristik iklim regional, mendukung perencanaan sumber daya air, serta membantu mitigasi risiko bencana hidrometeorologi (Wijayanto & Fathoni, 2021).

Dalam penelitian ini, proses clustering dilakukan dengan memanfaatkan hasil reduksi dimensi dari Principal Component Analysis (PCA). PCA digunakan sebagai tahap awal untuk mengekstraksi fitur-fitur utama dari data klimatologi yang telah dinormalisasi. Proses ini bertujuan untuk menyederhanakan data dengan mengurangi jumlah variabel, menghilangkan korelasi antarvariabel, serta mempertahankan informasi dominan yang merepresentasikan pola utama curah hujan.

Dalam penelitian ini, PCA digunakan untuk membentuk komponen utama dari variabel-variabel data klimatologi setelah melalui proses normalisasi. Komponen utama tersebut kemudian dimanfaatkan sebagai dasar dalam proses clustering. Meskipun clustering pada dasarnya bersifat tanpa label, hasil pengelompokan dapat diinterpretasikan dan diberi label tertentu berdasarkan karakteristik curah hujan yang dominan pada masing-masing cluster.

Dengan demikian, integrasi PCA dan clustering memungkinkan pengelompokan curah hujan berdasarkan pola yang terbentuk secara alami dari

data, bukan berdasarkan asumsi awal. Pendekatan ini memberikan hasil yang lebih objektif dan data-driven dalam mengidentifikasi pola curah hujan, sehingga dapat mendukung analisis klimatologi yang lebih akurat dan aplikatif.

### **2.3 Data Klimatologi**

Klimatologi termasuk dalam cabang ilmu atmosfer dan subbidang geografi fisik, yang merupakan bagian dari ilmu kebumian Studi klimatologi juga mencakup keterkaitan dengan oseanografi dan biogeokimia(Benny Hartanto et al., 2022). Data klimatologi adalah kumpulan informasi yang merekam kondisi cuaca dan iklim dalam jangka waktu yang panjang pada suatu wilayah tertentu yang diperoleh dari pengamatan Stasiun badan meteorologi,Klimatologi dan Geofisika dalam suatu daerah(Sarker, 2022).

Variabel-variaabel yang umum dicatat dalam data klimatologi meliputi curah hujan, kelembaban udara, temperatur, tekanan udara, kecepatan dan arah angin, serta penyinaran matahari. Data ini memiliki peran penting dalam berbagai bidang, seperti perencanaan tata ruang, pengelolaan sumber daya air, perencanaan pertanian, mitigasi bencana, hingga penelitian terkait perubahan iklim global(Abror et al., 2023)

Dalam konteks penelitian ini, data klimatologi digunakan sebagai dasar untuk menganalisis dan mengelompokkan pola curah hujan melalui metode Principal Component Analysis (PCA) dan Clustering, sehingga hasilnya dapat mendukung perencanaan pengelolaan sumber daya air dan strategi mitigasi bencana hidrometeorologi.

## 2.4 Silhouette Score

Silhouette Score merupakan salah satu metode validasi internal yang paling umum digunakan untuk mengevaluasi kualitas hasil clustering. Metode ini diperkenalkan untuk mengukur sejauh mana suatu objek data cocok dengan cluster tempat ia berada dibandingkan dengan cluster lain yang paling dekat. Dengan kata lain, Silhouette Score menilai kohesi (cohesion) dan separasi (separation) secara bersamaan dalam satu ukuran kuantitatif (Rousseeuw, 1987)

Konsep dasar Silhouette Score didasarkan pada perbandingan dua nilai utama, yaitu jarak rata-rata suatu data terhadap data lain dalam cluster yang sama dan jarak rata-rata terhadap data pada cluster terdekat yang berbeda. Jika suatu data memiliki jarak yang kecil terhadap anggota cluster-nya sendiri dan jarak yang besar terhadap cluster lain, maka data tersebut dianggap telah terkelompok dengan baik (Hendrastuty, 2024).

Rumus silhouette score

$$S(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (2.7)$$

a(i) = jarak rata rata antara i dan semua titik data lain dalam cluster yang sama  
b(i) = jarak rata rata dari i ke semua kluster terdekat

Nilai Silhouette berada pada rentang -1 hingga 1, dengan interpretasi sebagai berikut(Prasetyo et al., 2024):

1. Nilai mendekati 1 menunjukkan bahwa data terkelompok dengan sangat baik.
2. Nilai mendekati 0 menunjukkan bahwa data berada di batas antar klaster.
3. Nilai mendekati -1 menunjukkan bahwa data kemungkinan salah penempatan klaster.

Dalam praktiknya, rata-rata Silhouette Score dari seluruh data digunakan sebagai indikator kualitas keseluruhan hasil clustering. Semakin tinggi nilai rata-rata Silhouette Score, semakin baik kualitas pengelompokan yang dihasilkan. Oleh karena itu, Silhouette Score sering dimanfaatkan untuk menentukan jumlah cluster yang optimal, dengan cara membandingkan nilai Silhouette pada berbagai konfigurasi jumlah cluster (Saidah & Raiza Athifah Nisrina, 2024).

Dalam konteks analisis data klimatologi, Silhouette Score berperan penting dalam mengevaluasi sejauh mana pola curah hujan yang terbentuk benar-benar terpisah dengan baik antarcluster. Validasi ini memastikan bahwa cluster yang dihasilkan tidak hanya terbentuk secara matematis, tetapi juga memiliki tingkat kejelasan dan konsistensi internal yang baik

## 2.5 Validasi Eksternal

Validasi eksternal merupakan metode evaluasi clustering yang dilakukan dengan membandingkan hasil clustering dengan label acuan (ground truth) atau informasi kelas yang telah diketahui sebelumnya. Tujuan utama validasi eksternal adalah untuk menilai sejauh mana hasil clustering mencerminkan struktur data yang sebenarnya berdasarkan referensi eksternal tersebut. Berbeda dengan validasi internal yang hanya bergantung pada struktur internal data, validasi eksternal membutuhkan data tambahan berupa label atau kategori pembanding. Dengan adanya informasi ini, hasil clustering dapat dievaluasi secara lebih objektif, khususnya dalam konteks kesesuaian antara cluster yang terbentuk dan kelas yang diharapkan (Saidah & Raiza Athifah Nisrina, 2024).

Dalam penelitian klimatologi, validasi eksternal dapat dilakukan dengan membandingkan hasil clustering pola curah hujan dengan kategori iklim yang telah ditetapkan sebelumnya, misalnya berdasarkan zona iklim, klasifikasi curah hujan, atau hasil pengelompokan menggunakan metode supervised. Pendekatan ini memungkinkan peneliti untuk menilai sejauh mana hasil clustering mampu merepresentasikan kondisi iklim yang sebenarnya(Wijaya & Anondho, 2021).

Penggunaan validasi eksternal memberikan keunggulan dalam mengukur tingkat akurasi dan konsistensi hasil clustering terhadap acuan yang jelas. Namun, metode ini memiliki keterbatasan karena sangat bergantung pada ketersediaan dan kualitas data label acuan. Oleh karena itu, dalam banyak penelitian, validasi eksternal sering dikombinasikan dengan validasi internal untuk memperoleh evaluasi clustering yang lebih komprehensif(Wijaya & Anondho, 2021).

## **2.6 Konsep Pembentukan Tiga Cluster dalam Analisis Clustering**

Penentuan jumlah cluster dan rentang nilainya merupakan aspek penting dalam analisis clustering karena memengaruhi kejelasan pola dan interpretasi hasil pengelompokan data. Salah satu pendekatan yang dapat digunakan dalam menentukan rentang nilai cluster adalah dengan memanfaatkan hasil transformasi Principal Component Analysis (PCA). PCA tidak hanya berfungsi sebagai metode reduksi dimensi, tetapi juga mampu merepresentasikan distribusi data dalam ruang komponen utama yang mencerminkan variasi dominan dalam data(Nuraeni et al., 2024).

Almais et al. (2023) menjelaskan bahwa nilai koordinat data pada komponen utama, khususnya komponen utama pertama, dapat digunakan sebagai

dasar dalam pembentukan cluster berdasarkan rentang nilai tertentu. Komponen utama pertama memuat proporsi variasi terbesar dari data, sehingga distribusi nilainya dapat mencerminkan perbedaan karakteristik data secara signifikan. Dengan menganalisis sebaran nilai komponen utama tersebut, pola pemisahan alami data (*natural grouping*) dapat diidentifikasi.

Almais et al. (2023) menetapkan pembagian rentang nilai cluster berdasarkan distribusi nilai PCA sebagai berikut: cluster pertama merepresentasikan data dengan nilai komponen utama rendah, cluster kedua berada pada rentang nilai menengah, dan cluster ketiga mencerminkan data dengan nilai komponen utama tinggi. Penentuan batas antarcluster dilakukan berdasarkan perubahan signifikan pada distribusi nilai PCA, sehingga batas cluster bersifat data-driven dan tidak ditentukan secara

## **BAB III**

### **METODE PENELITIAN**

#### **3.1 Data Preparation**

Dalam penelitian ini, menggunakan data sekunder yang diperoleh dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) Stasiun Geofisika Kelas III Malang. Data tersebut berupa rekapitulasi data klimatologi. Dengan menggunakan data klimatologi dengan variabel pada tabel 3.1

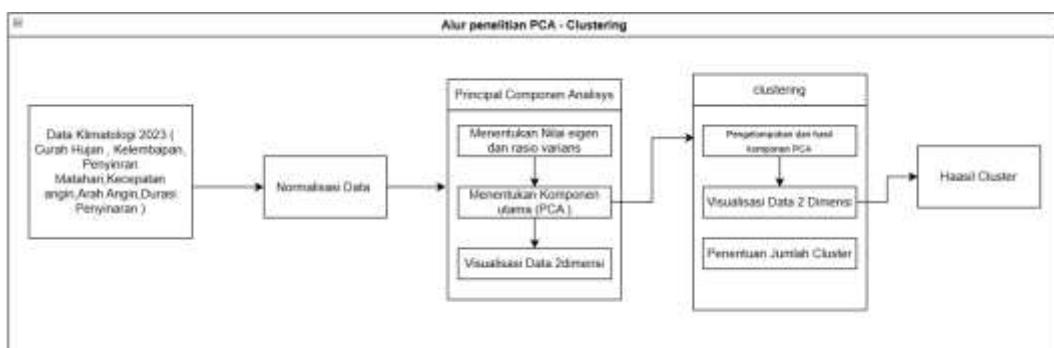
Tabel.3.1 Data Klimatologi

No	Curah Hujan (mm)	Kelembaban	Penyinaran Matahari	Tekanan (mbar)	Kecepatan Angin (0)	Arah Angin	Durasi Penyinaran Matahari
1	0.1	67.78	26.25	945.5	0.00075	315	0.000583
2	1.6	78.09	33.75	945.3	0.000472	180	0.00075
3	4.4	81.81	63.75	945.5	0.000417	135	0.001417
4	6.2	70.21	77.50	945.3	0.000528	180	0.001722

Pemilihan data ini berdasarkan pada pertimbangan bahwa variabel-variabel tersebut memiliki peran penting dalam menggambarkan kondisi iklim harian suatu wilayah. Selain itu, penggunaan data sekunder dari BMKG dinilai lebih akurat karena diperoleh melalui instrumen pengamatan resmi. Dengan adanya data tersebut, penelitian ini dapat melakukan analisis pola dan kecenderungan iklim secara lebih sistematis, sekaligus memberikan landasan yang kuat untuk mengidentifikasi hubungan antarvariabel klimatologi di wilayah Malang

### 3.2 Alur Penelitian

Bab ini memberikan penjelasan tentang proses yang digunakan dalam menerapkan metode PCA dan Clustering untuk melakukan segmentasi data klimatologi. Gambar 3.2 menunjukkan desain sistem yang digunakan dalam penelitian ini. Sistem ini bertujuan untuk melakukan segmentasi pola cuaca berdasarkan data klimatologi.



Gambar 3.2 Alur Penelitian

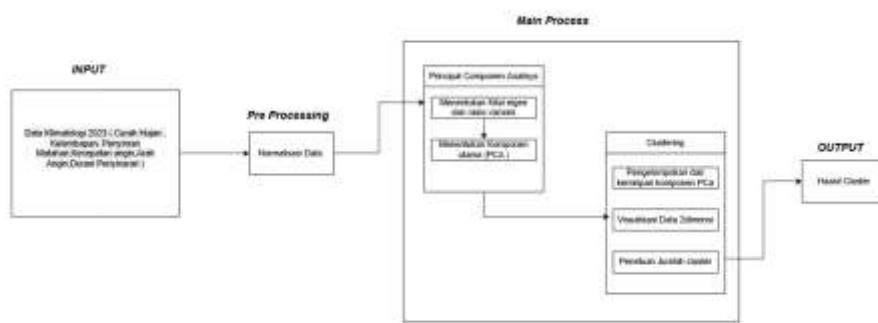
Tahap awal dalam penerapan PCA dimulai dengan input data klimatologi, Setelah itu normalisasi untuk menyamakan skala antar variabel agar tidak ada variabel yang memiliki nilai lebih tinggi hanya karena perbedaan satuan pengukuran. Setelah data dinormalisasi lakukan perhitungan rasio varians dari setiap variabel dan nilai eigen dari matriks kovarians untuk mengetahui seberapa besar peranan setiap variabel dalam menjelaskan keseluruhan variasi data. Nilai eigen menunjukkan besarnya varians yang dapat dijelaskan oleh masing-masing komponen utama (principal component). Rasio varians digunakan untuk melihat proporsi kontribusi tiap komponen utama terhadap keseluruhan variasi data.

Hasil perhitungan ini digunakan untuk memilih komponen utama PCA, yaitu komponen yang mampu menampung informasi paling banyak dari data asli.

Sehingga data yang semula kompleks dapat direduksi lebih sederhana tanpa kehilangan informasi penting. Data hasil reduksi kemudian digunakan sebagai masukan (*input*) untuk tahap Clustering. Melalui metode ini data hasil pca divisualisasikan dalam bentuk grafik dua dimensi untuk memperlihatkan hubungan antar kelompok pola hujan. Kemudian dilakukan pengelompokan data berdasarkan tingkat kemiripan karakteristik dengan menggunakan ukuran jarak dan metode penggabungan tertentu . Hasil pengelompokan Pada tahap akhir, dimasukkan ke dalam beberapa cluster dengan karakteristik yang serupa, yang selanjutnya dapat dianalisis dan diinterpretasikan berdasarkan data klimatologi dari BMKG.

### 3.3 Desain System

Desain sistem ini menggambarkan tahapan utama yang dimulai dari pengumpulan data klimatologi dari BMKG hingga proses pengelompokan data menggunakan metode clustering. Gambar 3.3 menyajikan alur kerja sistem yang dirancang untuk mengidentifikasi pola spasial dan temporal kondisi iklim secara lebih terstruktur dan sistematis.



Gambar 3.3 Desain System

### 3.3.1 Input

Tahap input dimulai dengan pengumpulan data klimatologi tahun 2023 yang terdiri dari variabel curah hujan, kelembapan udara, penyinaran matahari, kecepatan angin, arah angin, dan durasi penyinaran. Seluruh variabel ini dipilih karena memiliki pengaruh signifikan dalam menggambarkan kondisi iklim suatu wilayah. Data tersebut menjadi dasar bagi proses analisis lebih lanjut yang melibatkan reduksi dimensi dan pengelompokan karakteristik iklim.

### 3.3.2 Pre Processing

Pada tahap pre-processing, data klimatologi yang terkumpul terlebih dahulu melalui proses normalisasi untuk memastikan setiap variabel berada pada skala yang sebanding. Normalisasi menjadi langkah penting karena variabel klimatologi umumnya memiliki satuan dan rentang nilai yang berbeda, misalnya curah hujan dalam milimeter, kelembapan dalam persen, kecepatan angin dalam meter per detik, hingga durasi penyinaran dalam jam.

Rumus mean :

$$X_{Standart} = \frac{0,1+1,6+4,4+6,2}{4} = 3,075 \quad (3.1)$$

Dengan  $X$  adalah nilai data asli,  $mean(X)$  merupakan nilai rata-rata dari data, dan  $standar deviasi(X)$  adalah simpangan baku. Standar deviasi dihitung dengan rumus:

$$1. (0,1-3,075)2 = 8,85$$

$$2. (1,6 - 3,075) = 2,17$$

$$3. (4,4 - 3,075) = 1,76$$

$$4. (6,2 - 3,075) = 9,77$$

Rumus standar Deviasi :

$$\bar{X} = \sqrt{\frac{22,55}{3}} = \sqrt{7,52} = 2,74 \quad (3.2)$$

Perhitungan Standarscaler :

$$Z_1 = \frac{0,1 - 3,075}{2,74} = -1,09$$

$$Z_2 = \frac{1,6 - 3,075}{2,74} = -0,54$$

$$Z_3 = \frac{4,4 - 3,075}{2,74} = 0,48$$

$$Z_4 = \frac{6,2 - 3,075}{2,74} = 1,14$$

Hasilnya, seluruh variabel berada pada skala yang setara sehingga kontribusi tiap variabel terhadap analisis berikutnya bersifat proporsional. Tahap ini memastikan bahwa proses PCA dapat berjalan secara optimal tanpa dipengaruhi perbedaan skala antar variabel klimatolog

### 3.3.3 Main Process

Main Process diawali dengan penggunaan data yang telah dinormalisasi sebagai input untuk menjalankan Principal Component Analysis (PCA). Tahap PCA ini bertujuan mereduksi dimensi data serta mengidentifikasi variabel yang memiliki kontribusi terbesar terhadap variasi keseluruhan.

a. Menentukan Nilai Eigen dan Rasio Varians

Setelah melalui tahap normalisasi, langkah berikutnya dalam PCA adalah menghitung nilai eigen (*eigenvalue*) dan vektor eigen (*eigenvector*) dari matriks kovarians. Nilai eigen menggambarkan besarnya varians yang dijelaskan oleh setiap komponen utama, sedangkan vektor eigen menunjukkan arah dari komponen tersebut dalam ruang data.

Contoh Perhitungan manual menggunakan Matriks kovarians untuk 5 data pada variabel curah hujan dan kelembapan :

1.  $(-1,09)(-1,02) = 1,11$
2.  $(-0,54)(0,55) = -0,30$
3.  $(0,48)(1,12) = 0,54$
4.  $(1,14)(-0,65) = -0,74$

$$\sum (Z_{CH} \cdot Z_{CH}) = 0,61 \quad (3.4)$$

$$Cov ( CH, KL) = \frac{0,61}{3} 0,20 \quad (3.5)$$

Hasil matrik kovarian

$$\mathbf{C} = \begin{bmatrix} \text{Var}(CH) & \text{Cov}(CH, KL) \\ \text{Cov}(KL, CH) & \text{Var}(KL) \end{bmatrix} = \begin{bmatrix} 1,00 & 0,20 \\ 0,20 & 1,00 \end{bmatrix} \quad (3.6)$$

Nilai kovarians sebesar 0,20 menunjukkan bahwa curah hujan dan kelembapan memiliki hubungan linier positif yang lemah. Artinya, peningkatan

curah hujan cenderung diikuti oleh peningkatan kelembapan, meskipun hubungan tersebut tidak terlalu kuat.

Dari hasil matrik kovarian diatas lalu dihitung nilai eigen dengan rumus nilai eigen :

$$\det(C - \lambda I) = 0 \quad (3.7)$$

$$\begin{vmatrix} 1 - \lambda & 0,20 \\ 0,20 & 1 - \lambda \end{vmatrix} = 0$$

$$(1 - \lambda)^2 - (0,20)^2 = 0$$

$$(1 - \lambda)^2 = 0,04$$

$$1 - \lambda = \pm 0,20$$

Sehingga diperoleh dua nilai eigen 1,20 dan 0,80

Setelah menghitung matriks kovarians, tahap berikutnya adalah mencari nilai eigen (*eigenvalue*) dan vektor eigen (*eigenvector*). Nilai eigen merepresentasikan besarnya varians yang dapat dijelaskan oleh masing-masing komponen utama, sementara vektor eigen menunjukkan arah dari komponen tersebut dalam ruang data. Nilai varians yang tinggi menunjukkan data memiliki penyebaran luas, sedangkan varians rendah menunjukkan data lebih terkonsentrasi di sekitar nilai rata-rata.

$$\begin{aligned} \text{Rasio Varians}_{PC1} &= \frac{1,20}{2,00} = 0,60 \text{ (60\%)} \\ \text{Rasio Varians}_{PC2} &= \frac{0,80}{2,00} = 0,40 \text{ (40\%)} \end{aligned} \quad (3.7)$$

Dari rasio varian diatas nilai PC1 memiliki rasio varian 60% dan PC2 40% Semakin besar nilai eigen, semakin besar pula proporsi keragaman data yang mampu dijelaskan oleh komponen utama terkait. Dengan demikian, nilai eigen menjadi indikator penting dalam memilih dimensi yang relevan untuk analisis lebih lanjut.

b. Menetukan Komponen Utama

Setelah nilai eigen dan vektor eigen diperoleh, vektor eigen diurutkan berdasarkan nilai eigen terbesar. Komponen utama kemudian dibentuk dari kombinasi linier variabel asli menggunakan vektor eigen. Komponen dengan nilai eigen terbesar diprioritaskan karena mampu memuat variasi informasi paling signifikan dari data. Jumlah komponen utama yang digunakan dipilih berdasarkan variansi yang mampu merepresentasikan karakteristik utama data petir,

c. Membuat Grafik 2D

Jika data petir memiliki banyak variabel spasial-temporal, PCA akan mereduksi kompleksitasnya menjadi beberapa komponen utama. Dua komponen pertama (PC1 dan PC2) kemudian dipilih karena mampu merepresentasikan sebagian besar variasi pola Hujan yang terjadi pada data. Hasil reduksi ini divisualkan dalam bentuk *scatter plot* dua dimensi untuk memperlihatkan pola distribusi data secara lebih sederhana. Melalui grafik ini, kecenderungan pembentukan cluster maupun pemisahan antar kelompok data dapat diamati dengan lebih jelas, sehingga mendukung proses interpretasi hasil analisis.

### 3.3.4. Output

Tahap akhir penelitian menghasilkan output berupa cluster atau kelompok data klimatologi yang memiliki karakteristik iklim serupa. Hasil pengelompokan ini digunakan untuk mengidentifikasi pola iklim tertentu, membedakan wilayah berdasarkan kesamaan kondisi cuaca, serta memberikan gambaran yang lebih komprehensif mengenai variasi iklim di wilayah penelitian. Informasi yang dihasilkan dari proses clustering ini dapat dimanfaatkan sebagai dasar analisis klimatologi lanjutan maupun sebagai bahan pendukung dalam bidang perencanaan wilayah, pertanian, mitigasi bencana, dan studi lingkungan.

Penentuan cluster dalam penelitian ini dilakukan berdasarkan dua pendekatan utama, yaitu rentang nilai komponen utama sebagaimana direkomendasikan dalam paper rujukan, serta pola lonjakan (break) pada persebaran data hasil PCA. Setelah data klimatologi direduksi dimensinya menggunakan Principal Component Analysis (PCA), setiap data direpresentasikan dalam ruang baru yang dibentuk oleh komponen utama, khususnya PC1 dan PC2, yang memuat proporsi variansi terbesar.

Berdasarkan kajian literatur dan paper rujukan, pembentukan tiga cluster dipilih karena mampu merepresentasikan variasi utama data tanpa kehilangan struktur alami pola iklim. Jumlah tiga cluster ini umumnya diasosiasikan dengan kondisi iklim rendah, sedang, dan tinggi, atau dalam konteks curah hujan dapat diinterpretasikan sebagai pola curah hujan rendah, menengah, dan tinggi. Rentang nilai setiap cluster ditentukan dengan mengamati distribusi skor komponen utama,

khususnya nilai PC1 sebagai sumbu utama yang paling dominan menjelaskan variabilitas data.

Selain mengacu pada rentang nilai dari model 3 cluster , penentuan batas antar cluster juga mempertimbangkan lonjakan (discontinuity) pada persebaran data. Lonjakan ini terlihat sebagai perubahan jarak yang signifikan antar kelompok titik data pada visualisasi PCA dua dimensi. Titik-titik data yang memiliki jarak relatif rapat dikelompokkan ke dalam satu cluster, sedangkan adanya jarak yang cukup besar antar kelompok titik menjadi indikator batas alami antar cluster. Pendekatan ini sejalan dengan konsep pemisahan alami (natural breaks), di mana cluster terbentuk berdasarkan struktur internal data, bukan semata-mata pembagian interval yang bersifat arbitrer.

Pendekatan kombinasi antara rentang nilai komponen utama dari paper rujukan dan lonjakan persebaran data hasil PCA ini memastikan bahwa hasil clustering bersifat objektif, data-driven, serta selaras dengan teori dan penelitian sebelumnya. Hal ini memperkuat validitas hasil pengelompokan dan meningkatkan keandalan interpretasi pola iklim yang dihasilkan.

### **2.3 Skenario Pengujian**

Skenario pengujian dalam penelitian menggunakan 2 metode, yaitu menggunakan Silhouette Score dan validasi eksternal data.

#### **2.3.1 Silhouette Score**

Silhouette Score digunakan sebagai metode validasi internal untuk mengevaluasi kualitas hasil clustering berdasarkan kedekatan data dalam satu

kluster dan tingkat pemisahannya terhadap kluster lain. Pada penelitian ini, Silhouette Score dihitung menggunakan hasil clustering dari data klimatologi yang telah direduksi dimensinya dengan PCA. Sebagai contoh, pada hasil pengelompokan tiga kluster, salah satu kluster menunjukkan nilai Silhouette Score yang tinggi karena jarak antar data di dalam kluster tersebut relatif kecil, sedangkan jaraknya terhadap kluster terdekat lainnya cukup besar. Kondisi ini mengindikasikan bahwa data terkelompok dengan baik dan batas antar kluster terbentuk secara jelas. Nilai rata-rata Silhouette Score yang diperoleh kemudian digunakan untuk menilai kualitas keseluruhan clustering serta memastikan bahwa jumlah kluster yang dipilih sudah optimal.

### 2.3.2 Validasi eksternal data

Validasi eksternal dilakukan untuk menilai kesesuaian hasil clustering dengan kondisi klimatologi yang sebenarnya menggunakan data referensi dari BMKG. Pada tahap ini, hasil clustering berbasis PCA dibandingkan dengan kategori klimatologi yang terdapat pada data BMKG, khususnya terkait karakteristik curah hujan. Sebagai contoh, kluster pertama didominasi oleh data dengan curah hujan rendah menurut klasifikasi BMKG, kluster kedua merepresentasikan curah hujan sedang, dan kluster ketiga menggambarkan curah hujan tinggi. Kesesuaian karakteristik kluster dengan data referensi tersebut menunjukkan bahwa hasil clustering tidak hanya baik secara struktur internal, tetapi juga relevan secara ilmiah dan mampu merepresentasikan pola iklim aktual di wilayah penelitian.

## 2.4 Pseudocode Clustering

Pseudocode clustering adalah bentuk penulisan algoritma clustering yang dibuat tidak dalam bahasa pemrograman murni, melainkan dalam bentuk deskripsi langkah-langkah yang menyerupai kode.

Table 3.4 Pseudocode algoritma Clustering

<b>Algoritma Clustering menggunakan PCA-Clustering dalam label data</b>	
<b>Input :</b>	membaca data menggunakan fungsi <code>pd.read_csv ("Klimatologi.csv")</code>
<b>Process</b>	<ul style="list-style-type: none"> <li>Normalisasi data menggunakan fungsi <code>Standarscale()</code> and <code>fit_transform()</code></li> <li>Membuat rasio varians menggunakan fungsi <code>explained_variance_ratio_()</code></li> <li>Membuat eigenvalues menggunakan fungsi <code>Explained_variance()</code></li> <li>Membuat komponen pca menggunakan <code>PCA()</code></li> <li>Membuat data frame dari komponen PCA <code>pd.DataFrame()</code></li> <li>Visualisasi 2 dimensi pada PCA menggunakan <code>scatterplot()</code></li> <li>Visualisasi 3 dimensi pada PCA menggunakan <code>scatterplot()</code></li> <li>Membuat label data dengan menggunakan percabangan (<i>if-else</i>) dan perulangan (<i>for</i>). Tambahkan label cluster dataframe hasil pca</li> <li>Visualisasi hasil clustering menggunakan <code>scatterplot()</code></li> </ul>
<b>Output</b>	<ul style="list-style-type: none"> <li>Menggabungkan dataframe dari komponen PCA dengan label cluster menggunakan <code>concat()</code></li> <li>Hasil visualisasi dari proses pelabelan data menggunakan PCA yang divalidasi dengan data asli dari bmkg</li> </ul>

## **BAB IV**

### **HASIL DAN PEMBAHASAN**

#### **4.1 Preprocessing Data**

Tahap *preprocessing data* dilakukan untuk memastikan bahwa data klimatologi yang digunakan siap dan layak untuk dianalisis menggunakan metode Principal Component Analysis (PCA) dan .

Tahap Preprocessing data dilakukan dengan 3 tahap berikut :

##### **4.4.1 Teknik Pengumpulan Data**

Penelitian ini menggunakan data sekunder yang diperoleh dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) Stasiun Geofisika Kelas III Malang. Data tersebut berupa rekapitulasi data klimatologi harian yang terdiri dari 7 variabel.

Pemilihan variabel-variabel ini didasarkan pada pertimbangan bahwa masing-masing memiliki pengaruh yang signifikan terhadap kondisi iklim di suatu wilayah. Selain itu, penggunaan data dari BMKG menjamin keakuratan karena diperoleh melalui pengamatan instrumen resmi dan terstandar.

##### **4.4.2 Listwise Deletion**

Salah satu permasalahan yang ditemukan adalah adanya nilai kosong (*missing value*) pada variabel curah hujan. Kondisi ini umum terjadi dalam data klimatologi karena tidak setiap hari terjadi hujan, sehingga pencatatan curah hujan

tidak selalu menghasilkan nilai. Dari total 365 data klimatologi, terdapat 159 data pada variabel curah hujan yang tidak memiliki nilai.

khususnya pada hari-hari tanpa hujan sehingga menghasilkan banyak nilai kosong. Dari total 365 data klimatologi, terdapat 159 data pada variabel curah hujan yang tidak memiliki nilai. Untuk memastikan bahwa data yang digunakan pada tahap analisis benar-benar bersih dan bebas dari nilai hilang, penelitian ini menggunakan teknik drop data (listwise deletion), yaitu menghapus seluruh baris data yang mengandung missing value pada variabel curah hujan. Pendekatan ini dipilih karena nilai kosong pada variabel tersebut tidak dapat diestimasi secara akurat dan berpotensi menurunkan kualitas analisis.

Setelah proses penghapusan data hilang dilakukan, jumlah data yang dapat digunakan dalam analisis menjadi 203 data klimatologi dari total 365 data awal. Dengan demikian, dataset akhir yang digunakan adalah data yang sepenuhnya lengkap tanpa missing value sehingga lebih layak untuk diproses pada tahap analisis berikutnya.

## 4.2 Normalisasi Data

Normalisasi data merupakan tahapan penting dalam proses analisis data klimatologi karena variabel yang digunakan dalam penelitian ini memiliki satuan dan rentang nilai yang berbeda-beda. Variabel curah hujan dinyatakan dalam milimeter (mm), kelembapan dalam persen (%), tekanan udara dalam milibar (mbar), kecepatan angin dalam meter per detik (m/s), serta arah dan durasi penyinaran yang memiliki skala pengukuran tersendiri. Perbedaan skala tersebut

berpotensi menyebabkan variabel dengan rentang nilai yang lebih besar mendominasi hasil analisis, khususnya pada metode yang berbasis jarak dan variansi seperti Principal Component Analysis (PCA).

#### **4.2.1 StandarScaler**

Proses normalisasi pada penelitian ini dilakukan menggunakan metode StandardScaler, yaitu metode normalisasi yang mentransformasikan data berdasarkan rata-rata (mean) dan standar deviasi dari masing-masing variabel. Setiap nilai data dikurangi dengan nilai rata-rata variabel kemudian dibagi dengan standar deviasinya. Proses ini bertujuan untuk memusatkan data pada nilai nol dan menyamakan tingkat penyebaran data.

Setelah dilakukan normalisasi, setiap variabel memiliki nilai rata-rata (mean) sebesar 0, yang berarti bahwa data telah dipusatkan pada titik rata-rata variabel asli. Nilai 0 menunjukkan posisi rata-rata, sedangkan nilai positif dan negatif masing-masing menunjukkan bahwa data berada di atas atau di bawah rata-rata. Dengan demikian, perbedaan posisi data terhadap rata-rata dapat diamati secara langsung tanpa dipengaruhi oleh satuan asli variabel.

Selain itu, hasil normalisasi juga menghasilkan standar deviasi sebesar 1, yang menunjukkan bahwa penyebaran data dinyatakan dalam satuan deviasi standar. Besarnya nilai hasil normalisasi menunjukkan jarak suatu data terhadap rata-rata dalam satuan standar deviasi.

Hasil proses normalisasi menunjukkan bahwa seluruh variabel telah berada pada skala yang relatif seimbang

Tabel 4.1 Hasil Normalisasi

Curah Hujan (mm)	Kelembapan (%)	Penyinaran Matahari (%)	Tekanan (mbar)	Kecepatan Angin (m/s)	Arah Angin (°)	Durasi Penyinaran (Jam)
-0.750180	-3.254711	-0.851230	-0.961320	0.494522	1.813002	0.850134
-0.642231	-1.161194	-0.581399	-0.973439	-0.296557	0.238689	0.580342
-0.440728	-0.406295	0.497925	-0.961320	-0.454773	-0.286082	0.498828
-0.311190	-2.761714	0.992615	-0.973439	-0.138341	0.238689	0.993448

Normalisasi dengan StandardScaler menjadi penting karena variabel klimatologi yang digunakan dalam penelitian ini memiliki satuan dan rentang nilai yang berbeda, seperti curah hujan dalam milimeter, kelembapan dalam persen, tekanan udara dalam milibar, dan kecepatan angin dalam meter per detik. Tanpa normalisasi, variabel dengan rentang nilai yang lebih besar berpotensi mendominasi proses analisis.

### 4.3 Component Anaisys

Component Analysis dalam penelitian ini dilakukan menggunakan metode Principal Component Analysis (PCA) dengan tujuan untuk mereduksi dimensi data klimatologi yang memiliki banyak variabel sekaligus mengekstraksi informasi utama yang paling berpengaruh terhadap variasi data. PCA bekerja dengan mentransformasikan variabel-variabel asli yang saling berkorelasi menjadi sejumlah komponen utama baru yang bersifat tidak berkorelasi satu sama lain. Setiap komponen utama merepresentasikan kombinasi linier dari variabel asli dan

disusun berdasarkan besarnya variansi data yang mampu dijelaskan. Penerapan PCA pada penelitian ini dilakukan setelah seluruh data melalui tahap preprocessing dan normalisasi menggunakan StandardScaler. Langkah ini penting untuk memastikan bahwa setiap variabel memiliki kontribusi yang seimbang dalam pembentukan komponen utama, sehingga hasil PCA benar-benar merefleksikan struktur variasi data klimatologi yang sesungguhnya dan tidak dipengaruhi oleh perbedaan skala pengukuran antarvariabel.

#### **4.3.1 Menetukan Nilai eigen dan Varian Rasio**

Nilai eigenvalue dan variance ratio merupakan indikator utama dalam menentukan jumlah serta kontribusi setiap komponen utama pada hasil *Principal Component Analysis* (PCA).

Nilai eigenvalue diperoleh dari hasil dekomposisi matriks kovarians antarvariabel dan menunjukkan besarnya variasi data yang mampu dijelaskan oleh masing-masing komponen utama. Setiap komponen utama memiliki satu nilai eigenvalue, di mana nilai eigenvalue yang lebih besar menandakan bahwa komponen tersebut mengandung informasi yang lebih dominan dalam merepresentasikan data asli.

Sementara itu, variance ratio atau rasio variansi merupakan nilai yang diperoleh dari perbandingan antara eigenvalue suatu komponen utama dengan total eigenvalue seluruh komponen. Nilai ini menggambarkan proporsi atau persentase variasi total data yang dapat dijelaskan oleh masing-masing komponen utama, sehingga dapat digunakan untuk mengevaluasi tingkat kontribusi setiap komponen terhadap keseluruhan informasi data.

Dengan demikian, komponen utama dengan variance ratio yang tinggi dipertahankan karena mampu mewakili sebagian besar variasi data, sedangkan komponen dengan variance ratio yang rendah tidak dipertahankan karena kontribusinya relatif kecil.

4.2 Tabel hasil Eigenvalue dan Variance Ratio

Komponen	Eigenvalue	Variance Ratio
PC1	2.384272	0.338932
PC2	1.713957	0.243645
PC3	1.036050	0.147278
PC4	0.992376	0.141070
PC5	0.607374	0.086340
PC6	0.300614	0.042733
PC7	0.000011	0.000002

Berdasarkan hasil perhitungan nilai eigen dan variance ratio pada tabel di atas, terlihat bahwa beberapa komponen utama memiliki kontribusi yang cukup signifikan terhadap total variasi data. Komponen utama pertama (PC1) memiliki nilai eigen sebesar 2.384272 dengan variance ratio 0.338932, yang berarti komponen ini mampu menjelaskan sekitar 33,89% dari total variasi data. Hal ini menunjukkan bahwa PC1 menjadi komponen paling dominan dan berperan penting dalam merepresentasikan pola umum variasi pada dataset klimatologi.

Komponen utama kedua (PC2) memiliki nilai eigen sebesar 1.713957 dengan variance ratio 0.243645, atau sekitar 24,36% dari total variasi. Kontribusi ini masih cukup besar, sehingga PC2 melengkapi informasi tambahan yang tidak tercakup pada PC1. Secara bersama-sama, PC1 dan PC2 mampu menjelaskan sekitar 58,25% total variasi data.

Komponen berikutnya, yaitu PC3, memiliki nilai eigen sebesar 1.036050 dengan variance ratio 0.147278 (sekitar 14,73%). PC3 tetap memberikan kontribusi penting meskipun tidak sebesar dua komponen pertama. PC4 juga memiliki nilai eigen tinggi, yaitu 0.992376 dengan variance ratio 14,10%, sehingga empat komponen utama (PC1–PC4) secara kumulatif mampu menjelaskan sekitar 82,98% variasi dalam data.

Sementara itu, PC5 dan PC6 memberikan kontribusi yang lebih kecil, masing-masing sebesar 8,63% dan 4,27%. Meski demikian, keduanya masih menambah informasi variabilitas data, terutama untuk pola-pola yang lebih minor. Komponen terakhir, PC7, memiliki nilai eigen yang sangat kecil (0.000011) dengan variance ratio hampir nol (0.0002%), sehingga kontribusinya dapat diabaikan.

Secara keseluruhan, hasil ini menunjukkan bahwa sebagian besar variasi data dapat dijelaskan oleh 2 komponen utama pertama (PC1–PC2) dengan total kontribusinya sudah melebihi 50%, yang umumnya dianggap sangat memadai dalam analisis PCA. Dengan demikian, reduksi dimensi menjadi dua komponen utama sudah optimal untuk mempertahankan informasi penting dalam dataset, dan komponen-komponen ini dapat digunakan untuk analisis lanjutan seperti visualisasi, clustering, atau interpretasi faktor dominan dalam data klimatologi.

#### 4.3.2 Menentukan komponen utama

komponen utama dilakukan dengan mempertimbangkan besarnya nilai eigenvalue, rasio varians, serta faktor loading variabel.

Tabel 4.3.hasil faktor loading variabel

Variabel	PC1	PC2	PC3	PC4	PC5	PC6	PC7
----------	-----	-----	-----	-----	-----	-----	-----

Curah Hujan (mm)	-0.2546	0.0898	0.3880	0.8535	-0.2284	0.0291	-0.0000
Kelembapan (%)	-0.6180	0.3765	-0.2962	0.2803	0.5201	-0.2109	0.0000
Penyinaran Matahari (%)	0.9599	0.0330	0.0452	0.1917	0.2062	-0.0340	0.0024
Tekanan (mbar)	0.0145	0.9201	-0.0271	0.0466	0.0891	0.3838	-0.0000
Kecepatan Angin (m/s)	0.2576	0.8239	-0.0862	0.1117	0.3793	0.3099	0.0000
Arah Angin (°)	-0.1678	0.1915	0.8863	-0.3119	0.2185	-0.0981	0.0000
Durasi Penyinaran (Jam)	0.9599	0.0318	0.0455	0.1913	0.2067	-0.0339	0.0024

Hasil analisis faktor loading menunjukkan bahwa komponen utama pertama (PC1) didominasi secara kuat oleh dua variabel radiasi, yaitu Penyinaran Matahari dan Durasi Penyinaran, yang masing-masing memiliki loading sangat tinggi sebesar 0.9599. Nilai ini menegaskan bahwa PC1 hampir sepenuhnya menggambarkan intensitas penyinaran matahari dalam data klimatologi. Variabel lain seperti Kelembapan (-0.6180) dan Curah Hujan (-0.2546) hanya memberikan kontribusi tambahan yang jauh lebih kecil, sehingga perannya terhadap PC1 dapat dianggap minor.

Pada komponen utama kedua (PC2), variabel dengan pengaruh paling besar adalah Tekanan Udara (loading 0.9201) dan Kecepatan Angin (loading 0.8239). Kedua variabel ini menunjukkan hubungan kuat dalam pembentukan PC2, sehingga komponen utama kedua ini dapat diinterpretasikan sebagai dimensi yang merepresentasikan dinamika tekanan atmosfer dan intensitas angin. Variabel lain

seperti Kelembapan (0.3765) dan Arah Angin (0.1915) hanya memberikan kontribusi kecil.

Komponen utama ketiga (PC3) memperlihatkan loading tertinggi pada Arah Angin sebesar 0.8863, yang menunjukkan bahwa PC3 terutama menggambarkan variasi arah angin. Variabel lain seperti Curah Hujan (0.3880) dan Penyinaran (0.0452) berkontribusi jauh lebih kecil. Dengan dominasi yang jelas pada variabel arah angin, PC3 menjadi dimensi yang merepresentasikan pola dinamika arah angin secara spesifik.

Komponen utama keempat (PC4) didominasi oleh Curah Hujan dengan nilai loading tinggi sebesar 0.8535. Hal ini menunjukkan bahwa PC4 terutama menggambarkan variasi presipitasi, terpisah dari komponen sebelumnya yang menggambarkan radiasi, tekanan, atau angin. Variabel lain seperti Kelembapan (0.2803) dan Penyinaran (0.1917) memberikan kontribusi tambahan yang kecil.

Pada komponen utama kelima (PC5), variabel dengan kontribusi terbesar adalah Kelembapan (0.5201), disusul oleh Arah Angin (0.2185) dan Penyinaran (0.2062). Meskipun loading-nya tidak sebesar pada komponen sebelumnya, PC5 dapat diinterpretasikan sebagai dimensi minor yang menggambarkan variasi kelembapan.

Komponen utama keenam (PC6) memiliki kontribusi terbesar pada Tekanan Udara (0.3838) dan Kecepatan Angin (-0.3099), namun nilai loadingnya tetap relatif kecil dibanding komponen utama lain. Oleh karena itu, PC6 dapat dianggap sebagai komponen yang menangkap variasi minor terkait dinamika tekanan dan angin.

Sementara itu, komponen ketujuh (PC7) menunjukkan nilai loading mendekati nol pada semua variabel (sekitar 0.0000–0.0024), sehingga komponen ini tidak memberikan informasi bermakna terhadap struktur data dan dapat diabaikan dalam interpretasi.

Tabel 4.4 Hasil Komponen Utama

Komponen	Eigenvalue	Variance Ratio	Faktor Loading (Urutan dari terbesar → terkecil)
PC1	2.384272	0.338932	1. Penyinaran Matahari 2. Durasi Penyinaran 3. Kelembapan 4. Kecepatan Angin 5. Curah Hujan 6. Arah Angin 7. Tekanan
PC2	1.713957	0.243645	1. Tekanan 2. Kecepatan Angin 3. Kelembapan 4. Arah Angin 5. Curah Hujan 6. Penyinaran Matahari 7. Durasi Penyinaran
PC3	1.036050	0.147278	1. Arah Angin 2. Curah Hujan 3. Penyinaran Matahari 4. Durasi Penyinaran 5. Kelembapan 6. Kecepatan Angin 7. Tekanan
PC4	0.992376	0.141070	1. Curah Hujan 2. Kelembapan 3. Penyinaran Matahari 4. Durasi Penyinaran 5. Arah Angin 6. Kecepatan Angin 7. Tekanan

PC5	0.607374	0.086340	1. Kelembapan 2. Penyinaran Matahari 3. Durasi Penyinaran 4. Arah Angin 5. Tekanan 6. Curah Hujan 7. Kecepatan Angin
PC6	0.300614	0.042733	1. Tekanan 2. Kecepatan Angin 3. Kelembapan 4. Arah Angin 5. Penyinaran Matahari 6. Durasi Penyinaran 7. Curah Hujan
PC7	0.000011	0.000002	1. Penyinaran Matahari 2. Durasi Penyinaran variabel lainnya $\approx 0$ (tidak signifikan)

Berdasarkan hasil analisis Principal Component Analysis (PCA), komponen utama pertama hingga ketujuh memiliki nilai eigen, rasio varians, dan faktor loading yang berbeda-beda, yang secara bersama-sama menggambarkan struktur variasi data klimatologi pada penelitian ini.

- Komponen utama pertama (PC1) memiliki nilai eigen sebesar 2.384272 dengan rasio varians 0.338932 atau sekitar 33,89% dari total variasi data. Berdasarkan faktor loading, PC1 didominasi oleh variabel Penyinaran Matahari (0.9599) dan Durasi Penyinaran (0.9599) sebagai variabel dengan kontribusi terbesar, diikuti oleh Kelembapan (-0.6180), Kecepatan Angin (0.2576), Curah Hujan (-0.2546), serta kontribusi kecil dari Arah Angin (-0.1678) dan

Tekanan Udara (0.0145). Dominasi variabel penyinaran menunjukkan bahwa PC1 dapat diinterpretasikan sebagai komponen yang merepresentasikan dimensi radiasi matahari dan interaksinya dengan faktor atmosfer seperti kelembapan dan angin.

- b. Komponen utama kedua (PC2) memiliki nilai eigen 1.713957 dengan rasio varians 0.243645 (24,36%). PC2 didominasi oleh Tekanan Udara (0.9201) dan Kecepatan Angin (0.8239), dengan kontribusi tambahan dari Kelembapan (0.3765) dan Arah Angin (0.1915). Variabel penyinaran dan curah hujan memiliki loading yang jauh lebih kecil. Oleh karena itu, PC2 dapat diinterpretasikan sebagai komponen yang menggambarkan dinamika tekanan atmosfer dan intensitas angin, yang menjadi faktor meteorologis penting dalam variasi kondisi atmosfer.
- c. Komponen utama ketiga (PC3) memiliki nilai eigen 1.036050 dan rasio varians 0.147278 (14,73%). Variabel dengan loading terbesar adalah Arah Angin (0.8863), diikuti oleh Curah Hujan (0.3880). Komponen penyinaran, tekanan, dan kelembapan memberikan kontribusi relatif kecil. Dengan demikian, PC3 menggambarkan variasi yang terutama dipengaruhi oleh perubahan arah angin, dengan tambahan efek dari curah hujan.
- d. Komponen utama keempat (PC4) memiliki nilai eigen 0.992376 dengan rasio varians 0.141070 (14,11%). Loading terbesar pada PC4 terdapat pada Curah Hujan (0.8535), diikuti oleh Kelembapan

(0.2803) dan Penyinaran sekitar 0.191. Variabel lainnya memberikan kontribusi kecil. PC4 dapat diinterpretasikan sebagai komponen yang menggambarkan intensitas presipitasi, di mana variasi curah hujan menjadi faktor paling dominan pada komponen ini.

- e. Komponen utama kelima (PC5) memiliki nilai eigen 0.607374 dengan rasio varians 0.086340 (8,63%). Variabel dominan pada PC5 adalah Kelembapan (0.5201) dan Arah Angin (0.2185), disertai kontribusi sedang dari variabel penyinaran ( $\pm 0.206$ ). Komponen ini menggambarkan variasi tambahan pada kelembapan atmosfer dan pola interaksi minor dengan arah angin serta radiasi matahari.
- f. Komponen utama keenam (PC6) memiliki nilai eigen 0.300614 dengan rasio varians 0.042733 (4,27%). Loading terbesar pada PC6 berasal dari Tekanan Udara (0.3838), diikuti oleh Kecepatan Angin (-0.3099) dan Kelembapan (-0.2109). PC6 menggambarkan variasi minor dalam tekanan dan kondisi atmosfer, namun kontribusinya terhadap total variasi data relatif kecil.
- g. Komponen utama ketujuh (PC7) memiliki nilai eigen yang sangat kecil yaitu 0.000011 dengan rasio varians 0.000002, sehingga kontribusinya terhadap variasi data dapat diabaikan. Semua variabel memiliki faktor loading mendekati nol pada PC7, menandakan bahwa komponen ini tidak relevan dalam struktur utama data.

Secara keseluruhan, hasil PCA menunjukkan bahwa informasi utama dalam dataset paling banyak dijelaskan oleh dua komponen pertama (PC1–PC2), dengan PC1 dan PC2 sebagai komponen paling Komponen lainnya hanya menyumbang variasi kecil dan lebih bersifat pendukung.

#### **4.4. Visualisasi Grafik**

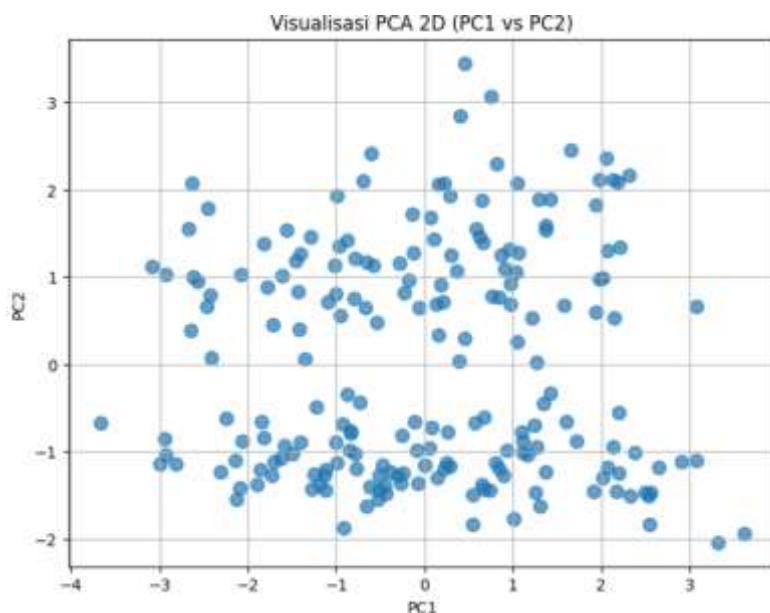
Visualisasi grafik merupakan tahapan penting dalam analisis Principal Component Analysis (PCA) karena berfungsi untuk menggambarkan struktur penyebaran data setelah proses reduksi dimensi dilakukan. Melalui visualisasi, pola hubungan antar data dapat diamati secara lebih jelas, termasuk tingkat variasi, kecenderungan pengelompokan, serta keberadaan data yang memiliki karakteristik berbeda atau ekstrem. Visualisasi ini juga membantu dalam mengevaluasi sejauh mana komponen utama mampu merepresentasikan variasi data klimatologi secara efektif.

Pada penelitian ini, visualisasi dilakukan dengan memproyeksikan data klimatologi ke dalam ruang komponen utama hasil PCA. Representasi grafis ini digunakan sebagai dasar awal untuk mengidentifikasi pola sebaran data dan potensi pembentukan klaster sebelum dilakukan proses clustering secara lebih formal.

##### **4.4.1 Visualisasi 2 dimensi**

Visualisasi dua dimensi menggunakan komponen utama pertama (PC1) dan komponen utama kedua (PC2) dilakukan untuk menggambarkan struktur penyebaran data klimatologi setelah proses reduksi dimensi dengan PCA. Representasi ini memungkinkan identifikasi pola dan hubungan antarsampel secara

lebih jelas dibandingkan visualisasi pada ruang variabel asli yang memiliki dimensi tinggi. Dengan memproyeksikan data ke dalam dua komponen utama yang memiliki kontribusi informasi terbesar, grafik ini memberikan gambaran awal mengenai variasi utama dalam dataset serta potensi pembentukan kelompok atau klaster berdasarkan kemiripan karakteristik klimatologi.



Gambar 4.1 Grafik 2 dimensi hasil PCA

Gambar 4.1 menunjukkan sebaran titik pada grafik memperlihatkan bagaimana data klimatologi tersebar setelah direduksi dimensinya dari variabel asli ke dua komponen utama yang memiliki kontribusi informasi terbesar. Terlihat bahwa titik-titik data tidak membentuk satu kelompok melainkan menyebar ke beberapa area dengan pola distribusi yang berbeda. Hal ini mengindikasikan adanya variasi yang cukup kuat antar sampel, terutama terkait karakteristik curah hujan, kelembapan, penyinaran matahari, dan parameter klimatologi lainnya.

Grafik dua dimensi ini menampilkan proyeksi setiap data klimatologi pada bidang PC1–PC2, di mana sumbu horizontal merepresentasikan PC1 sebagai komponen dengan kontribusi variansi terbesar, sedangkan sumbu vertikal merepresentasikan PC2 sebagai komponen dengan kontribusi variansi kedua terbesar. Penyebaran titik pada grafik menunjukkan bagaimana data terdistribusi setelah direduksi dari ruang variabel asli yang berdimensi tinggi ke dalam dua komponen utama.

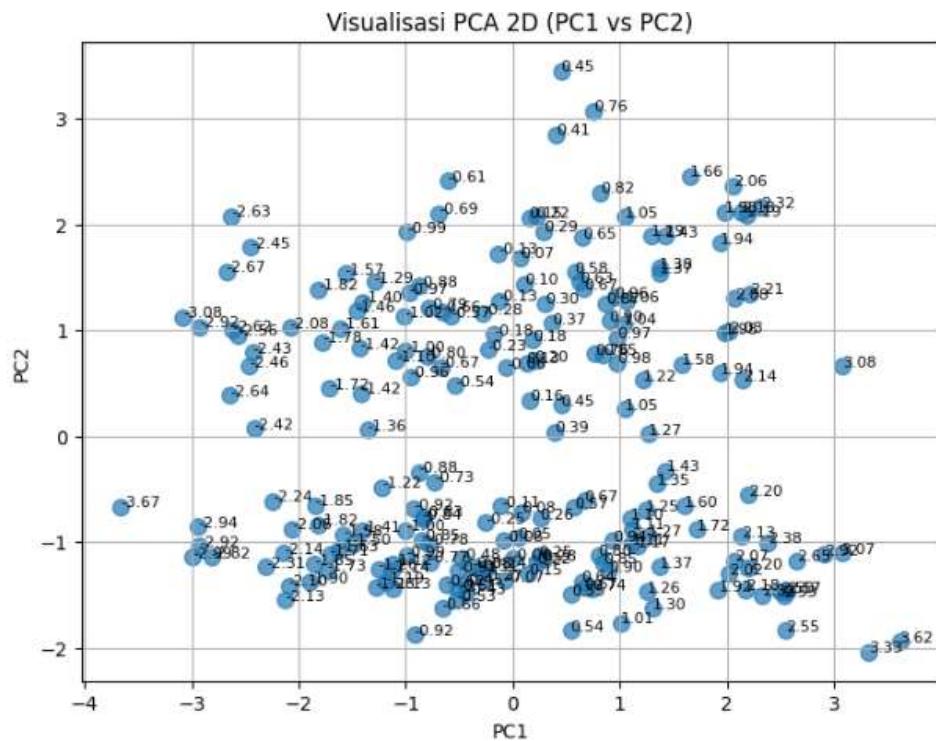
Berdasarkan hasil visualisasi, terlihat bahwa sebaran data cenderung lebih dominan pada arah PC1, yang menunjukkan bahwa variasi utama antar data klimatologi paling banyak dijelaskan oleh komponen ini. Titik-titik data membentuk pola penyebaran yang tidak sepenuhnya homogen, melainkan menunjukkan adanya area dengan kepadatan tinggi dan area dengan kepadatan rendah. Kondisi ini mengindikasikan adanya perbedaan karakteristik klimatologi antar kelompok data.

Secara keseluruhan, visualisasi dua dimensi ini memberikan gambaran awal bahwa data memiliki kecenderungan membentuk beberapa kelompok alami berdasarkan perbedaan nilai PC1. Temuan ini mendukung penggunaan metode clustering pada tahap selanjutnya, karena struktur penyebaran data menunjukkan pemisahan yang cukup jelas antar kelompok, sehingga analisis pengelompokan dapat dilakukan dengan lebih efektif dan interpretatif.

#### **4.4.2 Pengelompokan hasil visualisasi dengan cluster terdekat**

Pada tahap ini, hasil visualisasi dua dimensi dari komponen utama yang diperoleh melalui PCA dianalisis lebih lanjut untuk membentuk kelompok data

berdasarkan kedekatan posisi setiap titik pada grafik. Pendekatan ini bertujuan untuk mengidentifikasi pola pengelompokan alami yang muncul dari sebaran data, sehingga setiap titik dapat ditempatkan ke dalam klaster yang paling sesuai dengan jarak terdekatnya. Melalui proses ini, struktur hubungan antar data menjadi lebih jelas dan memudahkan interpretasi karakteristik masing-masing kelompok.



Gambar 4.2 Persebaran data grafik 2 dimensi

Gambar 4.2 menampilkan persebaran data klimatologi hasil proyeksi ke dalam dua komponen utama, yaitu PC1 dan PC2. Secara visual terlihat bahwa penyebaran titik lebih didominasi oleh arah horizontal (PC1), yang memiliki rentang nilai jauh lebih luas dibandingkan PC2. Hal ini sejalan dengan hasil PCA yang menunjukkan bahwa PC1 memiliki kontribusi variansi terbesar, sehingga perbedaan utama antar sampel klimatologi paling banyak ditentukan oleh komponen ini. Sebagian besar titik terkonsentrasi pada rentang PC1 antara -3

hingga 0 dengan nilai PC2 yang berkisar antara  $-1$  hingga 1, menggambarkan bahwa mayoritas data memiliki karakteristik yang relatif seragam setelah direduksi. Sementara itu, titik-titik dengan nilai PC1 positif tampak lebih tersebar dan jumlahnya lebih sedikit, khususnya pada rentang nilai 2 hingga 3.5, yang menunjukkan keberadaan sampel dengan karakteristik lebih ekstrem atau berbeda dibandingkan kelompok utama. Pada sisi vertikal, penyebaran PC2 memiliki rentang yang lebih sempit, sehingga variasi yang ditangkap oleh komponen kedua tidak sebesar PC1, meskipun tetap berperan dalam membedakan karakteristik minor seperti tekanan dan angin. Secara keseluruhan, pola penyebaran titik pada grafik menunjukkan kecenderungan pembentukan kelompok yang alami, di mana data dengan PC1 negatif membentuk kelompok padat, data pada rentang PC1 0–2 berada pada area transisi, dan data dengan PC1 lebih dari 2 membentuk kelompok kecil yang terpisah jelas. Struktur ini mendukung proses clustering pada tahap selanjutnya, karena distribusi titik terlihat memiliki batas visual yang jelas sesuai dengan perbedaan nilai pada komponen utama.

Dengan demikian, visualisasi ini menjadi dasar kuat bagi proses clustering pada tahap selanjutnya, karena struktur penyebaran titik sudah menunjukkan kecenderungan alami pembentukan beberapa kelompok yang berbeda secara jelas baik kelompok pusat, kelompok menengah, maupun kelompok ekstrem. Pola ini menunjukkan bahwa PCA berhasil mereduksi data dengan tetap mempertahankan struktur variasi penting dalam dataset klimatologi.

## 4.5 Model Jumlah Cluster

Penentuan jumlah cluster merupakan tahap penting dalam proses pengelompokan data karena berpengaruh langsung terhadap kualitas dan interpretasi hasil clustering. Jumlah cluster yang terlalu sedikit berpotensi menggabungkan data dengan karakteristik yang berbeda, sedangkan jumlah cluster yang terlalu banyak dapat menyebabkan pemisahan yang berlebihan sehingga struktur alami data menjadi kurang jelas. Oleh karena itu, penentuan jumlah cluster dalam penelitian ini dilakukan dengan mempertimbangkan pola sebaran data hasil PCA serta acuan dari penelitian sebelumnya.

Pada penelitian ini, jumlah cluster ditentukan berdasarkan pola distribusi nilai komponen utama pertama (PC1) yang menunjukkan pemisahan alami (natural breaks) antar kelompok data. Selain itu, pemilihan jumlah cluster juga mengacu pada penelitian sebelumnya yang menyatakan bahwa pembagian data klimatologi ke dalam tiga kelompok utama sudah mampu merepresentasikan variasi kondisi hujan secara umum, yaitu kondisi rendah, sedang, dan tinggi. Pendekatan ini diharapkan dapat menghasilkan struktur cluster yang sederhana namun tetap mencerminkan karakteristik utama data klimatologi.

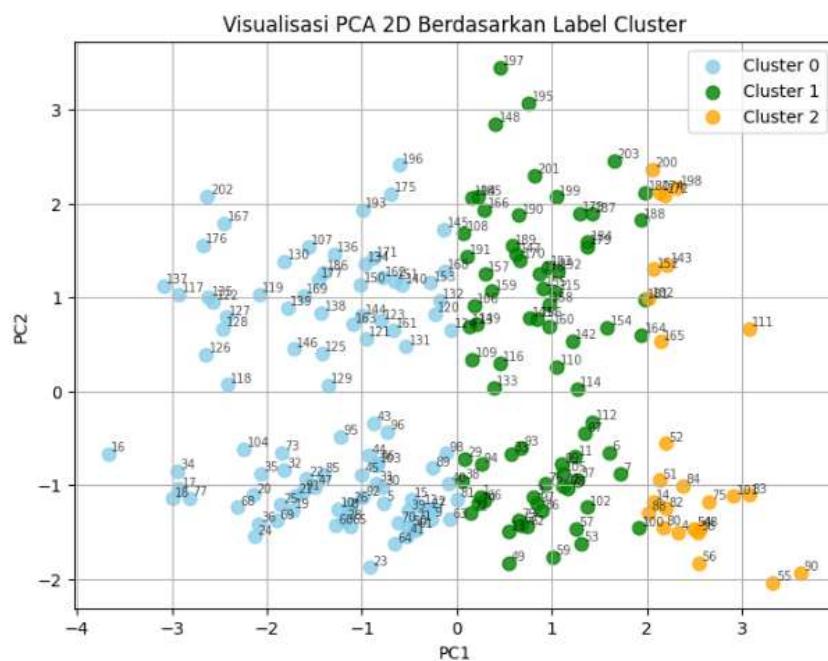
### 4.5.1 Model 3 Cluster

Penentuan jumlah klaster merupakan tahap penting sebelum proses pengelompokan dilakukan. Jumlah klaster ditentukan dengan rentang nilai berdasarkan penelitian yang dilakukan oleh almais (2023) bahwa cluster rentang cluster ada 3

Tabel 4.5. Rentang nilai clustering

Warna cluster	Rentang nilai	Jenis cluster
Biru	$n < 0$	Cluster 0
Hijau	$n < n \leq 2$	Cluster 1
Orange	$n > 2$	Cluster 2

Pada tahap ini, analisis dilakukan untuk mengidentifikasi jumlah kelompok yang mampu mewakili pola variasi data secara efektif tanpa mengurangi informasi penting yang terkandung di dalamnya. Berdasarkan hasil pengamatan visual terhadap struktur penyebaran data pada PCA,



Gambar 4.3 Visualisasi Cluster PCA

Gambar 4.3 memperlihatkan visualisasi persebaran data pada bidang PCA dua dimensi (PC1–PC2). Grafik tersebut tampak bahwa kelompok dengan nilai PC1

$< 0$  membentuk satu klaster besar yang terpisah jelas pada sisi kiri grafik dan ditandai dengan warna biru (Cluster 0). Sebaran titik pada area ini sangat padat, menunjukkan bahwa sebagian besar data memiliki karakteristik yang serupa dalam dimensi utama PCA.

#### 4.5.2 Model 4 Cluster

Selanjutnya, kelompok dengan nilai  $0 \leq PC1 < 2$  membentuk klaster kedua yang ditandai dengan warna hijau (Cluster 1). Wilayah ini merupakan zona transisi yang memisahkan cluster ekstrem kiri ( $PC1 < 0$ ) dan cluster kanan ( $PC1 \geq 2$ ). Titik-titik pada rentang ini membentuk kelompok yang cukup jelas dan berada di bagian tengah grafik. Hal ini menunjukkan bahwa data dalam rentang ini memiliki karakteristik menengah yang berbeda dari dua klaster lainnya.

Pada sisi paling kanan grafik terlihat klaster ketiga, yaitu kelompok dengan nilai  $PC1 \geq 2$  yang ditandai dengan warna oranye (Cluster 2). Titik-titik pada klaster ini tersebar cukup teratur dan posisinya jelas terpisah dari dua klaster lain, menunjukkan bahwa data pada rentang nilai PC1 tersebut memiliki pola yang berbeda atau ekstrem dibandingkan kelompok lainnya.

Secara keseluruhan, pola persebaran data pada grafik PCA menunjukkan pemisahan klaster yang sangat jelas berdasarkan nilai PC1 sebagai komponen utama dengan kontribusi varians tertinggi. Ketiga klaster tersebut mencerminkan tiga kondisi dominan pada data, mulai dari kategori ringan ( $PC1 < 0$ ), kategori sedang ( $0 \leq PC1 < 2$ ), hingga kategori intensitas tinggi atau ekstrem ( $PC1 \geq 2$ ).

Selain melakukan pengelompokan utama ke dalam tiga klaster, penelitian ini juga menambahkan analisis pembagian menjadi empat dan lima klaster untuk melihat struktur penyebaran data secara lebih rinci. Pembentukan jumlah klaster tambahan ini tidak dilakukan secara acak, tetapi didasarkan pada pola alami yang muncul dari distribusi nilai komponen utama pertama (PC1). Ketika seluruh nilai PC1 diurutkan, terlihat beberapa *natural breaks* atau lonjakan signifikan antar titik yang menandakan adanya batas kelompok yang jelas.

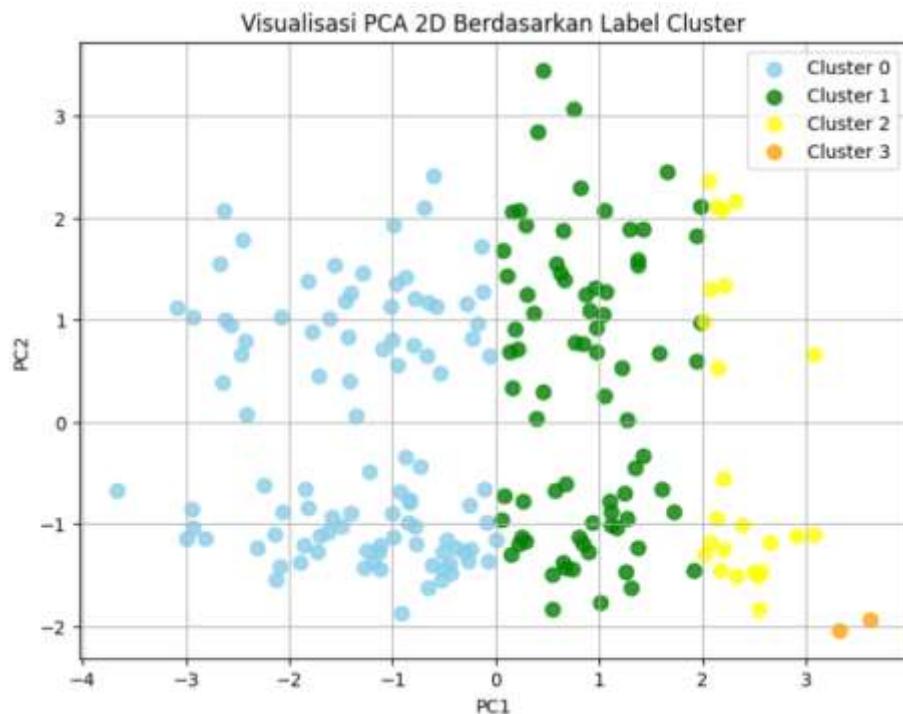
Dalam penentuan rentang nilai untuk 4 cluster, batas interval ditentukan berdasarkan pola penyebaran nilai PC1 yang menunjukkan adanya pemisahan alami (*natural breaks*). Ketika nilai PC1 diurutkan, terlihat empat kelompok yang terbentuk secara jelas: nilai  $PC1 < 0$  yang terkumpul padat di sisi kiri grafik, rentang 0–2 yang menjadi wilayah transisi, rentang 2–3,3 yang mulai membentuk kelompok baru yang lebih homogen, serta nilai di atas 3,3 yang terpisah jauh dari distribusi utama. Pola pemisahan ini menjadi dasar penetapan rentang setiap cluster, karena mencerminkan perubahan karakteristik data.

Tabel 4.5 Rentang nilai 4 cluster

Warna cluster	Rentang nilai	Jenis cluster
Biru	$n < 0$	Cluster 0
Hijau	$n < n \leq 2$	Cluster 1
Kuning	$2 < n \leq 3,3$	Cluster 2
Orange	$N > 3,3$	Cluster 3

Pada tahap ini, analisis dilakukan untuk mengidentifikasi jumlah kelompok yang mampu mewakili pola variasi data secara efektif tanpa mengurangi informasi

penting yang terkandung di dalamnya. Berdasarkan hasil pengamatan visual terhadap struktur penyebaran data pada PCA,



Gambar 4.5 Visualisasi Grafik 2D 4 Cluster

Pada Gambar 4.5 batas tambahan muncul pada nilai PC1 sekitar 3,3. Hal ini terlihat dari adanya lonjakan yang cukup besar antara titik dengan PC1 sekitar 3.07 dan 3.61, yang menunjukkan bahwa data pada nilai tersebut memiliki karakteristik berbeda dari kelompok PC1 2 – 3,3. Dengan pola ini, pembentukan 4 cluster menjadi lebih jelas, yaitu data dengan  $PC1 < 0$  membentuk kelompok besar yang homogen (Cluster 0), nilai 0 hingga kurang dari 2 sebagai kelompok transisi (Cluster 1), nilai 2 hingga kurang dari 3,3 sebagai kelompok dengan intensitas tinggi namun masih serupa (Cluster 2), dan nilai  $PC1 \geq 3,3$  sebagai kelompok yang benar-benar ekstrem dan terpisah dari lainnya (Cluster 3).

#### 4.5.3 Model 5 Cluster

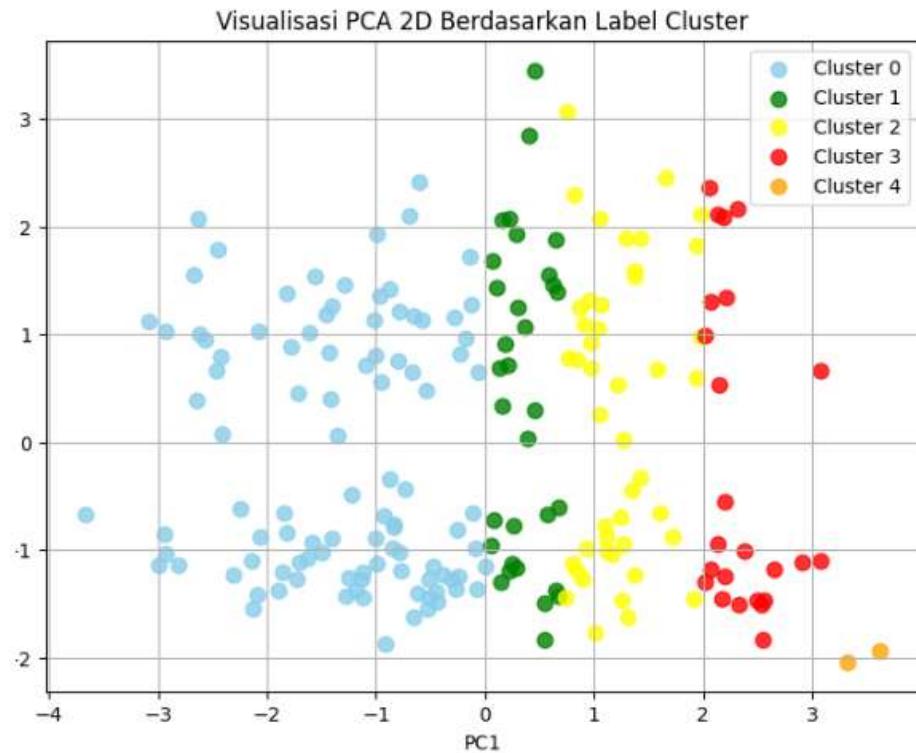
Dalam penentuan 5 cluster, pola pembentukan kelompok dianalisis berdasarkan perubahan nilai PC1 yang menunjukkan perbedaan karakteristik antar data. Cluster pertama ditetapkan pada  $PC1 < 0$  karena nilai negatif cenderung membentuk satu kelompok besar yang homogen. Lonjakan pertama muncul pada rentang 0 hingga 0.7, menunjukkan adanya pemisahan alami antara data bernilai negatif dan kelompok transisi awal. Rentang berikutnya, 0.7 hingga 2, membentuk kelompok kedua yang lebih stabil dengan distribusi nilai yang lebih rapat. Lonjakan berikutnya terjadi pada nilai sekitar 2, sehingga rentang 2 hingga 3.3 menjadi cluster tersendiri yang berbeda dari data sebelumnya. Lonjakan terbesar terlihat antara PC1 3.07 dan 3.61, yang memisahkan data bernilai sangat tinggi ke dalam cluster kelima. Dengan demikian, pembagian menjadi lima cluster mencerminkan struktur alami data berdasarkan variasi dan jarak antar nilai PC1.

Tabel 4.6 Rentang nilai 5 cluster

Warna cluster	Rentang nilai	Jenis cluster
Biru	$n < 0$	Cluster 0
Hijau	$0 < n \leq 0,7$	Cluster 1
Kuning	$0,7 < n \leq 2$	Cluster 2
Red	$2 < n \leq 3,3$	Cluster 3
Orange	$n > 3,3$	Cluster 4

Pada tahap ini, analisis dilakukan untuk mengidentifikasi jumlah kelompok yang mampu mewakili pola variasi data secara efektif tanpa mengurangi informasi

penting yang terkandung di dalamnya. Berdasarkan hasil pengamatan visual terhadap struktur penyebaran data pada PCA,



Gambar 4.6 Visualisasi 2D 5 Cluster

Pada Gambar 4.6 Visualisasi PCA pada grafik 2D (PC1 vs PC2) juga mendukung pembagian tersebut. Titik-titik dengan PC1 negatif tampak jelas membentuk satu kelompok besar di sisi kiri grafik, sedangkan kelompok dengan nilai PC1 sangat tinggi ( $\geq 3.3$ ) terpisah jauh dan berkumpul pada area paling kanan grafik. Kelompok tengah dengan rentang PC1 antara 0.7 hingga 2 terlihat membentuk area transisi, yang memvalidasi bahwa rentang tersebut memiliki karakteristik tersendiri. Pada model empat dan lima klaster, visualisasi tersebut memperlihatkan bagaimana titik-titik pada rentang nilai PC1 yang berbeda membentuk kelompok yang homogen, dengan jarak antar cluster yang semakin

jelas seiring bertambahnya jumlah pembagian. Dengan demikian, penentuan klaster tambahan ini merupakan hasil dari analisis pola sebaran PC1 dan bukan keputusan subjektif, melainkan mencerminkan struktur data yang alami sesuai dengan karakteristik klimatologi pada dataset ini.

#### 4.5 Evaluasi Model

Evaluasi kualitas pengelompokan dalam penelitian ini dilakukan menggunakan metrik *Silhouette Score*, yaitu ukuran yang digunakan untuk menilai seberapa baik suatu objek berada dalam klaster yang tepat. Metrik ini mempertimbangkan dua aspek utama, yaitu tingkat kekompakan data dalam satu klaster (*cohesion*) dan tingkat keterpisahan antar klaster (*separation*). Nilai *Silhouette Score* berada pada rentang  $-1$  hingga  $1$ , di mana nilai mendekati  $1$  menunjukkan bahwa data sangat sesuai dengan klasternya, sedangkan nilai mendekati  $0$  menandakan adanya tumpang tindih antar klaster. Sementara itu, nilai yang negatif menunjukkan bahwa sebagian data mungkin salah tempat karena lebih dekat dengan klaster lain dibandingkan dengan klaster asalnya.

Tabel 4.7 Hasil Uji validasi

Jumlah Cluster	Siluetto Score
3 Cluster	0,5193
4 Cluster	0,4964
5 Cluster	0.4015

Berdasarkan hasil perhitungan silhouette score, pemilihan jumlah cluster yang paling optimal terlihat pada jumlah 3 cluster dengan nilai *silhouette score*

0,5193, yang merupakan nilai tertinggi dibandingkan dengan pembagian 4 maupun 5 cluster. Nilai silhouette yang lebih besar menunjukkan bahwa titik-titik data dalam setiap cluster memiliki kemiripan internal yang kuat dan jarak yang jelas terhadap cluster lain. Ini mengindikasikan bahwa pembagian menjadi 3 cluster membentuk kelompok yang paling kompak dan terpisah dengan baik dibandingkan konfigurasi lainnya.

Selain itu, jumlah 3 cluster juga sesuai dengan kategori label BMKG, yang hanya membagi kondisi hujan ke dalam tiga kelas, yaitu hujan ringan, hujan sedang, dan hujan lebat. Dengan demikian, pembentukan 3 cluster secara alami mencerminkan struktur label referensi yang digunakan dalam data acuan. Ketika dilihat dari pola nilai PC1, cluster pertama ( $PC1 < 0$ ) berisi kelompok nilai rendah, cluster kedua ( $0 \leq PC1 < 2$ ) memiliki nilai sedang, sementara cluster ketiga ( $PC1 \geq 2$ ) memiliki nilai yang jauh lebih tinggi, sehingga menyatu dengan karakteristik kategori hujan yang semakin intens.

Jika dibandingkan dengan 4 atau 5 cluster, meskipun penambahan jumlah cluster menghasilkan pemisahan lebih detail, nilai silhouette score justru menurun menjadi 0,4964 untuk 4 cluster dan 0,4015 untuk 5 cluster. Penurunan ini menunjukkan bahwa semakin banyak cluster menyebabkan pembagian menjadi kurang stabil dan jarak antar cluster semakin tidak jelas. Artinya, pemisahan cluster tambahan tidak mencerminkan struktur alami data dan tidak meningkatkan kualitas pemodelan.

#### 4.6 Integrasi Islam

Pola yang mencerminkan adanya sistem yang tersusun dan teratur dalam proses penciptaan, sebagaimana ditegaskan dalam firman Allah Swt. dalam Surah Al-A‘lā ayat 3:

﴿ وَالَّذِي قَدَرَ فَهَدَى ۚ ۳ ﴾

“yang menentukan kadar (masing-masing) dan memberi petunjuk”(Q.S An Al-A‘lā:3).

Dalam Tafsir Tahlili Ayat tersebut menunjukkan bahwa setiap ciptaan memiliki ukuran, ketentuan, dan arah yang jelas. Keteraturan pola klimatologi yang dihasilkan melalui menggambarkan bahwa fenomena alam berlangsung sesuai dengan ketetapan dan hukum yang telah ditetapkan oleh Allah Swt., Dengan memahami pola yang terbentuk dari analisis PCA dan clustering, manusia sebagai khalifah di bumi dapat menjalankan amanahnya untuk menjaga keseimbangan alam dan memanfaatkan ilmu pengetahuan dengan sebaik-baiknya. Hal ini sejalan dengan firman Allah dalam Surah Al-‘Alaq ayat 1–5, yang menekankan pentingnya membaca, meneliti, dan menggali ilmu sebagai jalan untuk mngenal tanda-tanda kebesaran-Nya:

﴿ إِنَّ رَبَّكَ الَّذِي خَلَقَ ۚ ۱ۚ خَلَقَ الْإِنْسَانَ مِنْ غَمَّٰ ۲ۚ إِنَّ رَبَّكَ الْأَكْرَمُ ۳ۚ الَّذِي عَلَمَ بِالْقَمَّ ۴ۚ عَلَمَ الْإِنْسَانَ مَا لَمْ يَعْلَمْ ۵ۚ ﴾

“(1)Bacalah dengan (menyebut) nama Tuhanmu yang menciptakan!(2)Dia menciptakan manusia dari segumpal darah.(3)Bacalah! Tuhanmulah Yang Mahamulia(4)yang mengajar (manusia) dengan pena.(5)Dia mengajarkan manusia apa yang tidak diketahuinya.”(Q.S Al-‘Alaq 1-5)

Ayat ini menekankan bahwa ilmu merupakan kunci utama untuk memahami fenomena alam. Dalam konteks modern, analisis ilmiah seperti PCA dan clustering memungkinkan manusia membaca tanda-tanda perubahan lingkungan secara lebih akurat, memprediksi potensi bencana seperti banjir, kekeringan, atau cuaca ekstrem, serta merancang strategi mitigasi yang lebih matang.

Selain itu, pemahaman terhadap karakteristik setiap klaster klimatologi dapat membantu sektor pertanian, perencanaan tata ruang, dan pengelolaan sumber daya air agar lebih adaptif terhadap perubahan iklim. Analisis data yang akurat juga mendukung lembaga seperti BMKG dalam menyediakan informasi dan peringatan dini, sehingga masyarakat dapat mengambil langkah antisipatif demi menjaga keselamatan dan kelestarian lingkungan.

Pesan ini sejalan dengan hadis Rasulullah SAW yang menegaskan bahwa manusia bertanggung jawab terhadap segala yang berada dalam kekuasaannya:

لُكْمَ زَاعِ وَكُلْمَ مَسْئُولٌ عَنْ رَعِيَّتِهِ الْإِمَامُ زَاعِ وَمَسْئُولٌ عَنْ رَعِيَّتِهِ وَالرَّجُلُ زَاعِ فِي أَهْلِهِ وَهُوَ مَسْئُولٌ عَنْ رَعِيَّتِهِ

*“Setiap kalian adalah pemimpin dan akan dimintai pertanggungjawaban atas kepemimpinannya. Seorang imam adalah pemimpin dan akan dimintai pertanggungjawabannya dan demikian juga seorang pria adalah seorang pemimpin bagi keluarganya dan akan dimintai pertanggungjawaban atas kepemimpinannya.” (HR. Bukhari: 2278).*

Hadis ini mengingatkan bahwa menjaga bumi, mengelola sumber daya alam dengan bijaksana, serta menggunakan ilmu untuk mencegah kerusakan adalah bagian dari amanah dan pertanggungjawaban manusia di hadapan Allah. Dengan

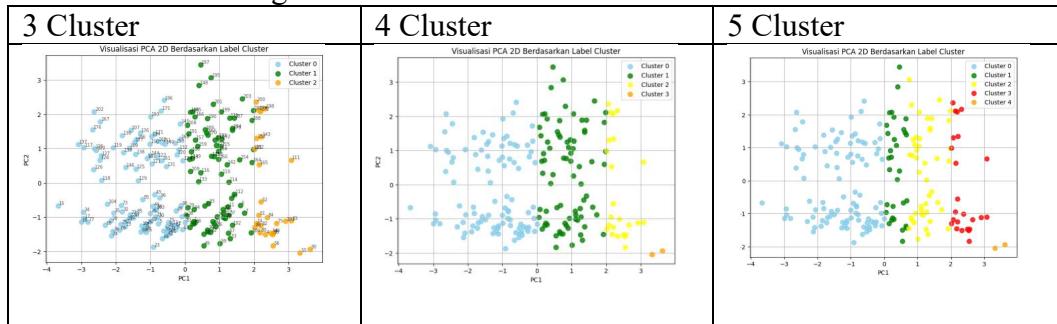
ilmu, manusia tidak hanya memahami ciptaan-Nya, tetapi juga menunaikan tugas mulia untuk memelihara bumi demi generasi sekarang dan yang akan datang.

#### 4.7 Hasil

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa penerapan Principal Component Analysis (PCA) berhasil mereduksi dimensi data klimatologi yang semula terdiri dari tujuh variabel menjadi sejumlah komponen utama yang lebih ringkas namun tetap mampu mempertahankan sebagian besar informasi penting. Hasil analisis menunjukkan bahwa dua komponen utama pertama (PC1 dan PC2) telah mampu menjelaskan lebih dari 58% variasi total data, sedangkan empat komponen pertama (PC1–PC4) secara kumulatif menjelaskan sekitar 82,98% variasi data. Hal ini menunjukkan bahwa PCA efektif dalam menyederhanakan struktur data tanpa menghilangkan karakteristik utama pola klimatologi.

Hasil reduksi dimensi menggunakan PCA kemudian dimanfaatkan sebagai dasar dalam proses clustering, dengan menggunakan skor komponen utama sebagai representasi data. Visualisasi PCA dua dimensi (PC1–PC2) menunjukkan bahwa data memiliki pola penyebaran yang tidak homogen dan memperlihatkan adanya pemisahan alami (natural breaks), terutama sepanjang sumbu PC1. Pola ini mengindikasikan bahwa data klimatologi secara alami cenderung membentuk beberapa kelompok berdasarkan perbedaan karakteristik utama yang direpresentasikan oleh komponen utama.

Tabel 4.8 Perbandingan 3visualisasi 3 Cluster



Pada Tabel 4.8 Menyajikan perbedaan pola cluster dalam tiap masing masing cluster kemudian dari pola cluster dilakukan pengujian beberapa model clustering, yaitu model 3 cluster, 4 cluster, dan 5 cluster. Evaluasi kualitas clustering menggunakan Silhouette Score menunjukkan bahwa model 3 cluster menghasilkan nilai tertinggi sebesar 0,5193, dibandingkan dengan model 4 cluster (0,4964) dan 5 cluster (0,4015). Nilai ini menunjukkan bahwa model 3 cluster memiliki tingkat kekompakkan internal yang lebih baik serta pemisahan antar cluster yang lebih jelas.

Keunggulan model 3 cluster juga diperkuat oleh hasil PCA, di mana pembagian cluster berdasarkan rentang nilai PC1 ( $< 0$ ,  $0-2$ , dan  $\geq 2$ ) sejalan dengan struktur variasi utama data. Setiap cluster merepresentasikan kondisi klimatologi yang berbeda secara jelas, mulai dari kondisi dengan intensitas rendah, kondisi menengah, hingga kondisi tinggi atau ekstrem. Selain itu, model 3 cluster juga selaras dengan klasifikasi curah hujan yang digunakan oleh BMKG, yaitu hujan ringan, hujan sedang, dan hujan lebat, sehingga memiliki relevansi klimatologis yang kuat.

Sementara itu, model 4 dan 5 cluster memberikan pembagian yang lebih rinci terhadap data, terutama dalam mengidentifikasi kelompok ekstrem. Namun,

penambahan jumlah cluster menyebabkan penurunan nilai Silhouette Score, yang menunjukkan bahwa batas antar cluster menjadi kurang tegas dan struktur alami data mulai terpecah secara berlebihan. Hal ini mengindikasikan bahwa meskipun model dengan jumlah cluster lebih banyak dapat digunakan untuk analisis eksploratif, model tersebut kurang optimal sebagai representasi utama pola data.

Secara keseluruhan, kombinasi PCA sebagai metode reduksi dimensi dan clustering berbasis skor komponen utama terbukti efektif dalam mengidentifikasi pola utama data klimatologi. PCA berperan penting dalam menyederhanakan data dan menonjolkan struktur variasi utama, sementara clustering memanfaatkan struktur tersebut untuk membentuk kelompok data yang bermakna. Berdasarkan hasil evaluasi dan interpretasi klimatologis, model 3 cluster ditetapkan sebagai model terbaik karena memberikan keseimbangan optimal antara kualitas statistik, kejelasan struktur data, dan kemudahan interpretasi.

## **BAB V**

### **KESIMPULAN**

#### **5.1 Kesimpulan**

Berdasarkan seluruh rangkaian analisis pada BAB IV, dapat disimpulkan bahwa proses pre-processing, reduksi dimensi dengan PCA, serta pengelompokan menggunakan clustering berhasil mengungkap pola utama dalam data klimatologi harian yang dianalisis. Tahap preprocessing yang meliputi penanganan missing value menggunakan litwise delete dan normalisasi dengan StandardScaler telah menghasilkan data yang bersih, terstandarisasi, serta siap dianalisis tanpa distorsi akibat perbedaan skala variabel.

Hasil PCA menunjukkan bahwa sebagian besar variasi data dapat dijelaskan oleh empat komponen utama pertama (PC1–PC2) dengan total kontribusi mencapai lebih dari 50%. Komponen PC1 didominasi oleh variabel penyinaran matahari, PC2 oleh tekanan dan kecepatan angin.

Visualisasi PCA 2 dimensi berhasil menggambarkan pola persebaran data yang menunjukkan kecenderungan pembentukan kelompok-kelompok alami berdasarkan nilai PC1. Analisis manual berdasarkan interval PC1 memperlihatkan bahwa pembagian 3, 4, dan 5 cluster semuanya dapat dibentuk berdasarkan *natural breaks* pada distribusi nilai PC1. Namun, hasil validasi menggunakan silhouette score menunjukkan bahwa pembagian 3 cluster merupakan yang paling optimal, dengan nilai silhouette 0,5193, lebih tinggi dibandingkan 4 cluster (0,4964) dan 5

cluster (0,4015). Hal ini menandakan bahwa struktur tiga klaster adalah yang paling kompak, paling terpisah, dan paling stabil secara matematis.

Selain itu, tiga cluster tersebut juga selaras dengan klasifikasi curah hujan menurut BMKG yang terdiri dari hujan ringan, sedang, dan lebat. Cluster 0 mencerminkan nilai PC1 yang rendah, cluster 1 merepresentasikan nilai sedang, sementara cluster 2 menunjukkan nilai PC1 tinggi yang menggambarkan kondisi yang lebih ekstrem. Keselarasan ini memperkuat validitas hasil clustering dalam konteks klimatologi.

Secara keseluruhan, hasil analisis menunjukkan bahwa metode PCA dan memberikan pemahaman komprehensif mengenai struktur data klimatologi dan mampu mengidentifikasi pola utama yang relevan. Informasi ini dapat digunakan untuk mendukung interpretasi kondisi cuaca, perencanaan mitigasi iklim, serta pengambilan keputusan berbasis data dalam pengelolaan lingkungan dan sumber daya alam.

## 5.2 Saran

Penelitian selanjutnya disarankan untuk menggunakan data klimatologi dengan rentang waktu pengamatan yang lebih panjang serta mencakup wilayah yang lebih luas. Dengan data yang lebih banyak dan beragam, pola curah hujan yang terbentuk diharapkan dapat digambarkan secara lebih jelas dan akurat. Selain itu, penelitian berikutnya juga dapat mencoba menggunakan metode pengelompokan lain, seperti K-Means, DBSCAN, maupun metode gabungan, sebagai pembanding terhadap metode yang digunakan dalam penelitian ini.

Perbandingan tersebut diharapkan dapat memberikan gambaran mengenai metode yang paling sesuai, sehingga hasil pengelompokan pola curah hujan yang diperoleh menjadi lebih baik dan mudah untuk diinterpretasikan.

## **DAFTAR PUSTAKA**

- Abror, N., Sri Rahayu, D., & Mustakim, M. (2023). Rancang Bangun Sistem Informasi Pemeliharaan Alat Kantor Berbasis Web Pada Bmkg Stasiun Meteorologi Sultan Syarif Kasim Ii Pekanbaru: Web-Based Office Maintenance System Design At Bmkg Meteorological Station Sultan Syarif Kasim Ii Pekanbaru. *Indonesian Journal of Informatic Research and Software Engineering (IJIRSE)*, 3(2), 103–110.  
<https://doi.org/10.57152/ijirse.v3i2.899>
- Aidoo, E. N., Appiah, S. K., Awashie, G. E., Boateng, A., & Darko, G. (2021). Geographically weighted principal component analysis for characterising the spatial heterogeneity and connectivity of soil heavy metals in Kumasi, Ghana. *Helijon*, 7(9), e08039.  
<https://doi.org/10.1016/j.helijon.2021.e08039>
- Akbar, J. M., Sabirin, M., Nugraha, G. S., Alamsyah, N., & Majapahit, J. (2022). *PENERAPAN METODE PRINCIPAL COMPONENT ANALYSIS (PCA) DAN K-NEAREST NEIGHBORS (KNN) UNTUK KLASIFIKASI DATA KANKER PARU-PARU*. 4(2).
- Almais, A. T. W., Susilo, A., Naba, A., Sarosa, M., Crysidian, C., Tazi, I., Hariyadi, M. A., Muslim, M. A., Basid, P. M. N. S. A., Arif, Y. M., Purwanto, M. S., Parwatinetyas, D., Supriyono, & Wicaksono, H. (2023). Principal Component Analysis-Based Data Clustering for Labeling of Level

- Damage Sector in Post-Natural Disasters. *IEEE Access*, 11, 74590–74601.  
<https://doi.org/10.1109/ACCESS.2023.3275852>
- Almais, A. T. W., Susilo, A., Naba, A., Sarosa, M., Juwono, A. M., Crysdiann, C., Muslim, M. A., & Wicaksono, H. (2024). Characterization of Structural Building Damage in Post-Disaster Using GLCM-PCA Analysis Integration. *IEEE Access*, 12, 146190–146201.  
<https://doi.org/10.1109/ACCESS.2024.3469637>
- Al-Qur'an. (2025). *An-Nūr: 43*. Quran NU Online. <https://quran.nu.or.id/an-nur/43>
- Arisandi, R., & Ruhiat, D. (2021). *Implementasi Ridge Regression untuk Mengatasi Gejala Multikolinearitas pada Pemodelan Curah Hujan Berbasis Data Time Series Klimatologi*.
- Arlandy, K. S., Faqih, A., & Rinaldi, A. R. (2025). Mengoptimalkan Kinerja Naïve Bayes Pada Ancaman Modern Dengan Menggunakan PCA Pada Data Intrusion Detection System (IDS). *Jurnal Ilmiah ILKOMINFO - Ilmu Komputer & Informatika*, 8(1), 25–37.  
<https://doi.org/10.47324/ilkominfo.v8i1.303>
- Benny Hartanto, Ningrum Astriawati, Supartini, & Damar Kuncoro Yekti. (2022). Pencarian dan Pemanfaatan Informasi Data Badan Meteorologi, Klimatologi, dan Geofisika (BMKG). *INSOLOGI: Jurnal Sains dan Teknologi*, 1(5), 553–564. <https://doi.org/10.55123/insologi.v1i5.906>
- Dewi, S., & Pakereng, M. A. I. (2023). IMPLEMENTASI PRINCIPAL COMPONENT ANALYSIS PADA K-MEANS UNTUK KLASTERISASI

TINGKAT PENDIDIKAN PENDUDUK KABUPATEN SEMARANG.

*JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika), 8(4),*

1186–1195. <https://doi.org/10.29100/jipi.v8i4.4101>

Gustoro, D., Sujono, J., & Karlina, K. (2022). Perbandingan Pola Distribusi Hujan Terukur dan Hujan Satelit PERSIANN di DAS Progo. *Jurnal Teknik Pengairan, 13(1)*, 23–35.

<https://doi.org/10.21776/ub.pengairan.2022.013.01.03>

Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.

Haqq, B. T. E., Antika, A., & Wulandari, S. P. (2021). *Analisis Faktor-faktor Volume Ekspor Hasil Perikanan Menurut Provinsi di Indonesia Tahun 2021 menggunakan Analisis Faktor. 3.*

Hayati, M., Hudori, M., & Sari, R. P. (2025a). *Sosialisasi dan Penerapan Prediksi Curah Hujan Tweedie-PCA untuk Mitigasi Bencana dan Pengelolaan Iklim di BMKG Lampung. 10.*

Hayati, M., Hudori, M., & Sari, R. P. (2025b). *Sosialisasi dan Penerapan Prediksi Curah Hujan Tweedie-PCA untuk Mitigasi Bencana dan Pengelolaan Iklim di BMKG Lampung. 10.*

Hediyati, D., & Suartana, I. M. (2021). Penerapan Principal Component Analysis (PCA) Untuk Reduksi Dimensi Pada Proses Clustering Data Produksi Pertanian Di Kabupaten Bojonegoro. *Journal of Information Engineering and Educational Technology, 5(2)*, 49–54.

<https://doi.org/10.26740/jieet.v5n2.p49-54>

- Hendrastuty, N. (2024). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa. *Jurnal Ilmiah Informatika dan Ilmu Komputer (JIMA-ILKOM)*, 3(1), 46–56.  
<https://doi.org/10.58602/jima-ilkom.v3i1.26>
- Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). *Data Clustering: A Review*. Prentice Hall.
- Jihan Bastini, F., Sudarman Supardi, & Mas'ud Sar. (2024). Kajian Pola Hujan Akibat Pengaruh Perubahan Iklim Pada Daerah Aliran Sungai Gilireng (Das Gilireng). *Jurnal TESLINK : Teknik Sipil Dan Lingkungan*, 6(2), 444–450. <https://doi.org/10.52005/teslink.v6i2.373>
- Jolliffe, I. T. (1986). *Principal Component Analysis*. Springer New York.  
<https://doi.org/10.1007/978-1-4757-1904-8>
- Khasanah, U., & Gardjito, E. (2024a). Analisis Variabilitas Curah Hujan Daerah Aliran Sungai (DAS) Kali Magok. *Jurnal Komposit*, 8(1), 39–44.  
<https://doi.org/10.32832/komposit.v8i1.14756>
- Khasanah, U., & Gardjito, E. (2024b). Analisis Variabilitas Curah Hujan Daerah Aliran Sungai (DAS) Kali Magok. *Jurnal Komposit*, 8(1), 39–44.  
<https://doi.org/10.32832/komposit.v8i1.14756>
- Liang, N., Tuo, Y., Deng, Y., & He, T. (2022). PCA-based SVM classification for simulated ice floes in front of sluice gates. *Polar Science*, 34, 100839.  
<https://doi.org/10.1016/j.polar.2022.100839>
- M. Audi Ikhsan, & Samsudin, S. (2024). Sistem Informasi Iklim dan Kualitas Udara pada Laboratorium Kalibrasi BMKG Berbasis Website. *SATESI*:

*Jurnal Sains Teknologi dan Sistem Informasi, 4(2).*

<https://doi.org/10.54259/satesi.v4i2.3299>

Mubarok, M. R., Harjanto, A. T. J., & Renaldy, R. (2025). PENINGKATAN PERFORMA DBSCAN DENGAN REDUKSI DIMENSI PRINCIPAL COMPONENT ANALYSIS (PCA) DALAM KLASTERISASI TINGKAT KEMISKINAN DI INDONESIA. *Jurnal Informatika Teknologi Dan Sains (Jinteks)*, 7(3), 1176–1184. <https://doi.org/10.51401/jinteks.v7i3.6129>

Musfiroh, M., Novitasari, D. C. R., Intan, P. K., & Wisnawa, G. G. (2023). Penerapan Metode Principal Component Analysis (PCA) dan Long Short-Term Memory (LSTM) dalam Memprediksi Prediksi Curah Hujan Harian. *Building of Informatics, Technology and Science (BITS)*, 5(1).

<https://doi.org/10.47065/bits.v5i1.3114>

Nuraeni, F., Kurniadi, D., & Diazki, M. H. (2024). Algoritma K-Nearest Neighbor pada Kasus Dataset Imbalanced untuk Klasifikasi Kinerja Karyawan Perusahaan. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 11(3), 557–568. <https://doi.org/10.25126/jtiik.938144>

Parding, K. M., Benestad, R. E., Dyrrdal, A. V., & Lutz, J. (2023). A principal-component-based strategy for regionalisation of precipitation intensity–duration–frequency (IDF) statistics. *Hydrology and Earth System Sciences*, 27(20), 3719–3732. <https://doi.org/10.5194/hess-27-3719-2023>

Prasetyo, Y. D., Hidayati, A. F., Maulida, E., & Silalahi, F. Y. (2024). Implementasi Algoritma K-Means Clustering dalam Penentuan Gangguan Tidur Seseorang berdasarkan Gaya Hidup. *Journal of Information*

- Engineering and Educational Technology, 8(1), 54–61.*  
<https://doi.org/10.26740/jieet.v8n1.p54-61>
- Pratama, J. P., & Darmawan, Y. (2025). *Analisis Pola Sebaran Spasial Curah Hujan di Provinsi Jawa Tengah Menggunakan Metode Poligon Thiessen untuk Mitigasi Bencana Banjir.* 22(1).
- Ritonga, A. S., & Muhandhis, I. (2021). TEKNIK DATA MINING UNTUK MENGKLASIFIKASIKAN DATA ULASAN DESTINASI WISATA MENGGUNAKAN REDUKSI DATA PRINCIPAL COMPONENT ANALYSIS (PCA). *Edutic - Scientific Journal of Informatics Education, 7(2).* <https://doi.org/10.21107/edutic.v7i2.9247>
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics, 20*, 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- Saidah, I. & Raiza Athifah Nisrina. (2024). Perilaku Narsis di Instagram: Studi Kasus pada Remaja Perempuan di Pamekasan. *DA'WA: Jurnal Bimbingan Penyuluhan & Konseling Islam, 4(1).*  
<https://doi.org/10.36420/dawa.v4i1.513>
- Santoso, A. B., Supriana, T., & Girsang, M. A. (2022). Pengaruh Curah Hujan terhadap Produksi Ubi Kayu di Indonesia. *Agro Bali : Agricultural Journal, 5(3)*, 520–528. <https://doi.org/10.37637/ab.v5i3.1051>
- Sarker, S. (2022). Fundamentals of Climatology for Engineers: Lecture Note. *Eng, 3(4)*, 573–595. <https://doi.org/10.3390/eng3040040>

- Sartika, D., Saluza, I., & Irfani, M. H. (2022). Perbandingan Akurasi Metode Principal Component Analysis (PCA) dan Correlation-Based Feature Selection (CFS) Pada Klasifikasi Perpanjangan Kontrak Karyawan Menggunakan Metode Naïve Bayes. *Jurnal Ilmiah Informatika Global*, 13(2). <https://doi.org/10.36982/jiig.v13i2.2292>
- Setiyowati, D. A., & Ariska, M. (2024). Analisis Pola Curah Hujan Di Pulau Jawa Dengan Menggunakan Empirical Orthogonal Function (EOF). *Jurnal Pendidikan Fisika dan Sains (JPFS)*, 7(2), 120–128. <https://doi.org/10.52188/jpfs.v7i2.788>
- Suhadi, S., Mabruroh, F., Wiyanto, A., & Ikra, I. (2023). ANALISIS FENOMENA PERUBAHAN IKLIM TERHADAP CURAH HUJAN EKSTRIM. *OPTIKA: Jurnal Pendidikan Fisika*, 7(1), 94–100. <https://doi.org/10.37478/optika.v7i1.2738>
- Sunarmi, N., Hasanah, R., Fitriana, R., & Hamidah, I. N. (2022). *Analisis Unsur Cuaca pada Pertanian Bawang Merah Kabupaten Nganjuk Tahun 2019 dengan Principal Component Analysis*. 2(1).
- Suryaman, S. A., Magdalena, R., & Sa'idah, S. (2021). Klasifikasi Cuaca Menggunakan Metode VGG-16, Principal Component Analysis Dan K-Nearest Neighbor. *Jurnal Ilmu Komputer dan Informatika*, 1(1), 1–8. <https://doi.org/10.54082/jiki.1>
- Wijaya, H., & Anondho, B. (2021). Analisis Faktor Eksternal Yang Dominan Terhadap Biaya Overhead Proyek Konstruksi. *JMTS: Jurnal Mitra Teknik Sipil*, 4(4), 811. <https://doi.org/10.24912/jmts.v4i4.12592>

Yuningsih, Y., Manessa, M. D. M., & Setiadi, H. (2024). Analisis Curah Hujan

Dominan Desa Ciputri: Risiko Bencana Hidroklimatologi. *Media*

*Komunikasi Geografi*, 25(1), 150–166.

<https://doi.org/10.23887/mkg.v25i1.73029>

Zapata-Sierra, A. J., Zapata-Castillo, L., & Manzano-Agugliaro, F. (2022). Water

resources availability in southern Europe at the basin scale in response to

climate change scenarios. *Environmental Sciences Europe*, 34(1), 75.

<https://doi.org/10.1186/s12302-022-00649-5>

# **LAMPIRAN**

PC1	Label_Cluster	Cluster
0,245141	Cluster 1	Ringan
-0,23609	Cluster 0	Ringan
0,801121	Cluster 1	Ringan
2,33316	Cluster 2	Ringan
-0,7739	Cluster 0	Sedang
1,60023	Cluster 1	Ringan
1,723121	Cluster 1	Ringan
-1,14015	Cluster 0	Ringan
-0,26948	Cluster 0	Ringan
-1,25578	Cluster 0	Ringan
1,245811	Cluster 1	Sedang
-0,3787	Cluster 0	Ringan
0,547423	Cluster 1	Ringan
2,067901	Cluster 2	Ringan
-0,47637	Cluster 0	Sedang
-3,66701	Cluster 0	Tinggi
-2,9246	Cluster 0	Ringan
-2,99338	Cluster 0	Sedang
-1,73099	Cluster 0	Sedang
-2,1391	Cluster 0	Ringan
-1,70711	Cluster 0	Ringan
-1,58214	Cluster 0	Ringan
-0,91741	Cluster 0	Ringan
-2,12577	Cluster 0	Ringan
-1,85438	Cluster 0	Ringan
-1,11636	Cluster 0	Ringan
-0,31027	Cluster 0	Ringan
-1,18817	Cluster 0	Ringan
0,077833	Cluster 1	Ringan
-0,77912	Cluster 0	Ringan
-0,85315	Cluster 0	Ringan
-1,81655	Cluster 0	Ringan
0,570334	Cluster 1	Ringan
-2,94023	Cluster 0	Ringan
-2,0613	Cluster 0	Sedang
-2,09586	Cluster 0	Ringan
1,266983	Cluster 1	Ringan
0,049774	Cluster 1	Ringan
-0,50912	Cluster 0	Sedang
-0,08077	Cluster 0	Sedang
-0,52627	Cluster 0	Ringan
1,111281	Cluster 1	Ringan

-0,87853	Cluster 0	Ringan
-0,9231	Cluster 0	Ringan
-1,00039	Cluster 0	Ringan
-0,83248	Cluster 0	Sedang
-1,49805	Cluster 0	Ringan
2,567873	Cluster 2	Sedang
0,54163	Cluster 1	Ringan
-0,50754	Cluster 0	Ringan
2,133731	Cluster 2	Ringan
2,196961	Cluster 2	Ringan
1,301979	Cluster 1	Sedang
2,499276	Cluster 2	Ringan
3,330702	Cluster 2	Ringan
2,552486	Cluster 2	Sedang
1,259564	Cluster 1	Ringan
2,529729	Cluster 2	Ringan
1,008595	Cluster 1	Ringan
-1,27978	Cluster 0	Ringan
-0,43061	Cluster 0	Ringan
0,743559	Cluster 1	Sedang
-0,07334	Cluster 0	Ringan
-0,66103	Cluster 0	Ringan
-1,12706	Cluster 0	Ringan
0,282877	Cluster 1	Sedang
0,670044	Cluster 1	Ringan
-2,30517	Cluster 0	Ringan
-1,90181	Cluster 0	Ringan
-0,61909	Cluster 0	Ringan
-0,45376	Cluster 0	Ringan
0,146659	Cluster 1	Ringan
-1,84871	Cluster 0	Sedang
0,22067	Cluster 1	Ringan
2,647926	Cluster 2	Ringan
0,938063	Cluster 1	Ringan
-2,81971	Cluster 0	Ringan
1,168732	Cluster 1	Ringan
0,644828	Cluster 1	Ringan
2,17692	Cluster 2	Ringan
-0,00021	Cluster 0	Ringan
2,199254	Cluster 2	Ringan
3,074354	Cluster 2	Ringan
2,380446	Cluster 2	Ringan
-1,41155	Cluster 0	Ringan

0,898548	Cluster 1	Ringan
1,346408	Cluster 1	Ringan
2,016737	Cluster 2	Sedang
-0,25342	Cluster 0	Sedang
3,616225	Cluster 2	Ringan
-1,62535	Cluster 0	Ringan
-0,99392	Cluster 0	Ringan
0,671514	Cluster 1	Ringan
0,255951	Cluster 1	Ringan
-1,22241	Cluster 0	Ringan
-0,72973	Cluster 0	Ringan
0,846289	Cluster 1	Ringan
-0,11277	Cluster 0	Tinggi
1,102021	Cluster 1	Ringan
1,920303	Cluster 1	Ringan
2,916557	Cluster 2	Ringan
1,370798	Cluster 1	Ringan
-0,84133	Cluster 0	Ringan
-2,24281	Cluster 0	Ringan
1,109768	Cluster 1	Ringan
0,182853	Cluster 1	Ringan
-1,56512	Cluster 0	Sedang
0,070155	Cluster 1	Ringan
0,162585	Cluster 1	Ringan
1,051941	Cluster 1	Ringan
3,077569	Cluster 2	Ringan
1,42963	Cluster 1	Ringan
0,128157	Cluster 1	Ringan
1,270013	Cluster 1	Ringan
1,03933	Cluster 1	Ringan
0,449754	Cluster 1	Ringan
-2,92429	Cluster 0	Sedang
-2,41693	Cluster 0	Ringan
-2,07985	Cluster 0	Sedang
-0,2323	Cluster 0	Ringan
-0,95863	Cluster 0	Ringan
-2,56231	Cluster 0	Sedang
-0,80352	Cluster 0	Ringan
-0,05902	Cluster 0	Ringan
-1,41774	Cluster 0	Ringan
-2,64194	Cluster 0	Ringan
-2,43161	Cluster 0	Ringan
-2,46295	Cluster 0	Ringan

-1,35695	Cluster 0	Ringan
-1,81659	Cluster 0	Sedang
-0,53646	Cluster 0	Ringan
-0,17686	Cluster 0	Ringan
0,391381	Cluster 1	Ringan
-0,96769	Cluster 0	Tinggi
-2,61804	Cluster 0	Ringan
-1,2946	Cluster 0	Ringan
-3,0847	Cluster 0	Tinggi
-1,42494	Cluster 0	Ringan
-1,77827	Cluster 0	Ringan
-0,57488	Cluster 0	Ringan
0,763308	Cluster 1	Ringan
1,215063	Cluster 1	Ringan
2,214942	Cluster 2	Ringan
-0,99986	Cluster 0	Ringan
-0,1342	Cluster 0	Ringan
-1,71546	Cluster 0	Ringan
0,626164	Cluster 1	Sedang
0,409765	Cluster 1	Ringan
0,20395	Cluster 1	Ringan
-1,01532	Cluster 0	Sedang
-0,65652	Cluster 0	Ringan
2,075413	Cluster 2	Ringan
-0,27638	Cluster 0	Ringan
1,581617	Cluster 1	Ringan
0,902398	Cluster 1	Ringan
0,847646	Cluster 1	Ringan
0,302004	Cluster 1	Tinggi
0,969886	Cluster 1	Ringan
0,367671	Cluster 1	Sedang
0,97816	Cluster 1	Ringan
-0,67051	Cluster 0	Ringan
-0,78723	Cluster 0	Ringan
-1,0954	Cluster 0	Ringan
1,943388	Cluster 1	Ringan
2,14431	Cluster 2	Ringan
0,290528	Cluster 1	Ringan
-2,44743	Cluster 0	Sedang
-0,12665	Cluster 0	Ringan
-1,61195	Cluster 0	Ringan
0,666473	Cluster 1	Ringan
-0,88094	Cluster 0	Ringan

2,18532	Cluster 2	Ringan
1,29035	Cluster 1	Ringan
2,132419	Cluster 2	Ringan
-0,69238	Cluster 0	Tinggi
-2,67306	Cluster 0	Ringan
-1,46037	Cluster 0	Ringan
0,871766	Cluster 1	Ringan
1,370153	Cluster 1	Ringan
1,983562	Cluster 1	Ringan
1,977606	Cluster 1	Ringan
2,025317	Cluster 2	Ringan
0,956245	Cluster 1	Ringan
1,375005	Cluster 1	Ringan
0,218611	Cluster 1	Ringan
-1,40072	Cluster 0	Ringan
1,425458	Cluster 1	Sedang
1,937931	Cluster 1	Ringan
0,578403	Cluster 1	Ringan
0,651383	Cluster 1	Ringan
0,104442	Cluster 1	Sedang
1,060489	Cluster 1	Ringan
-0,9907	Cluster 0	Ringan
0,154762	Cluster 1	Ringan
0,756831	Cluster 1	Ringan
-0,60536	Cluster 0	Ringan
0,449968	Cluster 1	Ringan
2,319136	Cluster 2	Ringan
1,050548	Cluster 1	Ringan
2,05717	Cluster 2	Ringan
0,819978	Cluster 1	Ringan
-2,62676	Cluster 0	Sedang
1,659638	Cluster 1	Ringan