

BAB II TINJAUAN PUSTAKA

A. Penelitian Terkait

Dalam penelitian yang akan dilakukan yaitu dengan judul “Prediksi Curah Hujan Harian Dengan Metode *K-Nearest Neighbor* (KNN) Menggunakan Python (Studi Kasus di Stasiun Meteorologi Tunggul Wulung Cilacap)”, terdapat beberapa penelitian yang terkait sebagai referensi bagi penulis yaitu sebagai berikut:

Tabel 1. Penelitian Terkait

No	Peneliti	Keterangan Penelitian	Perbedaan
(1)	(2)	(3)	(4)
1	Deden Martia Nanda, Tacbir Hendro Pudjiantoro dan Puspita Nurul Sabrina (2022) [6]	Judul: Metode <i>K-Nearest Neighbor</i> (KNN) Dalam Memprediksi Curah Hujan di Kota Bandung. Metode: <i>K-Nearest Neighbor</i> (KNN). Variabel: Temperatur rata-rata, kelembapan rata-rata, kecepatan angin rata-rata, dan curah hujan. Software: Microsoft Excel. Hasil Penelitian: Berdasarkan pengujian dengan nilai $k = 5$ menggunakan <i>K-Cross Validation</i> diperoleh akurasi sebesar 86,199%. Sementara itu, evaluasi melalui <i>Confusion Matrix</i> menghasilkan akurasi sebesar 85,78%.	Pada penelitian yang penulis lakukan, terdapat perbedaan pada: Lokasi Penelitian: Penelitian dilakukan di Kabupaten Cilacap. Variabel: Variabel tambahannya adalah tekanan udara rata-rata, dan lama penyinaran matahari. Software: Python

No	Peneliti	Keterangan Penelitian	Perbedaan
(1)	(2)	(3)	(4)
2	Nursobah, Siti Lailiyah, Bartolomius Harpad, dan Muhammad Fahmi (2022) [15]	Judul: Penerapan Data Mining Untuk Prediksi Perkiraan Hujan dengan Menggunakan Algoritma <i>K-Nearest Neighbor</i>. Metode: <i>K-Nearest Neighbor</i>. Variabel: Suhu, kelembapan udara dan kecepatan angin. Software: Microsoft Excel. Hasil Penelitian: Hasil prediksi pada data <i>testing</i> menunjukkan keputusan “tidak”. Dengan demikian, penerapan <i>data mining</i> dengan algoritma <i>K-Nearest Neighbor</i> terbukti dapat mendukung penyelesaian permasalahan.	Pada penelitian yang penulis lakukan, terdapat perbedaan pada: Lokasi Penelitian: Penelitian dilakukan di Kabupaten Cilacap. Objek penelitian: Objek pada penelitian adalah Curah hujan harian yang dikategorikan menjadi tidak hujan, hujan ringan, hujan sedang, hujan lebat, hujan sangat lebat, dan hujan ekstrem. Variabel: Variabel tambahannya adalah curah hujan harian, tekanan udara rata-rata, dan lama penyinaran matahari. Software: Python
3	Muhammad Yusuf R. R., Muhammad Valensyah A., dan Wawan Gunawan (2021) [14]	Judul: Penerapan Algoritma <i>K-Nearest Neighbor</i> (KNN) Dalam Memprediksi dan Menghitung Tingkat Akurasi Data Cuaca di Indonesia. Metode: <i>K-Nearest Neighbor</i> (KNN).	Pada penelitian yang penulis lakukan, terdapat perbedaan pada: Lokasi Penelitian: Penelitian dilakukan di Kabupaten Cilacap. Objek Penelitian: Objek pada penelitian adalah Curah hujan harian yang

No	Peneliti	Keterangan Penelitian	Perbedaan
(1)	(2)	(3)	(4)
		Variabel: Kelembapan, suhu, suhu minimum, suhu maksimum, kecepatan rata-rata, kecepatan maksimum, dan curah hujan. Software: Python. Hasil Penelitian: Penelitian ini menggunakan sebanyak 3.623 data, dengan perbandingan 80% untuk data <i>training</i> dan 20% untuk data <i>testing</i>. Hasil prediksi menunjukkan tingkat akurasi sebesar 89%.	dikategorikan menjadi tidak hujan, hujan ringan, hujan sedang, hujan lebat, hujan sangat lebat, dan hujan ekstrem. Variabel: Variabel digunakan adalah curah hujan harian, suhu udara rata-rata, kelembapan rata-rata, tekanan udara rata-rata, kecepatan angin rata-rata dan lama penyinaran matahari.
4	Vita Fatimah, Ahmad Zaenal Arifin, dan Suatmo Purto (2024) [13]	Judul: Peramalan Curah Hujan di Kabupaten Tuban Menggunakan Algoritma KNN. Metode: <i>K-Nearest Neighbor</i>. Variabel: Data tahunan dari Suhu, kelembapan, kecepatan angin, tekanan udara, dan curah hujan. Software: Python. Hasil Penelitian: Berdasarkan hasil prediksi, diperoleh RMSE sebesar 36.3942. Selain itu,	Pada penelitian yang penulis lakukan, terdapat perbedaan pada: Lokasi Penelitian: Penelitian dilakukan di Kabupaten Cilacap. Variabel: Data harian dari curah hujan, suhu udara rata-rata, kelembapan rata-rata, tekanan udara rata-rata, kecepatan angin rata-rata dan lama penyinaran matahari.

No	Peneliti	Keterangan Penelitian	Perbedaan
(1)	(2)	(3)	(4)
		nilai <i>R-Square</i> sebesar 0.6214 menunjukkan bahwa curah hujan di Kabupaten Tuban dipengaruhi oleh 4 variabel yang dimasukkan dalam model, sehingga algoritma <i>K-Nearest Neighbor</i> (KNN) mampu menjelaskan hubungan antara variabel <i>X</i> dalam meramalkan curah hujan. Adapun sisanya, yaitu sebesar 0,3786 dipengaruhi oleh variabel lain yang tidak dimasukkan dalam model.	
5	Muhammad Alviriza Ramadhan, Fetty Tri Anggraeny, dan Chrystia Aji Putra (2024) [16]	Judul: Klasifikasi Curah Hujan Harian Menggunakan Metode <i>K-Nearest Neighbor</i>. Metode: <i>K-Nearest Neighbor</i>. Variabel: Suhu minimal, suhu maksimal, suhu rata-rata, kelembapan rata-rata, penyinaran matahari, kecepatan angin, kecepatan angin rata-rata. Software: Python. Hasil Penelitian:	Pada penelitian yang penulis lakukan, terdapat perbedaan pada: Lokasi Penelitian: Penelitian dilakukan di Kabupaten Cilacap. Variabel: Variabel digunakan adalah curah hujan, suhu udara rata-rata, kelembapan rata-rata, tekanan udara rata-rata, kecepatan angin rata-rata dan lama penyinaran matahari.

No	Peneliti	Keterangan Penelitian	Perbedaan
(1)	(2)	(3)	(4)
		Berdasarkan hasil penelitian, dapat disimpulkan bahwa klasifikasi curah hujan harian menggunakan metode <i>K-Nearest Neighbor</i> menunjukkan tingkat akurasi yang cukup akurat. Hal ini ditunjukkan oleh nilai akurasi rata-rata yang berada diatas 80%, dengan nilai akurasi tertinggi sebesar 86,09% diperoleh pada nilai $k = 3$ dan perbandingan data <i>training</i> dan data <i>testing</i> yaitu 60:40.	
6	Andi Sadri Agung, A. Arfan Fauzi, Andi Akram Nur Risal, dan Fhatiah Adiba (2023) [17]	Judul : Implementasi Teknik <i>Data Mining</i> terhadap Klasifikasi Data Prediksi Curah Hujan BMKG Di Sulawesi Selatan. Metode: <i>K-Nearest Neighbor</i>. Variabel: Suhu Rata-rata, Kelembapan Rata-rata, dan Kecepatan Angin Rata-rata. Software: Python. Hasil Penelitian: Implementasi <i>data mining</i> menggunakan	Pada penelitian yang penulis lakukan, terdapat perbedaan pada: Lokasi Penelitian: Penelitian dilakukan di Kabupaten Cilacap. Variabel: Variabel tambahannya adalah curah hujan harian, tekanan udara rata-rata, dan lama penyinaran matahari.

No	Peneliti	Keterangan Penelitian	Perbedaan
(1)	(2)	(3)	(4)
		algoritma <i>K-Nearest Neighbor</i> (KNN), pada data prediksi curah hujan BMKG di provinsi Sulawesi Selatan menunjukkan hasil yang baik. Nilai $k = 3$ sebagai nilai k terbaik, menghasilkan akurasi tertinggi sebesar 82,21%. Dengan demikian, dapat disimpulkan bahwa data prediksi curah hujan BMKG memiliki tingkat akurasi data yang tergolong tinggi.	
7	Adinda Amalia, Ati Zaidiah, dan Ika Nurlaili Isnainiyah (2022) [18]	Judul: Prediksi Kualitas Udara Menggunakan Algoritma <i>K-Nearest Neighbor</i>. Metode: <i>K-Nearest Neighbor</i>. Variabel: PM10, PM25, SO₂, CO, O₃, NO₂, dan Max. Software: Python. Hasil Penelitian: Berdasarkan hasil penelitian dapat disimpulkan bahwa algoritma KNN efektif digunakan untuk memprediksi kualitas	Pada penelitian yang penulis lakukan, terdapat perbedaan pada: Lokasi Penelitian: Penelitian dilakukan di Kabupaten Cilacap. Objek Penelitian: Objek pada penelitian adalah Curah hujan harian yang dikategorikan menjadi tidak hujan, hujan ringan, hujan sedang, hujan lebat, hujan sangat lebat, dan hujan ekstrem. Variabel: Variabel digunakan adalah curah hujan

No	Peneliti	Keterangan Penelitian	Perbedaan
(1)	(2)	(3)	(4)
		udara berdasarkan Indeks Standar Pencemar Udara (ISPU). Evaluasi menggunakan <i>confusion matrix</i> menunjukkan bahwa nilai $k = 7$ memberikan performa terbaik dimana, dengan akurasi sebesar 96%, <i>presisi</i> 92%, <i>recall</i> 95%, dan <i>f-measure</i> 93%.	harian, suhu udara rata-rata, kelembapan rata-rata, tekanan udara rata-rata, kecepatan angin rata-rata dan lama penyinaran matahari.
8	Clariza Adelina Rachma (2022) [9]	Judul: Implementasi Algoritma <i>K-Nearest Neighbor</i> Dalam Penentuan Klasifikasi Tingkat Kedalaman Kemiskinan Provinsi Jawa Timur. Metode: <i>K-Nearest Neighbor</i>. Variabel: Indeks pengeluaran perkapita, rata-rata lama sekolah, laju pertumbuhan produk domestik regional bruto, dan tingkat pengangguran terbuka. Software: Microsoft Excel dan Rapidminer. Hasil Penelitian: Hasil klasifikasi menggunakan algoritma <i>K-Nearest Neighbor</i>	Pada penelitian yang penulis lakukan, terdapat perbedaan pada: Lokasi Penelitian: Penelitian dilakukan di Kabupaten Cilacap. Objek Penelitian: Objek pada penelitian adalah Curah hujan harian yang dikategorikan menjadi tidak hujan, hujan ringan, hujan sedang, hujan lebat, hujan sangat lebat, dan hujan ekstrem. Variabel: Variabel digunakan adalah curah hujan harian, suhu udara rata-rata, kelembapan rata-rata, tekanan udara rata-rata, kecepatan angin rata-

No	Peneliti	Keterangan Penelitian	Perbedaan
(1)	(2)	(3)	(4)
		menunjukkan bahwa algoritma ini dapat diimplementasikan dengan baik untuk permasalahan yang dikaji dalam penelitian ini. Berdasarkan evaluasi dengan nilai k terbaik yaitu, $k = 1$ dan $k = 2$, diperoleh nilai akurasi sebesar 76.67%.	rata dan lama penyinaran matahari. Software: Python
9	Aisha Alfani W. P. R, Fahrur Rozi dan Farid Sukmana (2021) [8]	Judul: Prediksi Penjualan Produk Unilever Menggunakan Metode <i>K-Nearest Neighbor</i>. Metode: <i>K-Nearest Neighbor</i>. Variabel: Produk dari Indofood, P&G, dan Nestlé. Software: Microsoft Excel Hasil Penelitian: Berdasarkan hasil penelitian, diperoleh nilai akurasi tertinggi sebesar 86,66%, sedangkan nilai akurasi terendah sebesar 40%. Dengan demikian, algoritma KNN dapat diterapkan untuk memprediksi penjualan produk Unilever di Toko Rizky Barokah Nganjuk.	Pada penelitian yang penulis lakukan, terdapat perbedaan pada: Lokasi Penelitian: Penelitian dilakukan di Kabupaten Cilacap. Objek Penelitian: Objek pada penelitian adalah Curah hujan harian Variabel: Variabel digunakan adalah curah hujan harian, suhu udara rata-rata, kelembapan rata-rata, tekanan udara rata-rata, kecepatan angin rata-rata dan lama penyinaran matahari. Software: Python

No	Peneliti	Keterangan Penelitian	Perbedaan
(1)	(2)	(3)	(4)
10	Irma Darmayanti, Pamukas Subarkah, Luky Rafi Anunggilarsa, dan Jali Suhaman (2021) [11]	Judul: Prodiksi Potensi Siswa Putus Sekolah Akibat Pandemi Covid-19 Menggunakan Algoritma <i>K-Nearest Neighbor</i>. Metode: <i>K-Nearest Neighbor</i>. Variabel: Rata-rata penghasilan orang tua selama tiga bulan terakhir, jumlah keluarga yang masih bersekolah, tunggakan keluarga yang masih bersekolah, status kepemilikan KIP, status perolehan Bantuan Langsung Tunai (BLT) di masa pandem, status perolehbantuan sembako dari pemerintah di masa pandemi. Software: WEKA Hasil Penelitian: Berdasarkan hasil penelitian, disimpulkan bahwa algoitme KNN mampu memprediksi siswa putus sekolah di masa pandemi, denagn nilai akurasi sebesar 87,42%, <i>precision</i> 88,2%, <i>recall</i> 87,4%	Pada penelitian yang penulis lakukan, terdapat perbedaan pada: Lokasi Penelitian: Penelitian dilakukan di Kabupaten Cilacap. Objek Penelitian: Objek pada penelitian adalah Curah hujan harian yang dikategorikan menjadi tidak hujan, hujan ringan, hujan sedang, hujan lebat, hujan sangat lebat, dan hujan ekstrem. Variabel: Variabel digunakan adalah curah hujan harian, suhu udara rata-rata, kelembapan rata-rata, tekanan udara rata-rata, kecepatan angin rata-rata dan lama penyinaran matahari. Software: Python

No	Peneliti	Keterangan Penelitian	Perbedaan
(1)	(2)	(3)	(4)
		dan <i>f-measure</i> 87%. Berdasarkan klasifikasi performa, nilai akurasi tersebut termasuk dalam kategori <i>Good Classification</i> .	
11	Muhammad Naja Maskuri, Harlina, Kadek Sukerti, dan R. M. Hardian Bhakti (2022) [19]	Judul: Penerapan Algoritma <i>K-Nearest Neighbor</i> (KNN) Untuk Memprediksi Penyakit Stroke. Metode: <i>K-Nearest Neighbor</i>. Variabel: Jenis kelamin, umur, hipertensi, penyakit jantung, pernah menikah, tipe pekerjaan, tipe tempat tinggal, terata kadar glukosa, BMI, dan status merokok. Software: Microsoft Excel Hasil Penelitian: Penelitian ini menggunakan 100 data, dengan pembagian 80% untuk data latih dan 20% untuk data uji. Berdasarkan pengujian dengan nilai $k = 9$, diperoleh akurasi sebesar 95%. Sehingga algoritma KNN menunjukkan tingkat	Pada penelitian yang penulis lakukan, terdapat perbedaan pada: Lokasi Penelitian: Penelitian dilakukan di Kabupaten Cilacap. Objek Penelitian: Objek pada penelitian adalah Curah hujan harian yang dikategorikan menjadi tidak hujan, hujan ringan, hujan sedang, hujan lebat, hujan sangat lebat, dan hujan ekstrem. Variabel: Variabel digunakan adalah curah hujan harian, suhu udara rata-rata, kelembapan rata-rata, tekanan udara rata-rata, kecepatan angin rata-rata dan lama penyinaran matahari. Software: Python

No	Peneliti	Keterangan Penelitian	Perbedaan
(1)	(2)	(3)	(4)
		akurasi baik dalam memprediksi penyakit stroke.	
12	Sri Ayu Rizky, Rolly Yasputra, dan Santoso (2021) [20]	Judul: Prediksi Kelancaran Pembayaran Cicilan Calon Debitur dengan Metode <i>K-Nearest Neighbor</i>. Metode: <i>K-Nearest Neighbor</i>. Variabel: Nomor pinjaman, jenis kelamin, status pernikahan, jenis usaha, jumlah pinjaman, dan jenis pinjaman. Software: Manual, Rapidminer, dan Website. Hasil Penelitian: Hasil perhitungan manual, pengolahan menggunakan Rapidminer serta aplikasi prediksi yang dikembangkan menunjukkan konsistensi hasil yang sama. Pengujian menggunakan Rapidminer menghasilkan akurasi 73,37% pada nilai $k = 9$.	Pada penelitian yang penulis lakukan, terdapat perbedaan pada: Lokasi Penelitian: Penelitian dilakukan di Kabupaten Cilacap. Objek Penelitian: Objek pada penelitian adalah Curah hujan harian yang dikategorikan menjadi tidak hujan, hujan ringan, hujan sedang, hujan lebat, hujan sangat lebat, dan hujan ekstrem. Variabel: Variabel digunakan adalah curah hujan harian, suhu udara rata-rata, kelembapan rata-rata, tekanan udara rata-rata, kecepatan angin rata-rata dan lama penyinaran matahari. Software: Python
13	Sriani, Aidil Halim Lubis, dan Ridho	Judul: Penerapan <i>Data Mining</i> untuk Prediksi	Pada penelitian yang penulis lakukan,

No	Peneliti	Keterangan Penelitian	Perbedaan
(1)	(2)	(3)	(4)
	Riky Nasution (2024) [21]	Penjualan Produk Elektronik Terlaris Menggunakan Metode <i>K-Nearest Neighbor</i>. Metode: <i>K-Nearest Neighbor</i>. Variabel: Harga dan Banyaknya yang terjual. Software: Website. Hasil Penelitian: Sistem prediksi produk elektronik terlaris yang menerapkan algoritma KNN dapat membantu perusahaan untuk meningkatkan penjualan serta mengoptimalkan pengolahan persediaan stok. Berdasarkan hasil pengujian, Sistem tersebut menghasilkan tingkat akurasi 85,8% dengan perbandingan data <i>training</i> dan data <i>testing</i> 80% : 20%.	terdapat perbedaan pada: Lokasi Penelitian: Penelitian dilakukan di Kabupaten Cilacap. Objek Penelitian: Objek pada penelitian adalah Curah hujan harian yang dikategorikan menjadi tidak hujan, hujan ringan, hujan sedang, hujan lebat, hujan sangat lebat, dan hujan ekstrem. Variabel: Variabel digunakan adalah curah hujan harian, suhu udara rata-rata, kelembapan rata-rata, tekanan udara rata-rata, kecepatan angin rata-rata dan lama penyinaran matahari. Software: Python

B. Landasan Teori

1. Curah Hujan

a. Definisi Curah Hujan

Hujan merupakan peristiwa jatuhnya hidrometeor berupa partikel-partikel air dari atmosfer menuju permukaan bumi. Hidrometeor dapat berupa butir-butir air maupun kristal es yang terbentuk di dalam awan kemudian jatuh sebagai presipitasi yang dikenal sebagai curah hujan

[2]. Curah hujan didefinisikan sebagai jumlah jatuhnya butir-butir atau tetes air hujan yang mencapai permukaan bumi pada periode tertentu. Lama hujan merujuk pada durasi terjadinya peristiwa hujan dalam interval waktu, sedangkan banyaknya hujan merupakan total volume air hujan yang jatuh dan mencapai permukaan bumi selama selang waktu pengamatan atau pengukuran [2]. Banyaknya curah hujan umumnya dinyatakan dalam satuan milimeter (mm).

Berdasarkan klasifikasi yang ditetapkan oleh Badan Meteorologi, Klimatologi, dan Geofisika (BMKG), curah hujan dibagi kedalam enam kategori curah hujan, yaitu ditunjukkan pada Tabel 2 sebagai berikut [22]:

Tabel 2. Kategori Curah Hujan

Kategori	Rentang Curah Hujan (mm/hari)
Tidak Hujan	$x < 5$
Hujan Ringan	$5 \leq x < 20$
Hujan Sedang	$20 \leq x < 50$
Hujan Lebat	$50 \leq x < 100$
Hujan Sangat Lebat	$100 \leq x < 150$
Hujan Ekstrem	$x \leq 150$

b. Pengukuran Curah Hujan di Stasiun Meteorologi

Curah hujan dalam pengamatan meteorologi permukaan (*sinoptik*) adalah jumlah curah hujan yang berupa air dan atau banyaknya air yang berasal dari hujan bekuan (salju, hail dan sebagainya).

Alat pengukuran hujan terdapat 2 (dua) tipe, yaitu sebagai berikut:

1) Tipe kolektor (penampung)

Alat pengukur curah hujan tipe kolektor (penampung) adalah Ombrometer.



Gambar 1. Ombrometer

2) Tipe perekam data (otomatis)

Alat pengukur curah hujan tipe perekam data (otomatis) adalah Ombrograf.



Gambar 2. *Ombrograf*

c. Faktor-Faktor yang Mempengaruhi Curah Hujan

Curah hujan dipengaruhi oleh berbagai faktor meteorologis yang berperan dalam proses pembentukan awan dan presipitasi. Faktor-faktor tersebut meliputi suhu udara, kelembapan udara, tekanan udara, kecepatan angin, dan lama penyinaran matahari [23]. Penjelasan masing-masing faktor adalah sebagai berikut:

1) Suhu Udara

Suhu udara merupakan kondisi yang menunjukkan tingkat panas dan dinginnya suatu tempat. Suhu juga dikenal dengan temperatur. Pengukuran suhu udara dilakukan menggunakan alat yang disebut Thermometer, dengan satuan derajat Celcius ($^{\circ}\text{C}$).

Suhu udara berpengaruh terhadap proses penguapan air dari permukaan bumi. Semakin tinggi suhu udara, maka laju penguapan air akan meningkat sehingga jumlah uap air di atmosfer bertambah. Peningkatan uap air ini dapat memperbesar peluang terbentuknya awan dan terjadinya hujan apabila kondisi atmosfer mendukung proses kondensasi.

2) Kelembapan Udara

Kelembapan udara adalah perbandingan antara massa uap air yang terdapat dalam suatu volume udara dengan massa uap air maksimum (jenuh) yang dapat dikandung oleh udara pada suhu yang sama. Kelembapan udara diukur dengan menggunakan alat yang disebut Higrometer dan dinyatakan dalam satuan Presentase (%).

Kelembapan udara menunjukkan kandungan uap air di atmosfer. Semakin tinggi kelembapan udara, maka semakin besar jumlah uap air yang tersedia untuk membentuk awan. Ketika udara mencapai kondisi jenuh, uap air akan mengalami

kondensasi menjadi butir-butir air yang kemudian dapat jatuh sebagai hujan. Oleh karena itu, kelembapan udara yang tinggi umumnya berkaitan dengan peluang curah hujan yang lebih besar.

3) Tekanan Udara

Tekanan udara merupakan gaya yang diberikan oleh massa udara terhadap setiap satuan luas permukaan. Tekanan ini berperan dalam pergerakan massa udara di atmosfer. Alat yang digunakan untuk mengukur dan mencatat tekanan udara secara otomatis adalah Barograph, dengan satuan Milibar (mb).

Tekanan udara memengaruhi pergerakan dan sirkulasi massa udara. Daerah bertekanan rendah biasanya menyebabkan udara bergerak naik. Udara yang naik akan mengalami pendinginan sehingga uap air mengalami kondensasi dan membentuk awan yang berpotensi menghasilkan hujan. Sebaliknya, daerah bertekanan tinggi cenderung menyebabkan udara bergerak turun sehingga menghambat pembentukan awan dan menurunkan kemungkinan hujan.

4) Kecepatan Angin

Angin adalah gerakan massa udara yang terjadi akibat adanya perbedaan tekanan udara. Udara bergerak dari daerah bertekanan tinggi menuju daerah bertekanan rendah. Kecepatan angin diukur menggunakan alat yang disebut Anemometer, dengan satuan Knot.

Kecepatan angin berperan dalam transportasi massa udara dan uap air dari suatu wilayah ke wilayah lainnya. Angin dapat membawa uap air dari daerah perairan menuju daratan sehingga meningkatkan potensi pembentukan awan hujan. Selain itu, pertemuan massa udara yang bergerak dengan arah berbeda dapat memicu proses konveksi yang meningkatkan peluang terjadinya hujan.

5) Lama Penyinaran Matahari

Lama penyinaran matahari merupakan durasi waktu ketika permukaan bumi menerima radiasi atau cahaya matahari secara langsung. Lama penyinaran matahari diukur dengan menggunakan alat Campbell-Stokes, dengan satuan waktu jam.

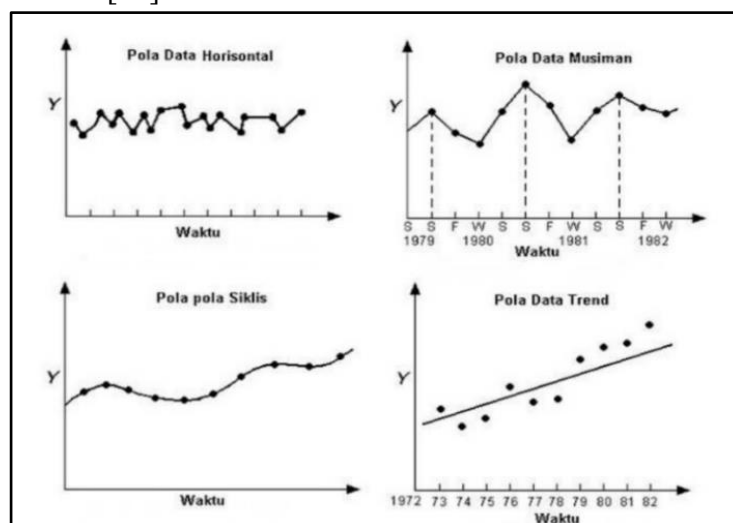
Lama penyinaran matahari memengaruhi energi panas yang diterima permukaan bumi. Penyinaran matahari yang lebih lama dapat meningkatkan suhu permukaan dan mempercepat proses penguapan air. Namun, pada saat terjadi hujan atau banyak awan, intensitas dan durasi penyinaran matahari biasanya berkurang karena tertutup oleh awan hujan.

Kelima faktor meteorologis tersebut saling berkaitan dalam mengendalikan proses siklus hidrologi di atmosfer. Perubahan pada suhu udara, kelembapan, tekanan udara, kecepatan angin, dan lama penyinaran matahari dapat memengaruhi pembentukan awan dan pada akhirnya menentukan besarnya curah hujan yang terjadi di suatu wilayah.

2. Prediksi

Prediksi atau peramalan merupakan suatu metode yang digunakan untuk memperkirakan nilai pada masa yang akan datang dengan mempertimbangkan data historis maupun data yang tersedia pada saat ini. Peramalan dapat dipahami sebagai proses sistematis dalam memperkirakan kemungkinan kejadian di masa depan berdasarkan informasi dari masa lalu, dengan tujuan meminimalkan tingkan kesalahan (*error*), yaitu selisih antara nilai aktual dan nilai hasil prediksi. Istilah peramalan memiliki makna yang sepadan dengan prediksi atau estimasi. Namun demikian, prediksi tidak dimaksudkan untuk memberikan kepastian mutlak mengenai apa yang akan terjadi di masa mendatang, melainkan untuk menghasilkan perkiraan yang sedekat mungkin dengan kondisi yang diperkirakan akan terjadi [24].

Deter waktu (*time series*) merupakan rangkaian pengamatan terhadap suatu variabel yang dikumpulkan secara beruntun dari waktu ke waktu, serta dicatat berdasarkan urutan kejadian dengan interval waktu yang konsisten. Data *time series* adalah data yang dikumpulkan, direkam, atau diamati mengikuti urutan waktu tertentu. Dengan memperhatikan pola data yang terbentuk, metode peramalan yang paling sesuai dan efektif dapat ditentukan. Secara umum terdapat 4 (empat) jenis pola data, yaitu sebagai berikut [25]:



Gambar 3. Jenis Pola Data

- a. *Horizontal (H)*
Pola data *horizontal* atau stasioner, terjadi ketika nilai data berfluktuasi disekitar suatu nilai rata-rata yang relatif konstan. Dalam kondisi ini, data cenderung stabil dan tidak menunjukkan kecenderungan peningkatan maupun penurunan yang signifikan, sehingga dikatakan stasioner terhadap nilai rata-rata.
- b. *Seasonality (S)*
Pola data *seasonality* atau musiman, muncul ketika data deret waktu dipengaruhi oleh faktor musiman yang berulang secara teratur dari satu periode ke periode berikutnya. Pola ini dapat dilihat dalam bentuk pengulangan pada interval waktu tertentu, seperti harian, mingguan, bulanan, tahunan, maupun pada periode kuartalan.
- c. *Cycles (C)*
Pola data *cycles* atau siklus, terjadi ketika data menunjukkan pergerakan naik dan turun dalam jangka waktu tertentu, namun tidak mengikuti periode yang tetap atau teratur. Dengan demikian, fluktuasi yang terjadi bersifat tidak konsisten dan dapat muncul secara acak dalam jangka waktu berbeda.
- d. *Trend (T)*
Pola data *trend*, terjadi apabila data pengamatan menunjukkan kecenderungan meningkat dan menurun dalam jangka waktu yang relatif panjang. Data yang memiliki pola *trend* umumnya dikategorikan sebagai pola data nonstasioner, karena nilai rata-ratanya berubah seiring waktu.

Prediksi merupakan suatu proses untuk memperkirakan nilai atau kondisi yang akan terjadi di masa mendatang berdasarkan data atau informasi yang tersedia pada masa sebelumnya. Dalam penelitian ilmiah, prediksi dilakukan melalui pendekatan analisis yang sistematis dengan memanfaatkan data histori serta pola hubungan antar variabel yang dapat diidentifikasi. Menurut pendekatan dalam bidang statistika, prediksi dilakukan dengan membangun model matematis atau statistik yang mampu menggambarkan kecenderungan suatu fenomena sehingga dapat digunakan untuk memperkirakan nilai pada periode berikutnya. Seiring dengan perkembangan teknologi, metode prediksi juga banyak dikembangkan dalam bidang Ilmu Data dan *Machine Learning* yang memungkinkan analisis data dalam jumlah besar untuk menghasilkan prediksi yang lebih akurat. Meskipun demikian, hasil prediksi tetap mengandung unsur ketidakpastian karena dipengaruhi oleh keterbatasan data, perubahan kondisi lingkungan, serta variabel lain yang mungkin tidak tercakup dalam model yang digunakan. Oleh karena itu, dalam penelitian, prediksi tidak dimaksudkan untuk memberikan kepastian mutlak, melainkan sebagai estimasi ilmiah yang dapat membantu dalam

memahami kemungkinan perkembangan suatu fenomena di masa yang akan datang.

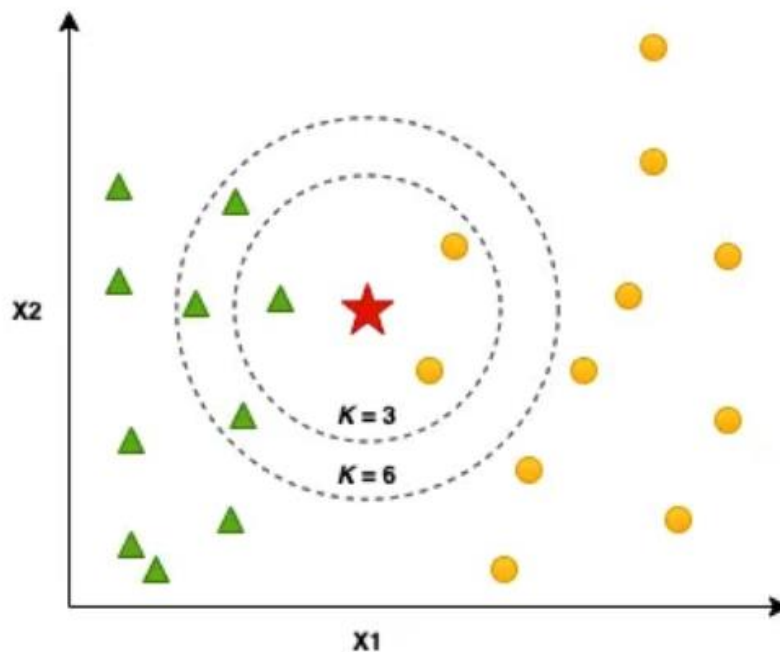
Metode prediksi dalam bidang Ilmu Data dan *Machine Learning* atau dapat disebut juga metode *heuristik* modern, dimana metode tersebut dikembangkan berdasarkan algoritma dalam bidang *soft computasi* (SC). Salah satu metode yang termasuk dalam kategori metode *heuristik* modern adalah metode *K-Nearest Neighbor* (KNN).

3. *K-Nearest Neighbor* (KNN)

Metode *K-Nearest Neighbor* (KNN) merupakan salah satu metode nonparametrik yang digagas oleh Thomas Cover dan digunakan dalam proses klasifikasi berdasarkan kedekatan dengan sejumlah tetangga terdekat (k). Metode ini bekerja dengan mengklasifikasikan suatu data yang belum diketahui kelasnya ke dalam kelas yang paling dominan di antara sejumlah tetangga terdekatnya. Kedekatan tersebut ditentukan melalui pengukuran jarak menggunakan matrik tertentu terhadap data latih (*training*) yang tersedia.

Algoritma *K-Nearest Neighbor* (KNN) termasuk dalam metode pembelajaran berbasis kedekatan objek, dimana proses klasifikasi dilakukan dengan mempertimbangkan objek-objek yang memiliki jarak paling dekat terhadap objek yang akan dianalisis. Prinsip kerja algoritma ini adalah menghitung jarak antara data yang akan dievaluasi dengan k tetangga (*neighbor*) terdekat pada data latih (data *training*), kemudian menentukan kelas berdasarkan mayoritas kelas dari tetangga tersebut. Dalam penerapannya, prediksi menggunakan metode *K-Nearest Neighbor* (KNN) mengacu pada pola atau kecenderungan yang muncul dari data yang diamati secara berulang dalam periode waktu tertentu. Dengan demikian, metode ini memanfaatkan kedekatan karakteristik antar data untuk menghasilkan keputusan klasifikasi atau prediksi.

Untuk mempermudah pemahaman mengenai konsep kerja metode *K-Nearest Neighbor* (KNN), dapat diperhatikan ilustrasi gambar dibawah ini [24]:



Gambar 4. *K-Nearest Neighbor* (KNN)

Gambar 4 di atas menggambarkan sebuah dataset yang terdiri atas dua kelas, yaitu kelas positif yang direpresentasikan oleh simbol segitiga berwarna hijau dan kelas negatif yang direpresentasikan oleh simbol lingkaran berwarna kuning. Dalam ilustrasi tersebut, bintang berwarna merah merupakan data uji yang akan ditentukan kelasnya menggunakan algoritma *K-Nearest Neighbor* (KNN). Pada gambar tersebut juga ditunjukkan penggunaan dua nilai k , yaitu $k = 3$ dan $k = 6$.

Algoritma *K-Nearest Neighbor* (KNN) bekerja dengan cara mengidentifikasi sejumlah tetangga terdekat dari data uji, kemudian menentukan kelas data tersebut berdasarkan mayoritas kelas dari tetangga terdekat yang ditentukan. Ketika nilai $k = 3$, maka jumlah tetangga terdekat yang dipertimbangkan adalah tiga data. Berdasarkan ilustrasi pada gambar, dua dari tiga tetangga terdekat termasuk dalam kelas negatif, sedangkan satu tetangga lainnya termasuk dalam kelas positif. Karena kelas yang paling dominan adalah kelas negatif, maka data uji yang direpresentasikan oleh bintang berwarna merah untuk $k = 3$ akan dikalsifikasikan/diprediksi sebagai kelas negatif.

Sementara itu, ketika $k = 6$, jumlah tetangga terdekat yang dipertimbangkan adalah enam data. Berdasarkan gambar tersebut, empat dari enam tetangga terdekat termasuk dalam kelas positif, sedangkan dua tetangga lainnya termasuk dalam kelas negatif. Karena mayoritas tetangga terdekat berasal dari kelas positif, maka data uji yang direpresentasikan oleh bintang berwarna merah untuk $k = 6$ akan dikalsifikasikan/diprediksi sebagai kelas positif.

Hubungan antara prediksi dan metode *K-Nearest Neighbor* (KNN) dapat dijelaskan dari peran KNN sebagai salah satu teknik yang digunakan untuk menghasilkan prediksi berdasarkan data yang tersedia. Metode *K-Nearest Neighbor* (KNN) melakukan prediksi dengan cara membandingkan data yang akan diprediksi dengan sejumlah data lain yang memiliki karakteristik paling mirip atau paling dekat. Secara umum, dapat disimpulkan bahwa prediksi merupakan tujuan atau hasil yang diperoleh, sedangkan KNN merupakan metode atau pendekatan yang digunakan untuk menghasilkan prediksi tersebut.

Hubungan prediksi dan klasifikasi pada metode *K-Nearest Neighbor* (KNN) sangat erat karena klasifikasi adalah salah satu bentuk prediksi yang dilakukan oleh KNN. Pada metode *K-Nearest Neighbor* (KNN), prediksi adalah proses utama sedangkan klasifikasi adalah bentuk prediksi yang menentukan label kelas berdasarkan mayoritas tetangga terdekat. Dengan demikian, klasifikasi merupakan salah satu hasil dari proses prediksi yang dilakukan oleh KNN.

Mencari tetangga atau jarak pada algoritma *K-Nearest Neighbor* (KNN) menggunakan jarak *Euclidean* yang didefinisikan sebagai berikut [26]:

$$D_{i,j} = \sqrt{(X_1^{latih} - X_1^{uji})^2 + \dots + (X_D^{latih} - X_D^{uji})^2}$$

$$= \sqrt{\sum_{i,j=1}^D (X_i^{latih} - X_j^{uji})^2}$$

Keterangan:

$D_{i,j}$: Jarak antar objek (jarak *euclidean*).

X_i^{latih} : Nilai pada data *training*.

X_j^{uji} : Nilai pada data *testing*.

D : Banyaknya data.

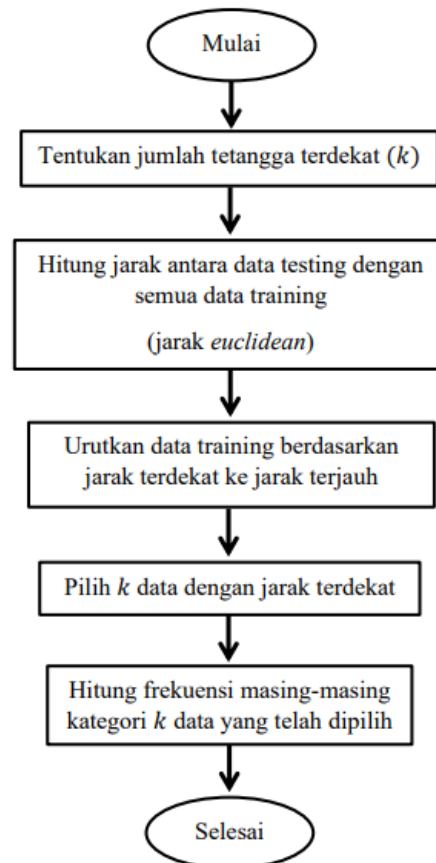
i, j : 1,2,3, ..., n .

Langkah-langkah dalam penerapan algoritma *K-Nearest Neighbor* (KNN) adalah sebagai berikut [26]:

- 1) Menentukan jumlah tetangga terdekat (k) yang akan dipilih atau digunakan.
- 2) Menghitung jarak antara data *testing* dengan seluruh data *training*. Pengukuran jarak yang digunakan adalah menggunakan jarak *Euclidean*.
- 3) Mengurutkan data *training* berdasarkan nilai jarak yang telah dihitung, dimulai dari yang dekat hingga yang paling jauh terhadap data *testing*.

- 4) Memilih sejumlah k data dengan jarak terdekat sebagai tetangga yang akan dijadikan dasar dalam proses klasifikasi/prediksi.
- 5) Menghitung frekuensi atau jumlah kemunculan masing-masing kategori atau kelas k data terpilih, untuk menentukan kelas mayoritas sebagai hasil klasifikasi/prediksi.

Adapun *flowchart* algoritma *K-Nearest Neighbor* (KNN), yaitu sebagai berikut:



Gambar 5. *Flowchart* Algoritma *K-Nearest Neighbor* (KNN)

Berikut ini merupakan contoh penerapan langkah-langkah dalam algoritma *K-Nearest Neighbor* (KNN):

- a) Tentukan jumlah tetangga terdekat (k) yang akan dipilih.
Tahap awal dalam algoritma KNN adalah menentukan jumlah tetangga terdekat. Penentuan nilai k tidak memiliki standar yang baku [8]. Pada penelitian ini nilai k ditentukan dengan menggunakan rumus berikut [12]:
$$k = \sqrt{n}$$
- b) Hitung jarak data *testing* dengan semua data *training*.
Pada tahap selanjutnya dilakukan perhitungan jarak, metode jarak *Euclidean* digunakan untuk menghitung jarak masing-masing

objek (data *testing*) terhadap data latih (*training*). Rumus jarak *Euclidean* yaitu sebagai berikut:

$$D_{i,j} = \sqrt{(X_1^{latih} - X_1^{uji})^2 + \dots + (X_D^{latih} - X_D^{uji})^2}$$

Perhitungan jarak *Euclidean* dengan contoh data *training* yang dapat dilihat pada Tabel 3 dan contoh data *testing* yang dapat dilihat pada Tabel 4 berikut ini:

Tabel 3. Contoh Data *Training*

Tanggal	Suhu	Kelembapan	Tekanan Udara	Kecepatan Angin	Intensitas Penyinaran Matahari	Curah Hujan	Kategori
01/07/2022	28	78	1008,8	8	9,8	0,0	Tidak Hujan
02/07/2022	26,8	89	1008,3	6	9,0	0,0	Tidak Hujan
03/07/2022	27,9	86	1007,6	3	2,8	28,8	Hujan Sedang
04/07/2022	28,5	85	1007,9	2	7,5	0,0	Tidak Hujan
05/07/2022	28,3	86	1009,1	3	7,6	0,0	Tidak Hujan
06/07/2022	27,5	83	1010,7	4	4,0	1,5	Tidak Hujan
07/07/2022	27,7	80	1010,8	4	6,8	0,0	Tidak Hujan
08/07/2022	26,8	78	1010,1	4	8,8	0,0	Tidak Hujan
09/07/2022	26,5	79	1009,8	5	8,0	1,3	Tidak Hujan

Tabel 4. Contoh Data *Testing*

Tanggal	Suhu	Kelembapan	Tekanan Udara	Kecepatan Angin	Intensitas Penyinaran Matahari	Curah Hujan	Kategori
10/07/2022	26,5	84	1009,8	4	7,5	4,6	?

Hasil perhitungan jarak antara data *training* terhadap data *testing* dengan ukuran jarak *euclidean* dapat dilihat pada Tabel 5 berikut ini:

Tabel 5. Contoh Perhitungan Jarak *Euclidean*

Tanggal	Jarak <i>Euclidean</i>	Keterangan	
		Curah Hujan	Kategori
01/07/2022	7,7845	0,0	Tidak Hujan
02/07/2022	5,8066	0,0	Tidak Hujan
03/07/2022	5,8247	28,8	Hujan Sedang
04/07/2022	3,5803	0,0	Tidak Hujan
05/07/2022	2,9718	0,0	Tidak Hujan
06/07/2022	3,8614	1,5	Tidak Hujan
07/07/2022	4,3378	0,0	Tidak Hujan
08/07/2022	6,1527	0,0	Tidak Hujan
09/07/2022	5,1235	1,3	Tidak Hujan

- c) Urutkan data *training* berdasarkan jarak dari yang terdekat ke yang terjauh.

Hasil perhitungan jarak yang sebelumnya sudah dilakukan kemudian diurutkan berdasarkan jarak *Euclidean* terdekat ke terjauh. Urutan perhitungan jarak dapat dilihat pada Tabel 6 berikut ini:

Tabel 6. Contoh Urutan Perhitungan Jarak

Tanggal	Jarak <i>Euclidean</i>	Kategori
05/07/2022	2,9718	Tidak Hujan
04/07/2022	3,5803	Tidak Hujan
06/07/2022	3,8614	Tidak Hujan
07/07/2022	4,3378	Tidak Hujan
09/07/2022	5,1235	Tidak Hujan
02/07/2022	5,8066	Tidak Hujan
03/07/2022	5,8247	Hujan Sedang
08/07/2022	6,1527	Tidak Hujan
01/07/2022	7,7845	Tidak Hujan

- d) Pilih k data dengan yang terdekat.

Pada contoh kasus ini terdapat data sebanyak 9 data. Berdasarkan rumus $k = \sqrt{n}$ maka nilai $k = \sqrt{9} = 3$. Sehingga banyaknya tetangga terdekat yang diambil berdasarkan k dapat dilihat pada Tabel 7 berikut ini:

Tabel 7. Contoh Banyaknya Tetangga Terdekat Berdasarkan k

Tanggal	Jarak <i>Euclidean</i>	Kategori
05/07/2022	2,9718	Tidak Hujan
04/07/2022	3,5803	Tidak Hujan
06/07/2022	3,8614	Tidak Hujan

- e) Hitung frekuensi masing-masing kategori k data yang telah dipilih

Berdasarkan k yang telah ditentukan terdiri dari kategori “Tidak Hujan”, “Tidak Hujan”, dan “Tidak Hujan”. Kategori yang paling banyak muncul (mayoritas) yaitu “Tidak Hujan”. Maka Prediksi data baru untuk tanggal 10 Juli 2022 adalah “Tidak Hujan”.

Metode *K-Nearest Neighbor* (KNN) memiliki sejumlah kelebihan dan kekurangannya dalam penerapannya. Beberapa kelebihan dan kekurangan metode tersebut dapat dijelaskan sebagai berikut [9]:

1) Kelebihan Metode *K-Nearest Neighbor* (KNN)

- a) Metode KNN relatif mudah untuk diimplementasikan dalam proses analisis data.
- b) Metode ini tidak memerlukan proses pembangunan model yang kompleks, penentuan parameter yang banyak, maupun asumsi tambahan terhadap distribusi data.
- c) KNN memiliki tingkat ketangguhan yang cukup baik terhadap data latih (*training*) yang mengandung *noise*.
- d) Metode ini dapat bekerja secara efektif terutama ketika jumlah data latih yang tersedia relatif besar.

2) Kekurangan Metode *K-Nearest Neighbor* (KNN)

- a) Metode *K-Nearest Neighbor* (KNN) memerlukan penentuan nilai parameter k yang tepat, yaitu jumlah tetangga terdekat yang digunakan dalam proses klasifikasi/prediksi, karena pemilihan nilai k sangat mempengaruhi hasil yang diperoleh.
- b) Proses pembelajaran dalam KNN didasarkan pada perhitungan jarak antar data. Oleh karena itu, metode ini memerlukan biaya komputasi yang relatif tinggi, terutama ketika harus menghitung jarak antara setiap data uji (*testing*) dengan seluruh data (*training*) menggunakan metode seperti *Euclidean Distance*.

4. Python

Python merupakan salah satu bahasa pemrograman komputer yang paling populer di dunia. Kugunaan Python salah satunya dapat digunakan untuk menganalisis data statistik. Python merupakan salah satu bahasa pemrograman yang populer untuk analisis data karena dukungan yang luas untuk pustaka analisis data seperti *pandas*, *numpy*, dan *matplotlib* [27]. Bahasa pemrograman Python selain diakses langsung menggunakan *software* Python dapat juga diakses melalui Google Coleb. Dengan adanya Google Coleb, bahasa pemrograman Python dapat digunakan dengan lebih praktis dan mudah tanpa perlu menginstall *software* Python. Berikut tampilan bahasa pemrograman Python melalui Google Coleb:

algoritma knn python colab - X | colab google - X | Percobaan K-NN Prediksi Curah - X | Algoritma K-Nearest Neighbor - X | +

colab.research.google.com/drive/1W42ByGJx8KjM6QyGpsu-dA0gEeAmGVm#scrollTo=0YW6s3PY_eel

Percobaan K-NN Prediksi Curah Hujan.ipynb ☆

File Edit Lihat Sisipkan Runtime Fitur Bantuan

Perintah + Kode + Teks ▶ Jalankan semua

Sambungkan kembali

Program Python untuk Prediksi Curah Hujan Harian dengan Metode KNN

```
[ ] import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns

[ ] from google.colab import drive
drive.mount('/content/drive')

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).

[ ] df = pd.read_excel("/content/drive/MyDrive/Data Penelitian Tugas Akhir Percobaan1.xlsx")
df
```

	Tanggal	Kategori	Curah Hujan	Suhu Rata-Rata	Kelambapan Rata-Rata	Tekanan Udara Rata-Rata	Kecepatan Angin Rata-Rata	Intensitas Penyinaran Matahari
0	2018-01-01	Sangat Lebat	133.5	28.0	79	1006.9	3	10.2
1	2018-01-02	Berawan	0.0	26.9	85	1007.2	2	10.8

Variabel Terminal

Type here to search

11:08 04/11/2025

Gambar 6. Tampilan Google Coleb untuk bahasa pemrograman Python