

BAB III

METODE PENELITIAN

A. Jenis Penelitian

Jenis penelitian ini adalah penelitian kuantitatif dengan pendekatan prediktif. Penelitian kuantitatif dipilih karena data yang digunakan bersifat numerik dan dianalisis secara statistik menggunakan model matematis. Pendekatan prediktif digunakan untuk meramalkan harga komoditas bawang merah berdasarkan pola historis dan pengaruh dari 6 variabel bebas, yaitu harga historis bawang merah, curah hujan (mm), suhu rata-rata ($^{\circ}\text{C}$), kelembapan rata-rata (%), volume panen bawang merah Kabupaten Solok (ton), dan hari besar keagamaan.

B. Prosedur dan Teknik Analisis Data

Penelitian ini menggunakan metode *Cross-Industry Standard Process for Data Mining* (CRISP-DM) sebagai pendekatan utama dalam pengolahan data dan pembangunan model. CRISP-DM yang memiliki 6 tahapan yaitu: pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi dan tahapan terakhir penyebaran hasil *data mining* (Sastya & Isna Nugraha, 2023). Berikut 6 tahapan tersebut:

1. *Business Understanding* (Pemahaman Bisnis)

Tahap awal bertujuan untuk memahami permasalahan penelitian dan menetapkan tujuan yang ingin dicapai. Oleh karena itu, peneliti melakukan riset langsung di Dinas Pangan Kota Padang dan Dinas Pertanian Kabupaten Solok untuk merumuskan masalah utama, mengidentifikasi kebutuhan data, serta menentukan kriteria keberhasilan yang relevan. Proses ini, dilakukan identifikasi terhadap permasalahan yang sering dihadapi oleh masyarakat, permasalahan ini muncul pada momen-momen tertentu seperti hari besar keagamaan, kondisi cuaca ekstrem, dan volume panen bawang merah yang berdampak pada ketidakstabilan. Ketidakpastian harga tersebut dapat merugikan banyak pihak, termasuk konsumen, pedagang, dan pemangku kebijakan karena tidak adanya sistem prediksi yang akurat. Setelah itu, dilakukan pemahaman teori yang cocok terkait penggunaan K-NN dalam

prediksi harga komoditas bawang merah. Peneliti mengacu pada berbagai sumber seperti jurnal ilmiah dan laporan penelitian untuk membangun landasan teori yang kuat. Tujuannya adalah untuk mengetahui perkembangan terkini dalam prediksi harga komoditas bawang merah dan menentukan algoritma dan pengujian yang relevan. Berdasarkan hasil identifikasi ini, solusi yang akan ditawarkan adalah penerapan metode *K-Nearest Neighbors* untuk prediksi harga komoditas bawang merah di Kota Padang.

2. *Data Understanding* (Pemahaman Data)

Aktivitas yang dilakukan meliputi pengumpulan, pengidentifikasi, dan pemahaman data yang akan digunakan pada penelitian ini. Proses yang dilakukan meliputi:

a. Observasi

Observasi dilakukan dengan mengunjungi langsung kantor Dinas Pangan Kota Padang untuk memperoleh data historis harga komoditas bawang merah di Kota Padang dan juga melakukan observasi ke Dinas Pertanian Kabupaten Solok untuk memperoleh data volume panen bawang merah tahun 2024 dan 2025.

b. Wawancara

Wawancara dilakukan dengan pihak Dinas Pangan Kota Padang dan Dinas Pertanian Kabupaten Solok untuk menggali informasi mengenai harga historis bawang merah di Kota Padang dan volume panen bawang merah di Kabupaten Solok. Melalui wawancara, peneliti dapat memperoleh penjelasan tentang faktor-faktor yang dianggap berpengaruh terhadap harga bawang merah di Kota Padang seperti, naik turunnya harga bawang merah pada momen-momen tertentu seperti hari besar keagamaan, kondisi cuaca dan volume panen bawang merah. Informasi ini akan membantu peneliti untuk menentukan variabel-variabel independen yang relevan dan digunakan dalam prediksi harga bawang merah. Wawancara juga memberikan wawasan kualitatif yang tidak selalu tersedia dalam data kuantitatif.

c. Studi literatur

Studi literatur dilakukan dengan membaca dan menganalisis berbagai sumber informasi, seperti artikel jurnal dan penelitian terdahulu yang relevan dengan penggunaan *K-Nearest Neighbors*. Peneliti juga mempelajari studi-studi sejenis yang telah menggunakan metode ini untuk memprediksi harga komoditas bawang merah dalam konteks yang serupa. Studi literatur ini bertujuan untuk memahami teori dasar *K-Nearest Neighbors*, memilih metode yang sesuai, dan mendapatkan gambaran tentang variabel-variabel umum yang sering digunakan dalam penelitian terkait. Selain itu, dengan mempelajari literatur, peneliti dapat mengidentifikasi celah penelitian sebelumnya yang akan menjadi dasar pembenaran untuk melakukan penelitian ini.

3. Persiapan Data (*Data Preparation*)

Pada tahap ini, data yang telah dikumpulkan akan diproses agar siap digunakan dalam pemodelan. Berikut prosesnya:

a. Seleksi data (*Data Selection*)

Pada tahap ini, peneliti mengidentifikasi dan mengekstraksi atribut-atribut penting dari kumpulan data mentah yang tersedia. Pemilihan data dilakukan berdasarkan relevansi terhadap permasalahan yang ingin diselesaikan, guna menghindari penggunaan atribut yang tidak berkaitan atau bersifat *redundant* yang dapat mengganggu proses analisis. Dengan melakukan seleksi data yang tepat, volume data yang digunakan menjadi lebih efisien, kualitas data meningkat, dan performa model yang dihasilkan pun lebih optimal. Seleksi data juga membantu mempercepat proses pemodelan dan mengurangi risiko *overfitting* akibat data yang tidak relevan.

b. Pembersihan Data (*Data Cleaning*)

Pada tahap ini, dilakukan identifikasi dan penanganan terhadap nilai-nilai yang hilang, data duplikat, serta nilai-nilai ekstrem atau *outlier* yang dapat memengaruhi hasil analisis. Proses ini juga mencakup pemeriksaan kesesuaian format data dan konsistensi antar atribut, agar data menjadi seragam dan dapat diolah lebih lanjut. Pembersihan data sangat penting

untuk meningkatkan kualitas dataset, mengurangi *noise*, dan memastikan bahwa model yang dibangun nantinya dapat memberikan hasil yang akurat dan dapat diandalkan. Tanpa proses pembersihan yang tepat, risiko terjadinya kesalahan prediksi atau interpretasi meningkat, karena model belajar dari informasi yang tidak benar atau tidak relevan.

c. Konstruksi Data (*Data Construction*)

Langkah ini dilakukan untuk memperkaya informasi dalam dataset, sehingga analisis atau pemodelan yang dilakukan dapat lebih akurat dan relevan. Konstruksi data dapat melibatkan berbagai teknik, seperti penggabungan beberapa atribut menjadi satu, pembuatan indikator berdasarkan kondisi tertentu, atau transformasi data temporal menjadi fitur yang lebih informatif. Dengan membangun fitur-fitur baru yang lebih representatif, proses ini membantu model dalam mengenali pola yang lebih kompleks dan meningkatkan kemampuan prediktifnya.

d. Integrasi Data (*Data Integration*)

Dalam penelitian ini, proses integrasi data dilakukan untuk menggabungkan berbagai sumber informasi yang beragam guna membentuk satu kesatuan dataset yang komprehensif. Data yang digunakan meliputi harga historis bawang merah, curah hujan harian, suhu rata-rata, kelembapan udara rata-rata, volume panen bawang merah dari daerah penghasil utama, serta daftar hari besar keagamaan. Keenam jenis data ini berasal dari lembaga yang berbeda, seperti Dinas Pangan Kota Padang, NASA POWER, Dinas Pertanian Kabupaten Solok, dan sumber kalender nasional, yang masing-masing memiliki format dan struktur data tersendiri. Oleh karena itu, tahap integrasi tidak hanya mencakup pengumpulan data, tetapi juga melibatkan penyelarasan waktu agar data antar variabel dapat dianalisis dalam kerangka waktu yang konsisten, misalnya dengan menyeragamkan semua data menjadi format harian.

Proses ini juga mencakup penyesuaian terhadap struktur data, seperti penyamaan nama kolom, konversi format tanggal, dan penanganan nilai kosong (*missing values*) yang muncul akibat perbedaan cakupan data antar

sumber. Hari besar keagamaan diberi penanda biner (1 dan 0) untuk menunjukkan kehadiran atau ketidakhadiran peristiwa tersebut dalam suatu tanggal. Penyesuaian skala nilai antar variabel juga dilakukan agar analisis prediktif tidak bias terhadap variabel tertentu. Dengan pendekatan integrasi ini, masing-masing jenis data tidak hanya berdiri sendiri, tetapi saling melengkapi satu sama lain untuk membentuk dataset yang utuh dan representatif. Dataset terpadu ini kemudian digunakan sebagai fondasi dalam membangun model prediksi harga bawang merah berbasis data cuaca dan variabel eksternal lainnya.

e. Transformasi Data (*Data Transformation / Formatting*)

Transformasi data untuk memastikan setiap variabel berada dalam bentuk yang sesuai dan optimal untuk analisis prediktif. Langkah awal dalam transformasi adalah melakukan normalisasi terhadap variabel numerik seperti curah hujan, suhu, dan kelembapan udara. Normalisasi ini bertujuan untuk menyetarakan skala antar variabel agar tidak terjadi dominasi salah satu fitur dalam perhitungan jarak atau bobot pada algoritma pembelajaran mesin, khususnya yang sensitif terhadap perbedaan skala. Selain itu, data kategorik seperti hari besar keagamaan diubah ke dalam format numerik menggunakan pendekatan representasi biner, yakni dengan memberikan nilai 1 pada tanggal yang termasuk hari besar dan 0 pada tanggal biasa, sehingga variabel ini dapat diolah dalam model statistik maupun algoritmik.

Transformasi juga mencakup rekonstruksi elemen waktu, seperti konversi tanggal ke dalam bentuk bulan atau kuartal, yang bertujuan untuk mengungkap pola musiman atau tren jangka menengah yang tidak terlihat pada skala harian. Di samping itu, dilakukan penyesuaian terhadap tipe data agar seluruh kolom memiliki konsistensi format dan dapat dikenali secara tepat oleh sistem analisis. Keseluruhan proses transformasi ini dirancang untuk meningkatkan kualitas dan kesiapan data, sehingga seluruh variabel dalam dataset dapat saling berinteraksi secara efektif dan menghasilkan model prediksi harga komoditas bawang merah yang lebih akurat dan andal.

4. Pemodelan (*Modeling*)

Dalam konteks penelitian ini, model prediktif dibangun untuk memprediksi harga bawang merah dengan mempertimbangkan sejumlah variabel independen, yakni data historis curah hujan, suhu rata-rata, kelembapan udara rata-rata, volume panen bawang merah dari daerah penghasil, serta penanda hari besar keagamaan. Seluruh variabel ini dipilih karena secara teoritis dan empiris telah menunjukkan keterkaitan yang signifikan terhadap dinamika harga, baik melalui mekanisme permintaan, maupun pola musiman yang berkembang di pasar.

Proses dalam Pemodelan:

a. Pemilihan Teknik Pemodelan

Dalam penelitian ini, teknik pemodelan yang dipilih adalah *K-Nearest Neighbors Regressor* (K-NN Regressor). Metode ini dipilih karena sesuai dengan karakteristik data dan tujuan utama penelitian, yaitu melakukan prediksi nilai harga bawang merah berdasarkan 6 variabel bebas, yaitu harga historis bawang merah, curah hujan (mm), suhu rata-rata (°C), kelembapan rata-rata (%), volume panen bawang merah Kabupaten Solok (ton), dan hari besar keagamaan.

b. Penyesuaian Format Data

Dalam penelitian ini, dilakukan penyesuaian format data melalui proses normalisasi skala menggunakan teknik seperti *min-max scaling*. Hal ini penting karena algoritma K-Nearest Neighbor sangat sensitif terhadap perbedaan skala antar fitur. Dengan menyamakan skala, setiap fitur memiliki kontribusi yang seimbang dalam perhitungan jarak menggunakan *euclidean distance*, sehingga tidak ada fitur yang secara tidak proporsional mempengaruhi hasil prediksi. Penyesuaian ini bertujuan untuk meningkatkan akurasi dan kestabilan model dalam memetakan pola historis harga bawang merah.

c. Penentuan Parameter Model

Sebelum melatih model akhir, dilakukan proses penentuan parameter optimal berupa jumlah tetangga terdekat (K) pada algoritma *K-Nearest Neighbors* (K-NN). Nilai K yang diuji berada pada rentang 1 hingga 20. Evaluasi setiap nilai K dilakukan menggunakan metode *k-fold cross validation* (sebanyak 5 *fold*), dengan menggunakan metrik evaluasi regresi berupa RMSE, MAE, dan R^2 . Nilai K terbaik dipilih berdasarkan kombinasi nilai RMSE dan MAE terendah serta R^2 tertinggi yang konsisten pada seluruh *fold*. Berdasarkan evaluasi, nilai $K = 4$ dipilih sebagai parameter terbaik karena memberikan nilai RMSE dan MAE terendah serta nilai R^2 tertinggi secara konsisten pada seluruh *fold*.

d. Pelatihan Model (*Training*)

Pada tahap ini dilakukan proses pelatihan terhadap model *K-Nearest Neighbors* (K-NN). Regressor menggunakan data historis harga bawang merah. Data yang telah melalui proses pra pemrosesan dan transformasi digunakan sepenuhnya dalam proses pelatihan dan evaluasi model dengan menerapkan metode *k-fold cross validation* sebanyak 5 *fold*. Model K-NN dilatih dengan memanfaatkan fitur-fitur independen yang meliputi curah hujan, suhu rata-rata, kelembapan rata-rata, jumlah panen bawang merah di wilayah produksi, hari besar nasional, dan tanggal.

Sementara itu, target (variabel dependen) yang diprediksi adalah harga bawang merah. Karena K-NN merupakan algoritma *lazy learning*, model tidak membentuk persamaan atau fungsi prediktif secara eksplisit, melainkan menyimpan seluruh data dan melakukan pencarian terhadap tetangga terdekat saat proses prediksi berlangsung. Melalui tahap pelatihan ini, sistem dibangun untuk mengenali pola keterkaitan antara kondisi cuaca, waktu panen, momen hari besar, dan dinamika harga bawang merah, sehingga mampu menghasilkan prediksi yang akurat saat diuji melalui validasi silang.

5. Pengujian (*Evaluation*)

Tahap pengujian dilakukan dengan menerapkan metode *k-fold cross validation* sebanyak 5 *fold*, sehingga seluruh data digunakan secara bergantian sebagai data uji. Evaluasi performa model dilakukan menggunakan tiga metrik regresi, yaitu *Root Mean Square Error (RMSE)*, *Mean Absolute Error (MAE)* dan *R-Squared (R^2)*. Hasil dari setiap *fold* dirata-ratakan untuk memperoleh gambaran menyeluruh mengenai akurasi model dalam memprediksi harga bawang merah. Dengan pendekatan ini, proses pengujian tidak hanya mengandalkan satu subset data uji, melainkan dilakukan secara menyeluruh dan sistematis untuk memastikan kestabilan dan generalisasi model.

6. Penyebaran (*Deployment*)

Tahap *deployment* merupakan langkah terakhir dalam proses CRISP-DM yang berfokus pada penerapan model yang telah dikembangkan ke dalam lingkungan nyata atau operasional untuk mendukung proses pengambilan keputusan. Dalam penelitian ini, proses *deployment* dilakukan menggunakan Streamlit, yaitu sebuah *framework* berbasis *python* yang memungkinkan pembuatan aplikasi web interaktif secara cepat dan efisien. Melalui Streamlit, model prediksi harga bawang merah dapat diakses melalui antarmuka yang sederhana dan mudah digunakan, sehingga pengguna seperti dinas terkait atau pelaku pasar dapat memperoleh hasil prediksi secara *real-time*. Dengan pendekatan ini, model yang dibangun tidak hanya bersifat teoritis, tetapi juga dapat langsung diimplementasikan secara praktis dan berkelanjutan.

C. Setting Penelitian

1. Lokasi penelitian

Penelitian ini dilakukan di Kota Padang.

2. Waktu penelitian

Waktu penelitian ini selama lima bulan, dari bulan Maret sampai Juli 2025.

Waktu pengerjaan penelitian yang dilakukan peneliti dapat dilihat pada tabel dibawah ini:

Tabel 3.1 Waktu penelitian

No	Kegiatan	Bulan																			
		1				2				3				4				5			
1	Identifikasi masalah																				
2	Pengumpulan data																				
3	Pengolahan data																				
4	Analisis data																				
5	Pembuatan laporan																				

D. Bahan Penelitian

Dataset penelitian ini terdiri dari 256 *record* harian (1 Agustus 2024 – 13 April 2025) dengan 6 variabel bebas, yaitu harga historis bawang merah, curah hujan, suhu rata-rata, kelembapan rata-rata, volume panen dan hari besar keagamaan.

1. Data harga historis komoditas bawang merah diperoleh dari Dinas Pangan Kota Padang.
2. Data curah hujan, temperatur suhu, dan kelembapan diperoleh dari situs NASA POWER (*Prediction of Worldwide Energy Resources*).
3. Data hari besar keagamaan diperoleh dari arsip kalender nasional.
4. Data volume panen bawang merah diperoleh dari Dinas Pertanian Kabupaten Solok.

E. Alat Penelitian

Alat dan instrumen yang digunakan untuk menyelesaikan penelitian ini terdiri dari perangkat lunak dan perangkat keras, yaitu:

1. Perangkat Lunak
 - a. Sistem Operasi Windows 11
 - b. Visual Studio Code, Browser (google chrome) dan Streamlit
 - c. Microsoft Word dan Microsoft Excel
2. Perangkat Keras
 - a. *Processor* AMD Ryzen 5 5500u *System type* 64-bit
 - b. RAM (*Random Access Memory*) 8 GB