
PRA-SKRIPSI

**PERBANDINGAN KINERJA MODEL SVR DAN
XGBOOST DALAM PREDIKSI HARGA BERAS
BERDASARKAN KUALITAS**

OKTAVIA NUR KHASANAH
NPM 22081010063

DOSEN PEMBIMBING

-
-

**KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA
SURABAYA
2025**

Halaman ini sengaja dikosongkan

BAB I

PENDAHULUAN

1.1. Latar Belakang

Indonesia dikenal sebagai negara agraris dan sebagian besar penduduknya bekerja pada bidang pertanian (Ruvananda & Taufiq, 2022). Sektor pertanian memegang peranan penting dalam struktur perekonomian Indonesia, khususnya dalam menyediakan lapangan kerja, memenuhi kebutuhan pangan, dan menjadi sumber pendapatan masyarakat pedesaan (Saputra & Kurniati, 2025). Di antara berbagai komoditas pertanian, padi merupakan salah satu komoditas utama yang memiliki kontribusi besar terhadap pemenuhan kebutuhan pangan masyarakat Indonesia. Berdasarkan data dari Food and Agriculture Organization (FAO), produksi beras Indonesia pada musim tanam 2024/2025 diperkirakan mencapai 34,6 juta ton. Hal ini menunjukkan bahwa padi memiliki posisi yang sangat penting dalam sistem pertanian nasional dan menjadi faktor kunci dalam upaya menjaga stabilitas pangan di Indonesia.

Beras merupakan makanan pokok yang paling banyak dikonsumsi oleh masyarakat Indonesia (Pragana et al., 2023). Selain sebagai sumber pangan pokok, beras juga menjadi sumber utama penghasilan bagi petani serta memenuhi kebutuhan hidup jutaan penduduk setiap hari. Peran beras sangat luas, meliputi aspek ekonomi, tenaga kerja, sosial, budaya, hingga politik. Oleh karena itu, kestabilan harga beras memiliki pengaruh besar terhadap kesejahteraan masyarakat, khususnya bagi kelompok berpenghasilan rendah, serta terhadap kestabilan ekonomi nasional secara keseluruhan. Ketimpangan dalam sistem distribusi maupun produksi sering kali menimbulkan ketidakseimbangan antara permintaan dan penawaran (Zainuddin & Nuryadin, 2024), yang pada akhirnya berdampak pada fluktuasi harga beras.

Permintaan dan penawaran beras di Indonesia sangat dipengaruhi oleh berbagai faktor, seperti musim tanam, tingkat produksi, biaya distribusi, serta kebijakan pemerintah terkait impor dan cadangan pangan nasional. Pada saat produksi melimpah, terutama setelah masa panen raya, penawaran beras cenderung meningkat sehingga harga dapat mengalami penurunan. Sebaliknya, ketika produksi menurun akibat faktor cuaca, gagal panen, atau keterlambatan distribusi, pasokan beras menjadi terbatas dan harga berpotensi naik. Kondisi ini menunjukkan bahwa keseimbangan antara

permintaan dan penawaran beras bersifat dinamis dan sangat sensitif terhadap perubahan kondisi ekonomi maupun lingkungan. Kondisi fluktuatif ini menunjukkan perlunya sistem yang mampu memprediksi pergerakan harga agar pemerintah dapat mengambil langkah antisipatif dalam menjaga kestabilan pasokan dan harga beras di pasar.

Prediksi harga beras menjadi salah satu upaya penting untuk mengantisipasi ketidakstabilan harga di masa depan. Dengan mengetahui pola perubahan harga berdasarkan data historis, pemerintah dan pelaku pasar dapat merencanakan strategi distribusi, stok, dan produksi secara lebih efisien. Dalam penelitian ini, data yang digunakan bersumber dari Badan Pusat Statistik (BPS) dalam rentang waktu 2020–2024, yang mencakup beberapa variabel seperti tahun, bulan, harga rata-rata, kualitas beras (premium, medium, dan rendah), serta jumlah produksi padi (dalam ton). Variabel-variabel tersebut digunakan untuk menganalisis hubungan antara kualitas dan produksi terhadap fluktuasi harga beras.

Dalam upaya meningkatkan akurasi prediksi harga, penerapan metode berbasis *machine learning* menjadi pilihan yang efektif. Dua model yang banyak digunakan untuk peramalan data non-linear adalah *Support Vector Regression (SVR)* dan *Extreme Gradient Boosting (XGBoost)*. Algoritma SVR merupakan metode yang sering digunakan untuk menyelesaikan permasalahan-permasalahan linear maupun nonlinear dalam kasus prediksi maupun klasifikasi (Ruliana et al., 2024). *Support Vector Regression* juga mampu memetakan hubungan kompleks antara variabel input dan output dengan generalisasi yang baik. Sedangkan XGBoost (Extreme Gradient Boosting) merupakan *tree boosting* yang dapat diskalakan yang telah terbukti sebagai metode *machine learning* yang sangat efektif dan luas digunakan untuk melakukan prediksi nilai dari suatu data (Rayadin et al., 2024). Model ini dikenal unggul dalam menangani data besar dan menghasilkan prediksi dengan tingkat kesalahan rendah. Dengan membandingkan performa kedua model tersebut, penelitian ini bertujuan untuk menghasilkan prediksi harga beras yang lebih akurat berdasarkan faktor kualitas dan jumlah produksi padi. Hasil prediksi ini diharapkan dapat membantu pemerintah, pelaku pasar, serta petani dalam merencanakan strategi produksi, distribusi, dan pengendalian harga yang lebih optimal. Selain itu, penelitian ini juga diharapkan memberikan kontribusi ilmiah dalam pengembangan model prediksi harga komoditas

pangan menggunakan pendekatan *machine learning*, sehingga dapat menjadi referensi bagi penelitian sejenis di masa mendatang.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah dalam penelitian ini Adalah sebagai berikut:

1. Bagaimana pengaruh kualitas beras dan jumlah produksi padi terhadap fluktuasi harga beras di Indonesia?
2. Bagaimana penerapan metode SVR dan XGBoost dalam memprediksi harga beras berdasarkan kualitas?
3. Model mana yang memberikan hasil prediksi paling akurat antara metode SVR dan XGBoost?

1.3. Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dijelaskan sebelumnya, tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut:

1. Menganalisis hubungan antara kualitas beras dan jumlah produksi padi terhadap perubahan harga beras di Indonesia.
2. Menerapkan model Support Vector Regression (SVR) dan Extreme Gradient Boosting (XGBoost) untuk melakukan prediksi harga beras berdasarkan kualitas dan jumlah produksi padi.
3. Membandingkan performa kedua model tersebut untuk menentukan model dengan akurasi terbaik dalam memprediksi harga beras.

1.4. Batasan Masalah

Agar penelitian ini lebih lebih terarah dan tidak menyimpang dari tujuan yang telah ditetapkan, maka perlu adanya batasan masalah. Adapun batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Data yang digunakan bersumber dari Badan Pusat Statistik (BPS) dalam rentang waktu tahun 2020–2024.
2. Variabel yang digunakan meliputi tahun, bulan, harga rata-rata beras, kualitas beras (premium, medium, rendah), dan jumlah produksi padi (dalam ton).

3. Model machine learning yang digunakan hanya terbatas pada Support Vector Regression (SVR) dan Extreme Gradient Boosting (XGBoost).
4. Penelitian difokuskan pada prediksi harga rata-rata beras di tingkat nasional, tidak mencakup perbandingan harga antar daerah atau provinsi.
5. Evaluasi model dilakukan menggunakan metrik seperti Mean Absolute Error (MAE), Root Mean Square Error (RMSE), dan R^2 Score.

1.5 Manfaat Penelitian

Penelitian yang dilakukan diharapkan mampu memberikan manfaat, yaitu:

1. Manfaat Akademis

- Menambah wawasan dan pengetahuan mengenai penerapan machine learning, khususnya model SVR dan XGBoost, dalam bidang pertanian.
- Menjadi referensi bagi penelitian selanjutnya yang berkaitan dengan prediksi harga komoditas pangan di Indonesia.

2. Manfaat Praktis

- Memberikan informasi bagi pemerintah dalam merumuskan kebijakan pengendalian harga dan cadangan pangan nasional.
- Membantu petani dan pelaku pasar dalam merencanakan strategi produksi dan distribusi berdasarkan prediksi harga yang lebih akurat.
- Menjadi bahan pertimbangan bagi lembaga ekonomi dan pelaku usaha dalam mengambil keputusan terkait pasokan dan investasi sektor beras.

Halaman ini sengaja dikosongkan

BAB II

TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Penelitian ini tidak terlepas dari acuan terhadap penelitian-penelitian sebelumnya. Penelitian acuan ini diharapkan dapat memberikan dasar teori, metode, serta referensi yang relevan dalam mendukung perancangan dan pelaksanaan penelitian ini, sehingga hasil yang diperoleh memiliki landasan ilmiah yang kuat. Melalui tinjauan pustaka terhadap penelitian terdahulu, penulis dapat memahami perkembangan penelitian sejenis, mengidentifikasi kelebihan dan kekurangan dari penelitian sebelumnya, serta menemukan *research gap* yang menjadi dasar pengembangan penelitian ini. Berikut adalah beberapa penelitian terdahulu yang relevan dengan topik yang akan diteliti:

1. Penelitian “Prediksi Harga Beras Berdasarkan Kualitas dengan Metode Long Short Term Memmory” oleh (Nafi’iyah & Wulandari, 2022).

Penelitian ini bertujuan untuk membangun model prediksi harga beras berdasarkan kualitas beras (premium, medium, dan luar kulit) menggunakan algoritma Long Short Term Memmory, salah satu metode *Neural Network* yang efektif untuk menganalisis data deret waktu (*time series*). Data yang digunakan dalam penelitian ini bersumber dari Badan Pusat Statistik (BPS) dalam rentan waktu 2019-2021, dengan total 36 data bulanan, terdiri dari 28 data latih dan 8 data uji. Model LSTM yang digunakan memiliki arsitektur dengan 4 neuron pada input layer, tiga hidden layer masing-masing 50 neuron, dan 1 neuron pada output layer.

Hasil pengujian menunjukkan bahwa model LSTM mampu memprediksi harga beras dengan tingkat kesalahan rata-rata yang relatif kecil, ditunjukkan oleh nilai Mean Absolute Error (MAE) dari harga beras premium, medium, dan luar kulit secara berurutan adalah 83.49, 89.6, 96.99. Hasil tersebut membuktikan bahwa LSTM efektif dalam memprediksi harga beras berdasarkan urutan waktu sebelumnya, karena mampu menangkap pola dan hubungan non-linear dari data historis harga beras.

2. Penelitian “Perbandingan Kinerja SVR dan XGBoost untuk Peramalan Emisi CO₂ Global berbasis Machine Learning” oleh (Rahman et al., 2025)

Penelitian ini bertujuan untuk membandingkan kinerja dua algoritma machine learning yaitu SVR dan XGBoost dalam meramalkan Emisi CO₂ berdasarkan data historis. Dataset yang digunakan mencakup berbagai variabel penting seperti konsumsi energi, Produk Domestik Bruto (PDB), dan populasi, yang diperoleh dari sumber terbuka seperti Our World in Data dan Global Carbon Project.

Model SVR dioptimalkan menggunakan Grid Search untuk menentukan kombinasi parameter terbaik (C, gamma, dan epsilon), sedangkan XGBoost digunakan sebagai pembanding karena kemampuannya dalam menangani hubungan non-linear dan interaksi antar fitur. Evaluasi model dilakukan dengan tiga metrik utama: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), dan Koefisien Determinasi (R²).

Hasil penelitian menunjukkan bahwa XGBoost secara signifikan lebih unggul dengan MAE sebesar 1745,70 dan RMSE sebesar 2663,18 serta nilai R² sebesar 0,93, dibandingkan SVR yang memiliki MAE sebesar 5476,54, RMSE 8153,37 dan R² sebesar 0,82. Visualisasi hasil prediksi juga mengindikasikan bahwa XGBoost lebih akurat dalam mengikuti pola fluktuasi data aktual, terutama dalam mendeteksi perubahan tajam. Temuan ini menyarankan bahwa XGBoost merupakan metode yang lebih tepat untuk digunakan dalam peramalan emisi CO₂ pada konteks global yang kompleks dan non-linear.

3. Penelitian “Evaluating the Performances of SVR and XGBoost for Short-Range Forecasting of Heatwaves Across Different Temperature Zones of India” oleh (Bhoopathi et al., 2024)

Penelitian ini bertujuan untuk membandingkan kinerja dua algoritma machine learning, yaitu Support Vector Regression (SVR) dan Extreme Gradient Boosting (XGBoost), dalam memprediksi suhu maksimum harian serta jumlah hari gelombang panas (heatwave days) di empat zona suhu yang berbeda di India berdasarkan data meteorologi selama periode 1991–2020, dengan sembilan variabel atmosfer utama seperti suhu udara, kelembaban, kecepatan angin, tekanan, dan radiasi matahari sebagai faktor prediksi.

Hasil pengujian menunjukkan bahwa SVR lebih unggul dalam memprediksi suhu jangka pendek (7 hari ke depan) karena kemampuannya

menangkap fluktuasi dan puncak suhu ekstrem dengan baik. Sementara itu, XGBoost menunjukkan performa lebih stabil dan akurat untuk prediksi jangka menengah (hingga 15 hari) karena memiliki kemampuan generalisasi yang lebih baik serta lebih tahan terhadap overfitting. Kedua model mengalami penurunan akurasi di wilayah bersuhu sedang hingga rendah, namun tetap menunjukkan hasil yang memuaskan pada daerah dengan suhu tinggi. Secara keseluruhan, penelitian ini menegaskan bahwa SVR dan XGBoost dapat menjadi alat prediksi yang efektif untuk peringatan dini gelombang panas, di mana SVR lebih cocok untuk peramalan jangka pendek, sedangkan XGBoost lebih sesuai untuk prediksi jangka panjang yang memerlukan stabilitas dan ketepatan yang tinggi.

4. Penelitian “Analisa Metode Backpropagation pada Prediksi Rata-rata Harga Beras Bulanan di Tingkat Penggilingan Menurut Kualitas” oleh (Pragana et al., 2023)

Penelitian ini untuk menganalisis dan menerapkan metode jaringan saraf tiruan (Artificial Neural Network) dengan algoritma Backpropagation dalam memprediksi rata-rata harga beras berdasarkan kualitasnya. Data yang digunakan berasal dari Badan Pusat Statistik (BPS) dengan periode tahun 2013 hingga 2022, yang mencakup harga beras di tingkat penggilingan menurut tiga kategori kualitas, yaitu premium, medium, dan luar kualitas. Model diuji menggunakan lima arsitektur jaringan yang berbeda, yaitu 15-15-1, 15-17-1, 15-19-1, 15-30-1, dan 15-31-1.

Hasil pengujian menunjukkan bahwa arsitektur terbaik adalah model 15-17-1, yang menghasilkan nilai Mean Squared Error (MSE) terkecil sebesar 0.00000105 dengan akurasi tertinggi mencapai 100% pada 5206 iterasi pelatihan. Hal ini menunjukkan bahwa metode backpropagation memiliki kemampuan yang sangat baik dalam mengenali pola data historis dan menghasilkan prediksi harga beras dengan tingkat kesalahan yang sangat rendah. Penelitian ini juga membuktikan bahwa JST dengan algoritma backpropagation dapat menjadi alat bantu yang andal untuk memperkirakan pergerakan harga beras di masa depan.

5. Penelitian “Perbandingan Akurasi Algoritma XGBoost dan SVR dalam Prediksi Harga Cryptocurrency” oleh (Fadhil, 2025)

Penelitian ini bertujuan untuk membandingkan efektivitas dua algoritma machine learning, yaitu Extreme Gradient Boosting (XGBoost) dan Support Vector Regression (SVR), dalam memprediksi harga berbagai mata uang kripto (cryptocurrency). Penelitian ini menggunakan data transaksi dari 10 jenis cryptocurrency yang dievaluasi dengan beberapa metrik akurasi, yaitu Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Squared Error (MSE), dan Mean Absolute Percentage Error (MAPE).

Hasil penelitian menunjukkan bahwa algoritma XGBoost memiliki kinerja yang lebih unggul dibandingkan SVR, ditandai dengan nilai MAPE yang jauh lebih rendah, berkisar antara 0–1% pada sebagian besar mata uang kripto yang diuji. Hal ini menunjukkan bahwa XGBoost mampu membaca pola pergerakan harga pasar dengan lebih baik dan memberikan prediksi yang lebih akurat. Sebaliknya, algoritma SVR menunjukkan performa yang bervariasi, di mana hasilnya cukup baik pada beberapa koin tetapi kurang stabil pada lainnya, sehingga dinilai kurang konsisten dalam memprediksi fluktuasi harga pasar kripto yang sangat dinamis.

2.2. Beras

Beras merupakan makanan pokok yang sangat penting di Indonesia dan memiliki peran penting dalam struktur pangan sebagai sumber gizi (Pragana et al., 2023). Selain berfungsi sebagai sumber energi dan nutrisi bagi sebagian besar penduduk, beras juga memiliki nilai sosial dan ekonomi yang tinggi karena menjadi komoditas strategis dalam menjaga stabilitas pangan nasional. Posisi beras yang sangat penting tersebut menjadikan beras sebagai komoditas dengan tingkat permintaan yang tinggi dan relatif stabil sepanjang tahun. Tingginya tingkat permintaan tersebut berpengaruh terhadap fluktuasi harga beras di pasaran, di mana salah satu faktor utama yang memengaruhi harga beras adalah kualitasnya.

Berdasarkan data dari Badan Pusat Statistik, kualitas beras dikategorikan menjadi tiga yaitu beras premium, beras medium, dan beras rendah. Kategori tersebut ditentukan didasarkan pada tingkat butir patah yang dihasilkan dari proses penggilingan, di mana beras premium memiliki kadar butir patah maksimum sebesar 10,00%, beras medium memiliki kadar butir patah antara 10,01% hingga 20,00%, dan beras rendah memiliki kadar butir patah antara 20,01% hingga 25,00%. Perbedaan tingkat butir patah tersebut mencerminkan mutu fisik beras yang berpengaruh

langsung terhadap nilai jualnya di pasaran. Semakin kecil persentase butir patah, maka semakin tinggi pula kualitas dan harga beras tersebut. Oleh karena itu, perbedaan kualitas beras menjadi aspek penting dalam analisis harga beras dan dapat dijadikan dasar dalam pengembangan model prediksi harga secara lebih akurat.

3.3 Prediksi

Prediksi adalah upaya memperkirakan sesuatu yang akan terjadi di masa depan dengan menggunakan berbagai informasi yang relevan pada waktu sebelumnya dengan menggunakan metode ilmiah (Rahayu, 2023). Tujuan dari prediksi adalah untuk memperoleh gambaran mengenai kondisi yang akan datang berdasarkan pola atau hubungan yang ditemukan dari data historis. Proses dari prediksi melibatkan analisis data, identifikasi variabel yang berpengaruh, serta penerapan model atau algoritma tertentu agar hasil yang diperoleh memiliki tingkat akurasi yang tinggi. Hasil prediksi tersebut dapat digunakan sebagai dasar dalam pengambilan keputusan, perencanaan strategi, maupun evaluasi kebijakan di berbagai bidang. Dengan adanya hasil prediksi yang akurat, pihak terkait dapat melakukan langkah antisipatif terhadap potensi perubahan kondisi di masa depan.

3.4 Machine Learning

Machine learning merupakan cabang ilmu bagian dari kecerdasan buatan (artificial intelligence), dengan pemrograman untuk memungkinkan komputer menjadi cerdas berperilaku seperti manusia, dan dapat meningkatkan pemahamannya melalui pengalaman secara otomatis (Retnoningsih & Pramudita, 2020). Proses pembelajaran machine learning dilakukan melalui algoritma yang menganalisis data dan menyesuaikan parameter model agar kesalahan prediksi dapat diminimalkan.

Berdasarkan teknik pembelajarannya, machine learning dapat dibedakan menjadi supervised learning, unsupervised learning, dan reinforcement learning. Supervised learning merupakan salah satu teknik machine learning yang menggunakan dataset (data training) yang sudah berlabel (labeled data) untuk melakukan pembelajaran pada mesin, sehingga mesin mampu mengidentifikasi label input dengan menggunakan fitur yang dimiliki untuk selanjutnya melakukan prediksi maupun klasifikasi (Retnoningsih & Pramudita, 2020). Unsupervised learning merupakan teknik machine learning yang digunakan untuk menganalisis data yang tidak memiliki label (unlabeled data). Sedangkan Reinforcement Learning adalah algoritma yang mengumpulkan

pengetahuan (sinyal penguatan) untuk memilih tindakan yang mengarah ke hasil tertinggi yang diharapkan (Alfarizi et al, 2023).

Dalam penelitian mengenai prediksi harga beras, pendekatan *supervised learning* banyak digunakan karena melibatkan data historis yang telah memiliki nilai target. Dua model yang umum digunakan dalam kategori ini adalah Support Vector Regression (SVR) dan Extreme Gradient Boosting (XGBoost). Model tersebut dipilih karena dikenal memiliki kemampuan tinggi dalam mengolah data nonlinier serta menghasilkan prediksi yang akurat.

1. Support Vector Regression (SVR)

Support Vector Regression (SVR) adalah algoritma supervised learning yang digunakan untuk memprediksi nilai variabel kontinu (Fadhil, 2025). Seperti konsep SVM, metode SVR juga mencari hyperplane terbaik berupa fungsi regresi dengan membuat error sekecil mungkin dengan memaksimalkan margin (Ruliana et al., 2024). SVR menggunakan konsep *epsilon-insensitive loss function*, di mana kesalahan yang berada di dalam batas toleransi (ϵ) tidak akan memengaruhi model, sedangkan kesalahan di luar batas tersebut akan dihitung sebagai penalti. Dengan cara ini, model tetap sederhana namun tetap mampu menghasilkan prediksi yang tepat.

Secara umum, cara kerja SVR dimulai dengan memetakan data masukan ke ruang berdimensi lebih tinggi menggunakan fungsi kernel (*kernel function*), seperti *linear*, *polynomial*, atau *radial basis function (RBF)*, agar hubungan nonlinier antarvariabel dapat direpresentasikan secara linear di ruang tersebut. Algoritma kemudian mencari *support vector*, yaitu titik-titik data terdekat dengan *hyperplane* yang menentukan posisi serta arah fungsi regresi.

Selama proses pelatihan, SVR melakukan optimasi untuk meminimalkan kompleksitas model sekaligus mengurangi kesalahan prediksi dengan menyesuaikan parameter regularisasi (C), yang mengontrol keseimbangan antara toleransi terhadap kesalahan dan lebar margin. Hasilnya adalah model regresi yang mampu melakukan prediksi dengan tingkat akurasi tinggi serta memiliki kemampuan generalisasi yang baik terhadap data baru. Dengan keunggulan tersebut, SVR banyak diterapkan dalam berbagai bidang, seperti prediksi harga komoditas, peramalan

ekonomi, dan analisis data nonlinier, karena kemampuannya dalam menghasilkan prediksi yang stabil dan akurat.

2. Extreme Gradient Boosting (XGBoost)

XGBoost merupakan model pengembangan dari gradient boosting, algoritma yang dapat menemukan solusi optimal untuk mengatasi permasalahan regresi dan klasifikasi berdasarkan Gradient Boosting Decision Tree (Andriansyah & Fridayanthie, 2023). XGBoost bekerja dengan menggabungkan beberapa model keputusan sederhana (weak learners) untuk membentuk model yang lebih kuat (Fadhil, 2025).

Secara sistematis, cara kerja XGBoost dimulai dengan membangun model awal yang sederhana, kemudian secara berulang menambahkan pohon keputusan baru untuk memperbaiki kesalahan prediksi dari model sebelumnya. Setiap pohon tambahan dibentuk berdasarkan *residual error* atau selisih antara nilai aktual dan nilai prediksi, sehingga model selanjutnya akan lebih fokus pada data yang sebelumnya sulit diprediksi. Proses ini terus dilakukan hingga model mencapai tingkat akurasi yang optimal atau jumlah iterasi yang telah ditentukan.

Selain memiliki mekanisme pembelajaran bertahap, XGBoost juga dilengkapi dengan fitur regularisasi yang berfungsi untuk mencegah terjadinya *overfitting*. Regularisasi dilakukan melalui parameter *lambda* (L2 regularization) dan *alpha* (L1 regularization), yang mengontrol kompleksitas model agar tetap seimbang antara akurasi dan kemampuan generalisasi. Parameter penting lainnya yang memengaruhi kinerja model antara lain *learning_rate* (*eta*) yang mengatur besarnya langkah pembaruan bobot pada setiap iterasi, *max_depth* yang menentukan kedalaman maksimum pohon keputusan, serta *n_estimators* yang mengatur jumlah pohon yang akan dibangun selama proses *boosting*. Selain itu, parameter *subsample* dan *colsample_bytree* digunakan untuk menentukan proporsi data dan fitur yang digunakan pada setiap iterasi, sehingga dapat meningkatkan efisiensi serta kemampuan generalisasi model. Dengan kombinasi teknik *gradient boosting*, regularisasi yang kuat, serta optimasi komputasi yang efisien, XGBoost mampu menghasilkan prediksi dengan tingkat akurasi tinggi, kecepatan pelatihan yang cepat, dan performa yang stabil. Oleh karena itu,

XGBoost menjadi salah satu algoritma yang banyak digunakan dalam berbagai bidang, seperti prediksi harga, analisis keuangan, dan pengolahan data berskala besar.

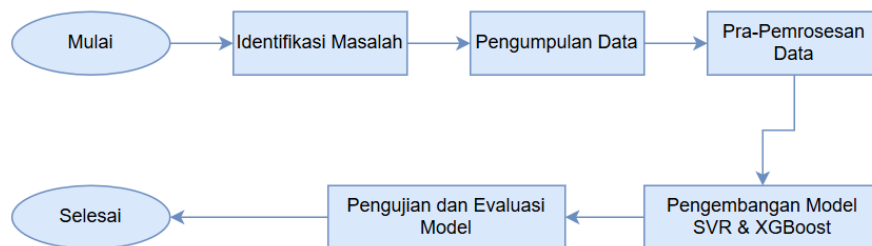
BAB III

METODOLOGI PENELITIAN

Bab ini menguraikan tahapan-tahapan penelitian mulai dari indentifikasi masalah hingga skenario pengujian. Setiap tahapan dirancang secara sistematis untuk memastikan proses penelitian berjalan terarah dan hasil yang diperoleh dapat dipertanggungjawabkan secara ilmiah.

3.1 Tahapan Penelitian

Tahapan penelitian ini menggambarkan alur kerja yang dilakukan untuk mencapai tujuan penelitian. Secara umum, proses penelitian meliputi beberapa tahap utama, yaitu:



3.2 Identifikasi Masalah

Tahap identifikasi masalah merupakan langkah awal dalam penelitian yang bertujuan untuk menemukan dan merumuskan permasalahan yang akan diselesaikan. Dalam konteks penelitian ini, permasalahan utama yang dihadapi adalah fluktuasi harga beras di Indonesia yang dipengaruhi oleh berbagai faktor, seperti kualitas beras, jumlah produksi padi, dan kondisi pasar. Ketidakstabilan harga beras ini dapat berdampak pada aspek ekonomi masyarakat dan kebijakan pemerintah dalam menjaga ketahanan pangan nasional.

Permasalahan tersebut menunjukkan perlunya suatu model prediksi harga beras yang akurat dan adaptif terhadap perubahan data. Metode konvensional seperti analisis statistik linier sering kali kurang mampu menangkap hubungan non-linear antara variabel, sehingga hasil prediksi kurang optimal. Oleh karena itu, penelitian ini

mengusulkan pendekatan berbasis *machine learning* untuk mempelajari pola kompleks dari data historis harga dan produksi beras.

Dua algoritma yang digunakan dalam penelitian ini adalah Support Vector Regression (SVR) dan Extreme Gradient Boosting (XGBoost). Kedua model tersebut dipilih karena memiliki kemampuan dalam menangani data non-linear dan menghasilkan prediksi dengan tingkat kesalahan yang rendah. Namun, masing-masing model memiliki karakteristik dan performa yang berbeda. Oleh karena itu, dilakukan perbandingan kinerja antara SVR dan XGBoost untuk mengetahui model mana yang memberikan hasil paling akurat dan efisien dalam memprediksi harga beras berdasarkan kualitas dan jumlah produksi padi.

3.3 Pengumpulan Data

Tahap pengumpulan data merupakan proses penting dalam penelitian yang bertujuan untuk memperoleh informasi yang relevan dan akurat sesuai dengan kebutuhan analisis. Pada penelitian ini, data yang digunakan bersumber dari Badan Pusat Statistik (BPS), yang menyediakan data rata-rata harga beras berdasarkan kualitas serta data produksi tanaman padi di Indonesia. Rentang waktu data yang digunakan adalah tahun 2020 hingga 2024, agar dapat menggambarkan tren harga dan produksi dalam beberapa periode terakhir.

Variabel yang digunakan dalam penelitian ini meliputi tahun, bulan, harga beras per bulan, kualitas beras (premium, medium, dan rendah), serta jumlah produksi padi. Data tersebut kemudian akan diolah dan dianalisis untuk membangun model prediksi harga beras menggunakan algoritma Support Vector Regression (SVR) dan Extreme Gradient Boosting (XGBoost). Dengan menggunakan data yang terukur dan terverifikasi, diharapkan model yang dikembangkan mampu memberikan hasil prediksi yang lebih akurat dan dapat digunakan sebagai dasar dalam pengambilan keputusan terkait kebijakan harga beras.

3.4 Pra-Pemrosesan Data

Pra-pemrosesan data merupakan tahap penting sebelum proses pemodelan dilakukan, dengan tujuan untuk memastikan bahwa data yang digunakan memiliki kualitas yang baik dan sesuai dengan kebutuhan algoritma *machine learning*. Data mentah yang diperoleh umumnya masih mengandung nilai yang hilang (*missing values*), data duplikat, atau ketidaksesuaian format yang dapat memengaruhi hasil

prediksi. Oleh karena itu, pada tahap ini dilakukan serangkaian proses untuk membersihkan, mentransformasi, dan menyiapkan data agar siap digunakan pada tahap pengembangan model. Proses pra-pemrosesan data dalam penelitian ini meliputi beberapa langkah sebagai berikut:

1. Pembersihan Data (Data Cleaning)

Pada tahap ini dilakukan pemeriksaan terhadap adanya data yang tidak valid, duplikat, atau tidak relevan. Nilai yang hilang pada variabel numerik seperti harga beras atau jumlah produksi diatasi dengan metode imputasi, misalnya menggunakan rata-rata (*mean imputation*). Selain itu, data yang mengandung kesalahan pencatatan atau berada di luar rentang wajar juga diperbaiki atau dihapus agar tidak menimbulkan bias pada model.

2. Transformasi Data (Data Transformation)

Data yang memiliki format tidak seragam, seperti satuan berat, volume, atau nilai mata uang, diseragamkan untuk memastikan konsistensi antarvariabel. Transformasi juga dilakukan terhadap variabel kategori, seperti kualitas beras, dengan menggunakan teknik *label encoding* atau *one-hot encoding* agar dapat diolah oleh model *machine learning*.

3. Normalisasi Data (Data Normalization)

Karena skala variabel dapat memengaruhi kinerja algoritma seperti SVR dan XGBoost, maka dilakukan proses normalisasi untuk menyamakan rentang nilai antarfitur. Normalisasi dilakukan menggunakan metode *Min-Max Scaler*, sehingga setiap nilai berada dalam rentang 0 hingga 1. Hal ini membantu model dalam mempercepat proses pembelajaran dan meningkatkan stabilitas hasil prediksi.

4. Pembagian Data (Data Splitting)

Setelah data bersih dan terstandardisasi, dataset dibagi menjadi dua bagian, yaitu data latih (*training data*) dan data uji (*testing data*). Pembagian dilakukan dengan proporsi 80% untuk data latih dan 20% untuk data uji. Data latih digunakan untuk melatih model agar dapat mengenali pola hubungan antara variabel input dan output, sedangkan data uji digunakan untuk mengukur kemampuan generalisasi model terhadap data baru yang belum pernah dilihat sebelumnya.

Tahapan pra-pemrosesan data ini berperan penting dalam menghasilkan dataset yang berkualitas, sehingga model SVR dan XGBoost dapat bekerja secara optimal dalam mempelajari pola dan menghasilkan prediksi harga beras yang akurat berdasarkan kualitas serta jumlah produksi padi.

3.5 Pengembangan Model

Tahap pengembangan model merupakan proses utama dalam penelitian ini yang bertujuan untuk membangun dan melatih model prediksi harga beras menggunakan algoritma *Support Vector Regression (SVR)* dan *Extreme Gradient Boosting (XGBoost)*. Kedua model ini dipilih karena memiliki kemampuan yang baik dalam menangani data non-linear serta menghasilkan prediksi dengan tingkat kesalahan yang rendah. Proses pengembangan dilakukan berdasarkan data yang telah melalui tahap pra-pemrosesan agar model dapat belajar secara optimal dari pola data historis.

1. Pengembangan Model Support Vector Regression

Dalam penelitian ini, SVR digunakan untuk memodelkan hubungan antara variabel input berupa kualitas beras dan jumlah produksi padi dengan variabel output berupa harga beras. Model Support Vector Regression bekerja dengan menentukan batas toleransi kesalahan (ϵ) di sekitar garis regresi dan berusaha meminimalkan deviasi dari batas tersebut. Proses pelatihan model SVR meliputi beberapa tahapan, yaitu:

- 1) Menentukan parameter utama seperti kernel, C (cost parameter), dan epsilon (ϵ).
- 2) Melakukan penyesuaian parameter (*hyperparameter tuning*) menggunakan metode *Grid Search* untuk memperoleh kombinasi parameter terbaik.
- 3) Melatih model menggunakan data latih (*training data*) dan menyimpan model hasil pelatihan untuk tahap evaluasi.

Dalam penelitian ini, digunakan fungsi kernel Radial Basis Function (RBF) karena mampu menangkap hubungan non-linear antara variabel. Model SVR yang telah dilatih kemudian digunakan untuk memprediksi harga beras pada data uji (*testing data*) dan hasilnya dibandingkan dengan model XGBoost.

2. Pengembangan Model Extreme Gradient Boosting (XGBoost)

Pada penelitian ini, XGBoost digunakan untuk memprediksi harga beras berdasarkan kualitas dan jumlah produksi padi. Setiap iterasi dalam proses

pelatihannya berfokus pada pengurangan kesalahan dari prediksi sebelumnya, sehingga menghasilkan model dengan kemampuan generalisasi yang tinggi. Tahapan pengembangan model XGBoost meliputi:

- 1) Penentuan parameter utama seperti learning rate, max depth, n_estimators, dan subsample.
- 2) Optimasi parameter menggunakan Grid Search untuk memperoleh konfigurasi terbaik.
- 3) Pelatihan model menggunakan data latih dengan meminimalkan fungsi loss (biasanya Mean Squared Error).
- 4) Pengujian model terhadap data uji untuk memperoleh hasil prediksi harga beras.

Melalui tahapan tersebut, model XGBoost diharapkan mampu mengenali pola hubungan kompleks antara variabel kualitas beras dan jumlah produksi terhadap harga beras, sehingga dapat menghasilkan prediksi yang lebih akurat, stabil, dan efisien dibandingkan metode regresi konvensional.

3.6 Pengujian dan Evaluasi Model

Tahap pengujian dan evaluasi model bertujuan untuk menilai kinerja model *Support Vector Regression (SVR)* dan *Extreme Gradient Boosting (XGBoost)* dalam memprediksi harga beras berdasarkan kualitas dan jumlah produksi padi. Setelah kedua model selesai dikembangkan dan dioptimasi, dilakukan proses pengujian menggunakan data uji (*testing data*) yang sebelumnya telah dipisahkan pada tahap pra-pemrosesan. Pengujian ini bertujuan untuk mengetahui sejauh mana model mampu melakukan prediksi terhadap data baru yang belum pernah dilihat sebelumnya.

Evaluasi dilakukan dengan membandingkan hasil prediksi model terhadap nilai aktual harga beras menggunakan beberapa metrik kinerja regresi, yaitu *Mean Absolute Error (MAE)*, *Root Mean Squared Error (RMSE)*, dan *R-squared (R^2)*.

1. Mean Absolute Error (MAE)

Metrik ini mengukur rata-rata selisih absolut antara nilai prediksi dan nilai aktual. Semakin kecil nilai MAE, semakin baik kinerja model dalam melakukan prediksi.

2. Root Mean Squared Error (RMSE)

RMSE digunakan menghitung akar dari rata-rata kesalahan kuadrat antara nilai aktual dan nilai prediksi. Nilai RMSE yang lebih kecil menunjukkan tingkat kesalahan prediksi yang lebih rendah.

3. R-squared (R^2)

Nilai R^2 menunjukkan proporsi variansi data yang dapat dijelaskan oleh model. Semakin mendekati 1, semakin baik kemampuan model dalam menjelaskan hubungan antara variabel input dan output.

Setelah hasil evaluasi diperoleh, dilakukan perbandingan kinerja antara model SVR dan XGBoost berdasarkan ketiga metrik tersebut. Model dengan nilai MAE dan RMSE yang lebih rendah serta nilai R^2 yang lebih tinggi dianggap memiliki performa terbaik. Analisis hasil pengujian ini bertujuan untuk menentukan model yang paling akurat dan stabil dalam memprediksi harga beras berdasarkan kualitas dan jumlah produksi padi. Selain itu, hasil evaluasi juga digunakan untuk menilai sejauh mana metode berbasis *machine learning* mampu meningkatkan akurasi prediksi dibandingkan pendekatan konvensional.

Halaman ini sengaja dikosongkan

DAFTAR PUSTAKA

- [1] Alfarizi et al. (2023). Karimah Tauhid, Volume 2 Nomor 1 (2023), e-ISSN 2963-590X. 2.
- [2] Andriansyah, D.- & Fridayanthie. (2023). Optimization of Support Vector Machine and XGBoost Methods Using Feature Selection to Improve Classification Performance. JOURNAL OF INFORMATICS AND TELECOMMUNICATION ENGINEERING, 6(2), 484–493. <https://doi.org/10.31289/jite.v6i2.8373>
- [3] Bhoopathi, S., Kumar, N., Somesh, & Pal, M. (2024). Evaluating the performances of SVR and XGBoost for short-range forecasting of heatwaves across different temperature zones of India. Applied Computing and Geosciences, 24, 100204. <https://doi.org/10.1016/j.acags.2024.100204>
- [4] Fadhil, N. (2025). PERBANDINGAN AKURASI ALGORITMA XGBOOST DAN SVR DALAM PREDIKSI HARGA CRYPTOCURRENCY.
- [5] Nafi'iyah, N., & Wulandari, P. A. (2022). Prediksi Harga Beras Berdasarkan Kualitas Beras dengan Metode Long Short Term Memory. 7(2).
- [6] Pragana, Manurung, D. W., & Agus Perdana Windarto. (2023). Analisa Metode Backpropagation Pada Prediksi Rata-rata Harga Beras Bulanan Di Tingkat Penggilingan Menurut Kualitas. Journal of Computing and Informatics Research, 2(3), 77–84. <https://doi.org/10.47065/comforch.v2i3.855>
- [7] Rahayu, C. A. (2023). PREDIKSI PENDERITA DIABETES MENGGUNAKAN METODE NAIVE BAYES. Jurnal Informatika dan Teknik Elektro Terapan, 11(3). <https://doi.org/10.23960/jitet.v11i3.3055>
- [8] Rahman, A., Paramarta, V., Ida, A. N., Akbar, M. H., & Simanjuntak, V. S. M. (2025). Perbandingan Kinerja SVR dan XGBoost untuk Peramalan Emisi CO₂ Global berbasis Machine Learning. Jurnal Komtika (Komputasi dan Informatika), 9(1), 37–44. <https://doi.org/10.31603/komtika.v9i1.13449>
- [9] Rayadin, M. A., Musaruddin, M., Saputra, R. A., & Isnawaty, I. (2024). Implementasi Ensemble Learning Metode XGBoost dan Random Forest untuk Prediksi Waktu Penggantian Baterai Aki. BIOS : Jurnal Teknologi Informasi dan Rekayasa Komputer, 5(2), 111–119. <https://doi.org/10.37148/bios.v5i2.128>

- [10] Retnoningsih, E., & Pramudita, R. (2020). Mengenal Machine Learning Dengan Teknik Supervised Dan Unsupervised Learning Menggunakan Python. BINA INSANI ICT JOURNAL, 7(2), 156. <https://doi.org/10.51211/biict.v7i2.1422>
- [11] Ruliana, R., Rais, Z., Marni, M., & Ahmar, A. S. (2024). Implementation of the Support Vector Regression (SVR) Method in Inflation Prediction in Makassar City. ARRUS Journal of Mathematics and Applied Science, 4(1), 28–35. <https://doi.org/10.35877/mathscience2608>
- [12] Ruvananda, A. R., & Taufiq, M. (2022). Analisis faktor-faktor yang mempengaruhi impor beras di Indonesia.
- [13] Saputra, G. D., & Kurniati, E. (2025). Analisis Implementasi Sistem Distribusi Dalam Menjaga Keseimbangan Permintaan Dan Penawaran Produk Pertanian Kakao Di Pesawaran, Provinsi Lampung. 2(2).
- [14] Zainuddin, Z., & Nuryadin, M. B. (2024). Konsep Permintaan dan Penawaran dalam Ekonomi Mikro Islam: Perspektif Fiqh Muamalah. Maro: Jurnal Ekonomi Syariah dan Bisnis, 7(2), 327–338. <https://doi.org/10.31949/maro.v7i2.11812>