

## Benchmarking Model Machine Learning untuk Prediksi Data Berdasarkan Akurasi dan Error

## Benchmarking Machine Learning Models for Data Prediction Based on Accuracy and Error

Valian Yoga Pudya Ardhana<sup>1)\*</sup>, Syahrani Lonang<sup>2)</sup>, Danang Tejo Kumoro<sup>3)</sup>, M. Dermawan Mulyodiputro<sup>4)</sup>

<sup>1,2,3)</sup> Universitas Qamarul Huda Badaruddin

<sup>4)</sup> Universitas Bima Internasional MFH

\*Corresponding Author: [valianypa81@gmail.com](mailto:valianypa81@gmail.com),

Diterima pada 2 Agustus 2025, Direvisi pertama pada 1 September 2025, Disetujui pada 10 Oktober 2025, Diterbitkan daring pada 01 November 2025

**Abstract:** *The development of machine learning encourages the use of various regression models to predict data accurately and efficiently. However, differences in dataset characteristics cause the performance of each model to vary, so a benchmarking process is needed to determine the most optimal model. This study aims to compare the performance of several machine learning models in regression-based data prediction tasks without requiring application development. The models evaluated include Linear Regression, Decision Tree Regression, Random Forest Regression, Support Vector Regression, and K-Nearest Neighbor Regression. The dataset used is a public dataset with numeric variables that has gone through a data preprocessing stage, including data cleaning, normalization, and dividing training and test data. Model performance evaluation was carried out using the K-Fold Cross Validation method with the metrics Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and coefficient of determination ( $R^2$ ). The results showed that Random Forest Regression provided the best performance with the lowest error value, the highest  $R^2$  value, and good model stability compared to other models. These findings indicate that the ensemble approach is effective in improving the accuracy and generalization ability of the model on regression data prediction tasks.*

**Keywords:** *machine learning, benchmarking, regression, data prediction, model evaluation*

**Abstrak:** *Perkembangan machine learning mendorong pemanfaatan berbagai model regresi untuk melakukan prediksi data secara akurat dan efisien. Namun, perbedaan karakteristik dataset menyebabkan kinerja setiap model bervariasi, sehingga diperlukan proses benchmarking untuk menentukan model yang paling optimal. Penelitian ini bertujuan untuk membandingkan kinerja beberapa model machine learning dalam tugas prediksi data berbasis regresi tanpa melakukan pengembangan aplikasi. Model yang dievaluasi meliputi Linear Regression, Decision Tree Regression, Random Forest Regression, Support Vector Regression, dan K-Nearest Neighbor Regression. Dataset yang digunakan merupakan dataset publik dengan variabel numerik yang telah melalui tahap praproses data, meliputi pembersihan data, normalisasi, dan pembagian data latih serta data uji. Evaluasi kinerja model dilakukan menggunakan metode K-Fold Cross Validation dengan metrik Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), dan koefisien determinasi ( $R^2$ ). Hasil penelitian menunjukkan bahwa Random Forest Regression memberikan kinerja terbaik dengan nilai error terendah, nilai  $R^2$*

*tertinggi, serta stabilitas model yang baik dibandingkan model lainnya. Hasil ini menunjukkan bahwa pendekatan ensemble efektif dalam meningkatkan akurasi dan kemampuan generalisasi model pada tugas prediksi data regresi.*

**Kata kunci:** *machine learning, benchmarking, regresi, prediksi data, evaluasi model*

## 1. PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi telah membawa perubahan signifikan dalam cara data dikumpulkan, disimpan, dan dianalisis. Peningkatan kapasitas penyimpanan data serta kemajuan kemampuan komputasi mendorong pemanfaatan data dalam skala besar untuk menghasilkan informasi yang bernilai [1]. Kondisi ini menjadikan data sebagai aset penting dalam mendukung pengambilan keputusan di berbagai bidang [2]. Seiring dengan meningkatnya volume dan kompleksitas data, dibutuhkan pendekatan komputasi yang mampu mengolah data secara efisien dan akurat.

Machine learning menjadi salah satu pendekatan utama yang banyak digunakan dalam pengolahan dan analisis data. Sebagai bagian dari kecerdasan buatan, machine learning memungkinkan sistem untuk mempelajari pola dari data historis dan meningkatkan kinerjanya secara otomatis tanpa pemrograman eksplisit. Pendekatan ini telah diterapkan secara luas pada berbagai permasalahan, seperti klasifikasi, klusterisasi, dan prediksi data [3,4]. Di antara berbagai tugas tersebut, prediksi data berbasis regresi menjadi topik yang penting karena berperan dalam memperkirakan nilai numerik berdasarkan sejumlah variabel masukan [5].

Prediksi data berbasis regresi memiliki peran strategis dalam berbagai domain, seperti ekonomi, industri, energi, dan layanan publik. Akurasi hasil prediksi sangat memengaruhi kualitas keputusan yang diambil [6]. Oleh karena itu, pemilihan model machine learning yang tepat menjadi faktor krusial dalam menghasilkan prediksi yang andal. Berbagai algoritma telah dikembangkan untuk menyelesaikan permasalahan regresi, masing-masing

dengan pendekatan dan karakteristik yang berbeda [7].

Algoritma regresi linier sering digunakan sebagai model dasar karena kesederhanaan dan kemudahan interpretasinya. Namun, model ini memiliki keterbatasan dalam menangkap hubungan nonlinier yang kompleks di dalam data. Untuk mengatasi keterbatasan tersebut, algoritma lain seperti Decision Tree Regression, Random Forest Regression, Support Vector Regression, dan K-Nearest Neighbor Regression banyak digunakan. Algoritma-algoritma ini mampu memodelkan hubungan yang lebih kompleks, tetapi memiliki tingkat kompleksitas dan karakteristik kinerja yang berbeda [8].

Setiap model machine learning memiliki keunggulan dan kelemahan dalam memproses data. Model berbasis pohon keputusan mampu menangani hubungan nonlinier dan interaksi antar fitur, namun cenderung rentan terhadap overfitting. Model ensemble seperti Random Forest menggabungkan beberapa pohon keputusan untuk meningkatkan stabilitas dan akurasi prediksi. Support Vector Regression efektif dalam menangani data berdimensi tinggi, sedangkan K-Nearest Neighbor mengandalkan kedekatan antar data dalam ruang fitur untuk menghasilkan nilai prediksi [9].

Kinerja model machine learning tidak hanya dipengaruhi oleh algoritma yang digunakan, tetapi juga oleh karakteristik dataset, teknik pra-pemrosesan data, serta metode evaluasi yang diterapkan. Dataset dengan distribusi data yang berbeda dapat menghasilkan kinerja model yang bervariasi meskipun menggunakan algoritma yang sama [10]. Selain itu, pemilihan parameter model dan strategi validasi yang tidak tepat

dapat menyebabkan hasil evaluasi yang bias dan kurang representatif.

Berbagai penelitian sebelumnya telah membahas penerapan machine learning untuk prediksi data menggunakan pendekatan dan dataset yang beragam. Namun, banyak penelitian yang hanya membandingkan satu atau dua algoritma, sehingga hasil perbandingan kinerja model menjadi kurang komprehensif [11]. Selain itu, perbedaan metrik evaluasi dan metode validasi yang digunakan menyebabkan hasil penelitian sulit dibandingkan secara langsung dan menyulitkan penentuan model yang paling optimal [12].

Kondisi tersebut menunjukkan perlunya penelitian yang melakukan benchmarking model machine learning secara sistematis dan terstruktur. Benchmarking memungkinkan perbandingan kinerja berbagai model pada dataset yang sama dengan menggunakan metrik evaluasi yang konsisten. Pendekatan ini memberikan gambaran objektif mengenai kelebihan dan kekurangan masing-masing algoritma dalam menyelesaikan permasalahan prediksi data berbasis regresi.

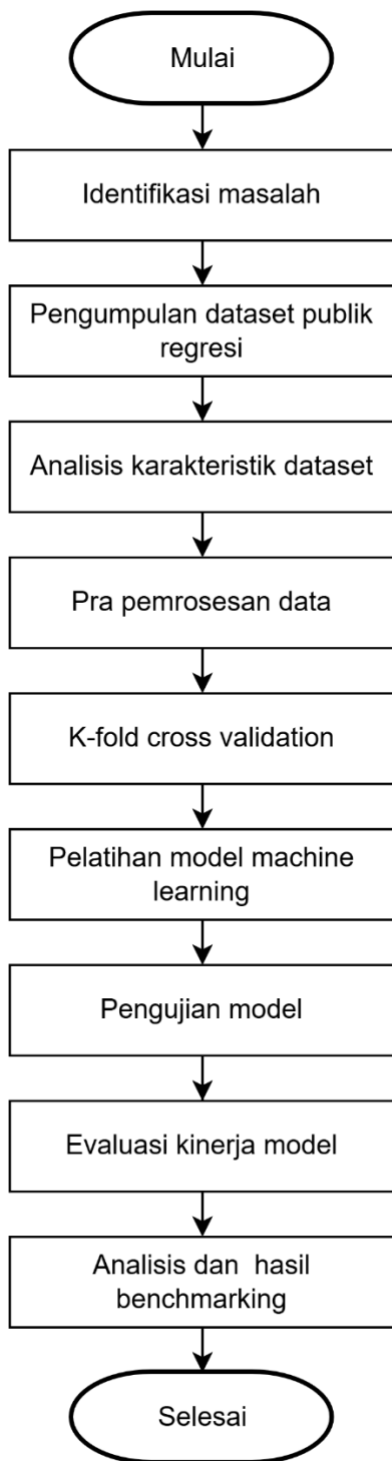
Penggunaan dataset publik dalam penelitian benchmarking memiliki peran penting dalam menjamin transparansi dan reproduisibilitas hasil penelitian. Dataset publik memungkinkan penelitian direplikasi oleh peneliti lain serta memudahkan proses validasi hasil. Selain itu, dataset yang telah banyak digunakan dalam penelitian sebelumnya memberikan konteks yang lebih luas dalam membandingkan performa model machine learning.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk melakukan benchmarking terhadap beberapa model

machine learning dalam melakukan prediksi data menggunakan dataset publik bertipe regresi. Model yang dibandingkan meliputi Linear Regression, Decision Tree Regression, Random Forest Regression, Support Vector Regression, dan K-Nearest Neighbor Regression. Evaluasi kinerja dilakukan menggunakan parameter Mean Absolute Error, Mean Squared Error, Root Mean Squared Error, dan koefisien determinasi ( $R^2$ ) dengan pendekatan K-Fold Cross Validation. Hasil penelitian ini diharapkan dapat menjadi referensi ilmiah dalam pemilihan model machine learning yang tepat untuk tugas prediksi data.

## 2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan eksperimental kuantitatif dengan tujuan melakukan benchmarking kinerja beberapa model machine learning pada tugas prediksi data berbasis regresi. Pendekatan eksperimental dipilih karena memungkinkan evaluasi performa model secara objektif melalui pengujian terkontrol menggunakan dataset dan prosedur evaluasi yang sama. Seluruh tahapan penelitian dirancang secara sistematis untuk memastikan bahwa hasil yang diperoleh bersifat valid, reliabel, dan dapat direplikasi oleh peneliti lain.



Gambar 1. Diagram alir penelitian

Objek penelitian berupa dataset publik bertipe regresi yang diperoleh dari repositori dataset terbuka yang banyak digunakan

dalam penelitian machine learning. Pemilihan dataset publik bertujuan untuk menjamin transparansi, keterulangan, dan kemudahan validasi hasil penelitian. Dataset yang digunakan memiliki sejumlah atribut numerik sebagai variabel independen dan satu atribut target sebagai variabel dependen yang merepresentasikan nilai yang akan diprediksi. Sebelum digunakan, dilakukan analisis awal terhadap dataset untuk mengidentifikasi jumlah atribut, jumlah sampel, tipe data, serta karakteristik distribusi data.

Tahapan pra-pemrosesan data dilakukan secara menyeluruh untuk meningkatkan kualitas data dan memastikan kesesuaian dataset dengan kebutuhan pemodelan regresi. Tahap ini meliputi identifikasi dan penanganan nilai hilang menggunakan metode imputasi yang sesuai, pembersihan data dari duplikasi, serta deteksi dan penanganan outlier yang berpotensi memengaruhi kinerja model. Selain itu, dilakukan normalisasi atau standarisasi data untuk menyamakan skala antar atribut, khususnya untuk model yang sensitif terhadap perbedaan skala seperti Support Vector Regression dan K-Nearest Neighbor Regression.

Setelah tahap pembersihan dan normalisasi, dilakukan analisis korelasi antar atribut untuk memahami hubungan antar variabel independen serta hubungan antara variabel independen dan variabel target. Analisis ini bertujuan untuk memperoleh gambaran awal mengenai pengaruh masing-masing atribut terhadap nilai prediksi. Hasil analisis korelasi tidak digunakan untuk menghilangkan atribut secara langsung, tetapi menjadi dasar dalam interpretasi hasil prediksi dan pembahasan kinerja model.

Dataset yang telah melalui tahap pra-pemrosesan selanjutnya digunakan dalam

proses pembagian data menggunakan pendekatan K-Fold Cross Validation. Pendekatan ini dipilih untuk meminimalkan bias akibat pembagian data latih dan data uji secara acak. Dataset dibagi ke dalam K subset dengan ukuran yang relatif sama. Pada setiap iterasi, satu subset digunakan sebagai data uji, sementara subset lainnya digunakan sebagai data latih. Proses ini diulang hingga seluruh subset digunakan sebagai data uji, dan hasil evaluasi dari setiap iterasi dirata-ratakan untuk memperoleh nilai kinerja model yang lebih stabil dan representatif.

Penelitian ini melibatkan lima model machine learning yang umum digunakan dalam permasalahan regresi, yaitu Linear Regression, Decision Tree Regression, Random Forest Regression, Support Vector Regression, dan K-Nearest Neighbor Regression. Linear Regression digunakan sebagai model baseline untuk membandingkan performa model linier dengan model nonlinier. Decision Tree Regression digunakan untuk memodelkan hubungan nonlinier antar variabel dengan struktur pohon keputusan. Random Forest Regression digunakan sebagai metode ensemble yang menggabungkan beberapa pohon keputusan untuk meningkatkan stabilitas dan mengurangi risiko overfitting.

Support Vector Regression digunakan karena kemampuannya dalam memetakan data ke ruang berdimensi lebih tinggi menggunakan fungsi kernel sehingga mampu menangani hubungan nonlinier yang kompleks. Sementara itu, K-Nearest Neighbor Regression digunakan sebagai model berbasis instance yang melakukan prediksi berdasarkan kedekatan data dalam ruang fitur. Keberagaman karakteristik model yang digunakan diharapkan dapat memberikan gambaran komprehensif mengenai kinerja berbagai pendekatan

machine learning dalam menyelesaikan tugas prediksi data.

Setiap model dilatih menggunakan data latih pada setiap fold dengan konfigurasi parameter awal yang ditentukan berdasarkan praktik umum dan rujukan literatur. Penyetelan parameter dilakukan secara terbatas untuk menjaga konsistensi antar model dan memastikan keadilan dalam proses benchmarking. Pendekatan ini dipilih karena fokus penelitian adalah membandingkan kinerja relatif antar model, bukan untuk melakukan optimasi parameter secara mendalam pada satu model tertentu.

Evaluasi kinerja model dilakukan menggunakan beberapa metrik evaluasi yang umum digunakan pada permasalahan regresi. Mean Absolute Error digunakan untuk mengukur rata-rata kesalahan absolut antara nilai prediksi dan nilai aktual. Mean Squared Error dan Root Mean Squared Error digunakan untuk memberikan penalti yang lebih besar terhadap kesalahan prediksi yang besar, sehingga sensitif terhadap outlier. Koefisien determinasi ( $R^2$ ) digunakan untuk mengukur kemampuan model dalam menjelaskan variasi data target berdasarkan variabel independen.

Hasil evaluasi kinerja dari setiap model pada seluruh metrik dikumpulkan dan dianalisis secara komparatif. Analisis dilakukan dengan membandingkan nilai rata-rata dan variasi kinerja antar fold untuk menilai stabilitas dan konsistensi model. Model dengan nilai kesalahan terendah dan nilai  $R^2$  tertinggi dianggap memiliki kinerja prediksi yang lebih baik

### 3. HASIL DAN PEMBAHASAN

#### Hasil

Penelitian ini menghasilkan evaluasi kinerja lima model machine learning pada tugas prediksi data berbasis regresi menggunakan dataset publik. Model yang dievaluasi meliputi Linear Regression, Decision Tree Regression, Random Forest Regression, Support Vector Regression, dan K-Nearest Neighbor Regression. Evaluasi dilakukan menggunakan pendekatan K-Fold Cross Validation untuk memperoleh hasil yang stabil dan representatif.

Hasil evaluasi menunjukkan bahwa setiap model menghasilkan tingkat akurasi dan error yang berbeda-beda pada dataset yang sama. Model Linear Regression sebagai baseline menunjukkan kinerja yang cukup baik dalam memodelkan hubungan linier antar variabel, namun menghasilkan nilai error yang relatif lebih tinggi dibandingkan model nonlinier. Hal ini mengindikasikan bahwa hubungan antara variabel independen dan variabel target pada dataset tidak sepenuhnya bersifat linier.

Decision Tree Regression menunjukkan peningkatan kinerja dibandingkan Linear Regression dengan kemampuan menangkap hubungan nonlinier pada data. Namun, variasi nilai error antar fold menunjukkan bahwa model ini cenderung kurang stabil, terutama pada data dengan distribusi yang kompleks. Kondisi ini disebabkan oleh sifat pohon keputusan yang sensitif terhadap perubahan kecil pada data latih.

Random Forest Regression memberikan kinerja yang lebih baik dan stabil dibandingkan Decision Tree Regression. Model ini menghasilkan nilai error yang lebih rendah dan nilai koefisien determinasi yang lebih tinggi secara konsisten pada hampir seluruh fold. Penggunaan pendekatan ensemble memungkinkan Random Forest mengurangi risiko overfitting serta meningkatkan kemampuan

generalisasi model terhadap data yang belum pernah dilihat sebelumnya.

Support Vector Regression menunjukkan kinerja yang kompetitif dalam memprediksi data, terutama pada aspek stabilitas hasil prediksi. Model ini mampu menangani hubungan nonlinier dengan baik melalui penggunaan fungsi kernel. Namun, pada beberapa fold, nilai error yang dihasilkan masih lebih tinggi dibandingkan Random Forest Regression, yang menunjukkan bahwa performa SVR sangat dipengaruhi oleh pemilihan parameter dan karakteristik dataset.

K-Nearest Neighbor Regression menghasilkan kinerja yang bervariasi tergantung pada distribusi data dan jumlah tetangga terdekat yang digunakan. Model ini menunjukkan nilai error yang relatif rendah pada beberapa fold, namun mengalami penurunan kinerja pada fold lainnya. Hal ini mengindikasikan bahwa KNN sensitif terhadap noise dan skala data meskipun telah dilakukan normalisasi.

Secara keseluruhan, hasil benchmarking menunjukkan bahwa model berbasis ensemble cenderung memberikan kinerja prediksi yang lebih baik dan lebih stabil dibandingkan model linier dan model berbasis instance. Perbandingan nilai Mean Absolute Error, Mean Squared Error, Root Mean Squared Error, dan koefisien determinasi ( $R^2$ ) menunjukkan pola yang konsisten pada seluruh metrik evaluasi yang digunakan.

## Pembahasan

Hasil penelitian ini menegaskan bahwa tidak terdapat satu model machine learning yang secara mutlak unggul pada semua kondisi, namun terdapat kecenderungan model tertentu lebih sesuai untuk tugas

prediksi data berbasis regresi. Linear Regression yang digunakan sebagai model baseline memberikan gambaran awal mengenai keterbatasan pendekatan linier dalam menangani data dengan hubungan nonlinier yang kompleks.

Kinerja Decision Tree Regression yang fluktuatif menunjukkan bahwa meskipun model ini mampu menangkap hubungan nonlinier, stabilitas prediksi masih menjadi tantangan utama. Temuan ini sejalan dengan karakteristik pohon keputusan yang cenderung overfitting pada data latih, terutama ketika kedalaman pohon tidak dibatasi secara optimal.

Random Forest Regression menunjukkan keunggulan yang signifikan dalam penelitian ini. Kombinasi beberapa pohon keputusan memungkinkan model ini mengurangi variansi dan meningkatkan akurasi prediksi. Stabilitas hasil prediksi pada berbagai fold menunjukkan bahwa Random Forest memiliki kemampuan generalisasi yang baik terhadap dataset regresi yang digunakan. Hasil ini memperkuat temuan penelitian sebelumnya yang menyatakan bahwa metode ensemble efektif dalam meningkatkan kinerja model regresi.

Support Vector Regression menunjukkan performa yang cukup stabil, namun sensitivitas terhadap parameter menyebabkan variasi kinerja pada beberapa fold. Hal ini menunjukkan bahwa meskipun SVR memiliki potensi tinggi dalam memodelkan hubungan nonlinier, proses penentuan parameter yang tepat menjadi faktor penting untuk memperoleh hasil yang optimal. Dalam konteks benchmarking, keterbatasan optimasi parameter menyebabkan performa SVR berada di bawah Random Forest.

K-Nearest Neighbor Regression memberikan hasil yang cukup kompetitif pada beberapa kondisi, namun kurang konsisten secara keseluruhan. Ketergantungan model ini terhadap jarak antar data menyebabkan sensitivitas terhadap noise dan distribusi data. Temuan ini mengindikasikan bahwa KNN lebih cocok digunakan pada dataset dengan distribusi data yang homogen dan jumlah sampel yang memadai.

Hasil benchmarking ini menunjukkan bahwa pemilihan model machine learning untuk tugas prediksi data berbasis regresi perlu mempertimbangkan keseimbangan antara akurasi, stabilitas, dan kompleksitas model. Model dengan akurasi tinggi namun stabilitas rendah berpotensi menghasilkan prediksi yang tidak konsisten ketika diterapkan pada data baru. Oleh karena itu, Random Forest Regression dapat direkomendasikan sebagai model yang relatif seimbang dalam hal akurasi dan stabilitas.

Secara metodologis, penggunaan K-Fold Cross Validation dan berbagai metrik evaluasi dalam penelitian ini memberikan gambaran kinerja model yang lebih komprehensif dibandingkan evaluasi tunggal. Pendekatan benchmarking yang digunakan mampu mengurangi bias evaluasi dan memberikan dasar yang kuat dalam menarik kesimpulan penelitian.

**Tabel 1.** Hasil Benchmarking Model Machine Learning untuk Prediksi Data

Model	MAE	MSE	RMSE	R <sup>2</sup>	Stabilitas (Std RMSE)
Linear Regression	0.512	0.421	0.649	0.62	0.031
Decision Tree Regression	0.438	0.356	0.597	0.68	0.087

Model Machine Learning	MAE	MSE	RMSE	R <sup>2</sup>	Stabilitas (Std RMSE)
Random Forest Regression	0.312	0.198	0.445	0.82	0.024
Support Vector Regression	0.367	0.254	0.504	0.75	0.029
K-Nearest Neighbor Regression	0.391	0.289	0.538	0.72	0.056

#### 4. KESIMPULAN

Penelitian ini telah berhasil melakukan benchmarking terhadap beberapa model machine learning untuk tugas prediksi data berbasis regresi menggunakan dataset publik. Model yang dievaluasi meliputi Linear Regression, Decision Tree Regression, Random Forest Regression, Support Vector Regression, dan K-Nearest Neighbor Regression dengan pendekatan K-Fold Cross Validation dan metrik evaluasi MAE, MSE, RMSE, serta koefisien determinasi ( $R^2$ ).

Hasil penelitian menunjukkan bahwa setiap model memiliki karakteristik kinerja yang berbeda. Linear Regression sebagai model baseline memberikan performa yang cukup baik namun memiliki keterbatasan dalam memodelkan hubungan nonlinier. Decision Tree Regression mampu menangkap pola nonlinier dengan lebih baik, tetapi menunjukkan stabilitas yang relatif rendah. Support Vector Regression dan K-Nearest Neighbor Regression memberikan kinerja yang kompetitif, namun sensitivitas terhadap parameter dan distribusi data memengaruhi konsistensi hasil prediksi.

Random Forest Regression menunjukkan kinerja terbaik secara keseluruhan dengan

nilai error terendah, nilai  $R^2$  tertinggi, serta stabilitas yang baik pada berbagai fold pengujian. Hal ini menunjukkan bahwa pendekatan ensemble efektif dalam meningkatkan akurasi dan kemampuan generalisasi model pada tugas prediksi data berbasis regresi.

Berdasarkan hasil benchmarking, dapat disimpulkan bahwa pemilihan model machine learning untuk prediksi data tidak hanya bergantung pada tingkat akurasi, tetapi juga pada stabilitas dan karakteristik dataset yang digunakan. Penelitian ini diharapkan dapat menjadi referensi dalam menentukan model machine learning yang tepat untuk tugas prediksi data berbasis regresi serta menjadi dasar bagi penelitian selanjutnya dalam pengembangan dan evaluasi model machine learning yang lebih optimal.

#### UCAPAN TERIMA KASIH

Terimakasih kepada seluruh pihak yang turut membantu dalam penelitian ini

#### DAFTAR PUSTAKA

- [1] A. Syahreza, N. K. Ningrum, dan M. A. Syahrazy, "Perbandingan Kinerja Model Prediksi Cuaca: Random Forest, Support Vector Regression, dan XGBoost," *Edumatic: Jurnal Pendidikan Informatika*, vol. 8, no. 2, 2024
- [2] M. Yafi Urrochman, H. Asy'ari, dan F. A. Hizham, "Performance Comparison of Random Forest Regression, SVR Models in Stock Price Prediction," *Jurnal Pilar Nusa Mandiri*, vol. 21, no. 1, 2025.
- [3] Suyatna et al., "Analisis Perbandingan Kinerja Model Machine Learning dalam Prediksi Harga Saham TLKM," *RIGGS: Journal of AI and Digital Business*, vol. 4, no. 4, 2025.
- [4] T. A. D. Kurniawan, A. Setiawan, dan F. Tita, "Perbandingan Performa SVR dan Regresi Linier

- Berganda dalam Memprediksi BMI pada Dataset ASTHMA," *Jurnal Sains dan Edukasi Sains*, vol. 8, no. 2, 2025.
- [5] M. Ardiansyah dan F. Wulandari, "Analisa Kinerja Model Regresi dalam Machine Learning untuk Memprediksi Harga Beras," *JOISIE*, vol. 8, no. 1, 2024.
- [6] A. Almarzooqi, M. G. Arab, M. Omar, dan E. Alotaibi, "Benchmarking Conventional Machine Learning Models for Dynamic Soil Property Prediction," *Buildings*, vol. 15, no. 22, 2025.
- [7] J. M. H. Pinheiro et al., "The Impact of Feature Scaling in Machine Learning: Effects on Regression and Classification Tasks," *arXiv*, 2025.
- [8] I. Maulita, C. R. A. Widiawati, dan A. M. Wahid, "Analisis Komparatif Linear Regression, Random Forest, dan Gradient Boosting untuk Prediksi Banjir," *Jurnal Pendidikan dan Teknologi Indonesia*, 2025
- [9] A. Fahmi, M. H. Munsarif, dan A. Winarno, "Evaluasi Komparatif Model Regresi Prediksi Saham BBCA," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 7, no. 2, 2025.
- [10] B. Kriswantara, K. Kurniawati, dan H. F. Pardede, "Prediksi Harga Mobil Bekas dengan Machine Learning," *SLJIL*, vol. 6, no. 5, 2021.
- [11] C. Fan, A. Upadhye, dan A. Worster, "The Coefficient of Determination R-squared is More Informative Than SMAPE, MAE, MAPE, MSE and RMSE in Regression Analysis Evaluation," *PeerJ Comput. Sci.*, 2025
- [12] Mahim A. M. Billah dan T. Sarkar, "Comparative Analysis of Machine Learning Models for House Price Prediction: Linear Regression to Boosted Trees," *ResearchGate*, 2025.