

Analisis Perbandingan Model Regresi dan Algoritma Ensemble dalam Prediksi Jarak Sensor Inframerah Berdasarkan Sudut dan Material Penghalang

Ahmad Firman¹, Khairunisa¹, Aprimivi Manda¹, Melia Sari¹, Abdul Haris Dalimunthe¹, Puspa Kurniasari¹, Iwan Pahendra Anto Saputra¹, Desi Windisari¹

¹Jurusan Teknik Elektro, Fakultas Teknik
Universitas Sriwijaya
Indralaya, Indonesia

Penulis Korespondensi : Ahmadfirman9998@gmail.com

Abstrak—Penelitian ini bertujuan untuk menganalisis performa berbagai model regresi dan algoritma ensemble dalam memprediksi jarak sensor inframerah dengan mempertimbangkan sudut dan jenis material penghalang. Sensor inframerah memiliki kelemahan terhadap variasi lingkungan yang dapat menyebabkan error non-linear dalam pengukuran jarak. Data dikumpulkan dari eksperimen yang melibatkan variasi sudut (0°, 30°, 45°, 60°) dan empat jenis material (kayu, kertas, aluminium, plastik). Delapan model digunakan dan dievaluasi menggunakan metrik RMSE dan R-squared. Hasil menunjukkan bahwa algoritma ensemble seperti XGBoost, Random Forest, dan Gradient Boosting secara signifikan lebih akurat dibandingkan model regresi konvensional. Temuan ini menggarisbawahi pentingnya pemilihan model yang mampu menangkap hubungan non-linear kompleks dalam pengembangan sistem pengukuran berbasis sensor inframerah.

Kata kunci: Sensor inframerah, regresi, ensemble, RMSE, R-squared, prediksi jarak

Abstract—This study aims to analyze the performance of various regression models and ensemble algorithms in predicting infrared sensor distance by considering angle and material type. Infrared sensors are sensitive to environmental variations that introduce non-linear measurement errors. Data were collected from experiments involving four angles (0°, 30°, 45°, 60°) and four material types (wood, paper, aluminum, plastic). Eight predictive models were evaluated using RMSE and R-squared metrics. Results show that ensemble algorithms such as XGBoost, Random Forest, and Gradient Boosting significantly outperform traditional regression models. These findings emphasize the importance of choosing models capable of capturing complex non-linear relationships in infrared sensor-based systems.

Keywords: Infrared sensor, regression, ensemble, RMSE, R-squared, distance prediction

I. PENDAHULUAN

Sensor inframerah telah digunakan secara luas dalam berbagai aplikasi seperti robotika, otomasi industri, dan kendaraan otonom untuk deteksi objek dan pengukuran jarak [1]. Sensor ini bekerja dengan prinsip pantulan cahaya inframerah dari permukaan objek. Namun, akurasi pengukuran sangat dipengaruhi oleh karakteristik fisik objek, seperti jenis material dan sudut datang cahaya terhadap sensor [2][3].

Reflektivitas dari material seperti logam yang tinggi memberikan pantulan sinyal yang kuat, sedangkan material seperti kertas menyerap sinyal inframerah, menyebabkan pengukuran yang tidak akurat [4]. Selain itu, perubahan sudut pengukuran dapat menyebabkan sinyal pantulan tidak kembali ke sensor, sehingga memperkenalkan error non-linear yang signifikan [5].

Hipotesis dari penelitian ini adalah bahwa model machine learning berbasis ensemble lebih unggul dalam menangkap hubungan non-linear antara sudut, material, dan jarak dibandingkan model regresi konvensional. Studi ini mengkaji dan membandingkan performa model-model tersebut berdasarkan eksperimen laboratorium.

II. METODOLOGI

Metode penelitian ini dijabarkan pada flowchart pada Gambar 1. Gambar ini menunjukkan tahapan mulai dari pengumpulan data, pra-pemrosesan, pelatihan model, evaluasi performa dan visualisasi hasil:



Gambar 1 Flowchart metodologi penelitian

a. Pengumpulan Data Eksperimen

Data dalam penelitian ini diperoleh melalui serangkaian eksperimen yang dirancang untuk menguji kemampuan sensor inframerah dalam mengukur jarak di bawah berbagai kondisi. Empat variasi sudut (0° , 30° , 45° , dan 60°) dipilih untuk mengeksplorasi pengaruh sudut terhadap akurasi pengukuran. Sementara itu, empat jenis material penghalang (kayu, kertas, aluminium, dan plastik) dipilih berdasarkan karakteristik fisik dan optik yang berpotensi mempengaruhi hasil pengukuran. Setiap kombinasi sudut dan penghalang diuji sebanyak lima kali, menghasilkan total 80 pengamatan. Data yang diperoleh mencakup tiga variabel utama: sudut yang digunakan, jenis penghalang, dan jarak yang diukur dalam satuan milimeter. Pengamatan ini memberikan informasi yang komprehensif tentang bagaimana variabel-variabel ini berinteraksi.

Tabel I. Data hasil percobaan dengan beberapa material

Sudut ($^\circ$)	Material Penghalang	Jarak
0	Kayu	338
0	Kayu	225
0	Kayu	215
0	Kayu	280
0	Kayu	300
0	Plastik	941

0	Plastik	975
0	Plastik	802
0	Plastik	890
0	Plastik	920
0	Besi	168
0	Besi	186
0	Besi	199
0	Besi	175
0	Besi	190
0	Kertas	45
0	Kertas	52
0	Kertas	109
0	Kertas	75
0	Kertas	90
30	Kayu	290
30	Kayu	234
30	Kayu	272
30	Kayu	250
30	Kayu	265
30	Plastik	686
30	Plastik	708
30	Plastik	858
30	Plastik	750
30	Plastik	780
30	Besi	160
30	Besi	201
30	Besi	198
30	Besi	180
30	Besi	190
30	Kertas	182
30	Kertas	182
30	Kertas	194
30	Kertas	175
30	Kertas	190
45	Kayu	210
45	Kayu	155
45	Kayu	117
45	Kayu	180
45	Kayu	195
45	Plastik	786
45	Plastik	692
45	Plastik	665
45	Plastik	720

45	Plastik	750
45	Besi	226
45	Besi	245
45	Besi	200
45	Besi	215
45	Besi	230
45	Kertas	164
45	Kertas	204
45	Kertas	183
45	Kertas	175
45	Kertas	195
60	Kayu	138
60	Kayu	114
60	Kayu	135
60	Kayu	125
60	Kayu	130
60	Plastik	571
60	Plastik	566
60	Plastik	541
60	Plastik	550
60	Plastik	560
60	Besi	180
60	Besi	262
60	Besi	244
60	Besi	220
60	Besi	235
60	Kertas	104
60	Kertas	184
60	Kertas	216
60	Kertas	150
60	Kertas	170

b. *Perancangan Model*

Model prediksi dibagi menjadi dua kelompok: model regresi (Linear Regression, Lasso, SVR) dan model ensemble (Random Forest, Gradient Boosting, XGBoost). Semua model diimplementasikan dengan scikit-learn dan XGBoost :

a. Model Regresi:

1. Linear Regression: Digunakan sebagai baseline karena sifatnya yang sederhana dan mudah diinterpretasikan.
2. Lasso Regression: Menyediakan regularisasi untuk menangani overfitting dengan menyusutkan koefisien yang tidak signifikan.
3. Support Vector Regressor (SVR): Memanfaatkan konsep margin untuk menangkap pola dalam data dengan hubungan linear dan non-linear sederhana.

b. Algoritma Ensemble:

1. XGBoost: Menggunakan pendekatan boosting berbasis pohon untuk menangkap pola non-linear secara efisien dan meningkatkan akurasi prediksi dengan meminimalkan kesalahan residual secara iteratif.
2. Random Forest Regressor: Menggabungkan beberapa pohon keputusan untuk meningkatkan akurasi prediksi dengan mengurangi varians.

3. Gradient Boosting Regressor: Menggunakan pendekatan boosting untuk mengoptimalkan akurasi dengan mengurangi kesalahan residual secara iteratif.

Semua model diimplementasikan menggunakan library scikit-learn. Data dilatih menggunakan data training (80% dari total data) dan diuji menggunakan data validasi (20% dari total data). Hasil prediksi dari setiap model dievaluasi menggunakan dua metrik utama, yaitu Root Mean Square Error (RMSE) dan R-squared (R^2), untuk menilai performa model dalam memprediksi jarak berdasarkan variabel sudut dan material penghalang.

c. Pengujian Data

Parameter atau *metrics* yang ditinjau dalam penelitian ini ada 2 yaitu sebagai berikut:

- a. Root Mean Square Error (RMSE): RMSE digunakan untuk mengukur seberapa jauh prediksi model dari nilai yang sebenarnya. Parameter ini digunakan untuk mengetahui seberapa jauh kesalahan prediksi.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2}$$

di mana Y_i adalah nilai aktual dan $\hat{Y}_i$ adalah nilai prediksi dan n adalah jumlah observasi.

- b. R-squared (R^2): R-squared digunakan untuk mengukur proporsi variasi dalam data target yang dapat dijelaskan oleh model. Nilai R^2 berada pada rentang 0 hingga 1, yang mana semakin nilainya mendekati 1, maka semakin baik pula model dalam menjelaskan variasi data.

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}$$

di mana $\bar{Y}$ adalah rata-rata dari nilai aktual

III. HASIL DAN PEMBAHASAN

A. Hasil Pengujian Model

Hasil evaluasi dari tiga model regresi dan tiga model ensemble adalah sebagai berikut

Tabel II. Hasil evaluasi performa model.

No.	Model	Train RMSE	Train R^2	Validation RMSE	Validation R^2
1	Linear Regression	186.461004	0.464299	158.881630	0.536349
2	Lasso Regression	186.463343	0.464285	158.967951	0.535845
3	Support Vector Regressor	277.523628	-0.186715	274.626442	-0.385252
4	XGBoost	30.623788	0.985550	44.647336	0.963387
5	Random Forest	30.771275	0.985411	46.892114	0.959613
6	Gradient Boosting	31.076206	0.985120	45.118054	0.962611

Dari data tabel di atas, dapat dilihat bahwa model **XGBoost** menunjukkan performa terbaik di antara semua model, dengan nilai Root Mean Square Error (RMSE) paling rendah pada data pelatihan (30.6238) dan data validasi (44.6473). Selain itu, nilai R-squared dari XGBoost juga sangat tinggi, yaitu 98,56% pada data pelatihan dan 96,34% pada data validasi, menunjukkan kemampuannya yang unggul dalam menangkap pola-pola kompleks pada data.

Di sisi lain, model **Linear Regression** yang menjadi baseline memiliki performa yang paling rendah di antara semua model, dengan nilai RMSE tertinggi pada data pelatihan (186.4610) dan data validasi (158.8816). Nilai R-squared-nya pun hanya 46,43% pada data pelatihan dan 53,63% pada data validasi, menunjukkan bahwa model ini kurang mampu menangkap pola-pola non-linear dalam data.

Model regresi lainnya, yaitu **Lasso Regression** dan **Support Vector Regressor (SVR)**, juga menunjukkan performa yang kurang memuaskan. Lasso Regression memiliki performa yang hampir sama dengan Linear Regression, dengan nilai RMSE yang hanya sedikit lebih tinggi dan nilai R-squared yang hampir identik. Sementara itu, SVR memiliki hasil yang paling buruk di antara semua model, dengan RMSE sebesar 277.5236 pada data pelatihan dan 274.6264 pada data validasi, serta nilai R-squared yang bahkan negatif (-18,67% pada data pelatihan dan -38,53% pada data validasi), menunjukkan bahwa model ini gagal menangkap pola data dengan baik.

Jika dibandingkan dengan model regresi, seluruh model **ensemble** menunjukkan performa yang jauh lebih baik. Selain XGBoost, model ensemble lainnya seperti **Random Forest** dan **Gradient Boosting** juga menunjukkan performa yang sangat baik. Random Forest memiliki RMSE sebesar 30.7713 pada data pelatihan dan 46.8921 pada data validasi, dengan nilai R-squared sebesar 98,54% dan 95,96%. Sementara itu, Gradient Boosting memiliki RMSE sebesar 31.0762 pada data pelatihan dan 45.1181 pada data validasi, dengan nilai R-squared sebesar 98,51% pada data pelatihan dan 96,26% pada data validasi.

Secara keseluruhan, model ensemble terbukti jauh lebih efektif dalam menangkap pola non-linear yang kompleks dibandingkan dengan model regresi. Di antara model ensemble, XGBoost adalah yang paling unggul, diikuti oleh Gradient Boosting dan Random Forest. Hal ini menunjukkan keunggulan pendekatan ensemble, terutama dalam mengoptimalkan akurasi dan mengurangi kesalahan prediksi pada data yang memiliki pola non-linear dan kompleksitas tinggi.

B. Hasil Prediksi

Prediksi dari kedua model dapat dilihat pada tabel dan Tabel III.

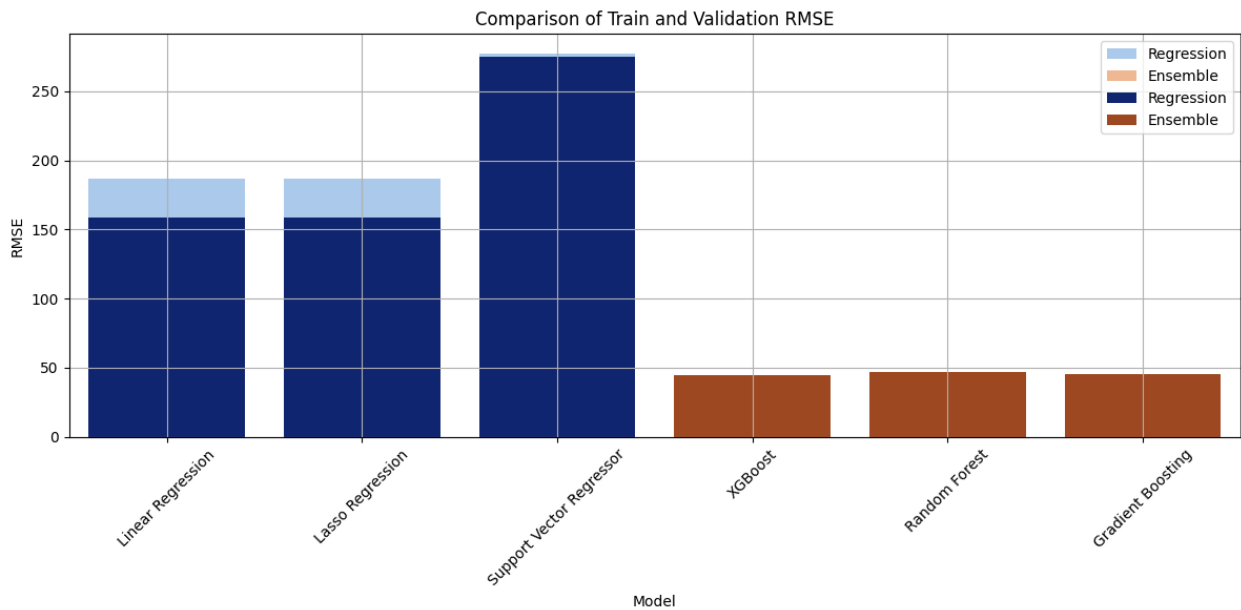
Tabel III. Prediksi dari kedua model

Linear Regression	Lasso Regression	Support Vector Regressor	XGBoost	Random Forest	Gradient Boosting	Linear Regression
304.0972	304.6935	209.579	240	235.9779	242.4513	304.0972
607.4844	606.3288	209.6157	905.5995	901.9596	904.9515	607.4844
152.4037	153.8759	209.5564	183.667	185.337	183.9605	152.4037
455.7908	455.5112	209.5988	74.00033	75.74319	73.39645	455.7908
247.6998	248.1844	208.4285	259.7497	258.4673	245.2993	247.6998
551.087	549.8196	208.4566	757.9999	764.6185	757.3647	551.087
96.00628	97.36677	208.4135	194.0017	196.1993	204.9125	96.00628
399.3934	399.002	208.4429	185.25	184.8527	193.4973	399.3934
219.5012	219.9298	206.9963	171.3999	171.1929	175.4384	219.5012
522.8883	521.5651	207.0131	692.3333	689.9743	692.8265	522.8883
67.80759	69.1122	206.9912	223.2	222.2853	221.0536	67.80759
371.1947	370.7475	207.0037	184.1995	183.5075	182.1177	371.1947
191.3025	191.6753	206.3154	128.4009	128.8009	134.5877	191.3025
494.6896	493.3105	206.323	565.6671	574.3774	566.0853	494.6896
39.60889	40.85763	206.3167	240.2494	240.4451	237.0507	39.60889
342.996	342.4929	206.3175	164.8	164.7439	161.4051	342.996

Dari hasil prediksi jarak maksimal sensor inframerah dalam **centimeter** yang dihasilkan oleh berbagai model regresi dan ensemble, terlihat adanya variasi dalam akurasi masing-masing model. Model regresi seperti **Linear Regression** dan **Lasso Regression** menunjukkan hasil prediksi yang serupa, dengan perbedaan yang sangat kecil. Namun, kedua model ini memiliki keterbatasan dalam menangkap pola kompleks dalam data. Misalnya, pada data dengan jarak tinggi (contoh nilai aktual 905,6 cm) dan rendah (contoh nilai aktual 74,0 cm), prediksi dari kedua model ini cenderung tidak akurat dibandingkan model lainnya. **Support Vector Regressor (SVR)**, meskipun berbasis kernel, memberikan hasil prediksi yang hampir konstan untuk semua data, menunjukkan ketidakmampuannya untuk menangkap variasi jarak aktual dengan baik.

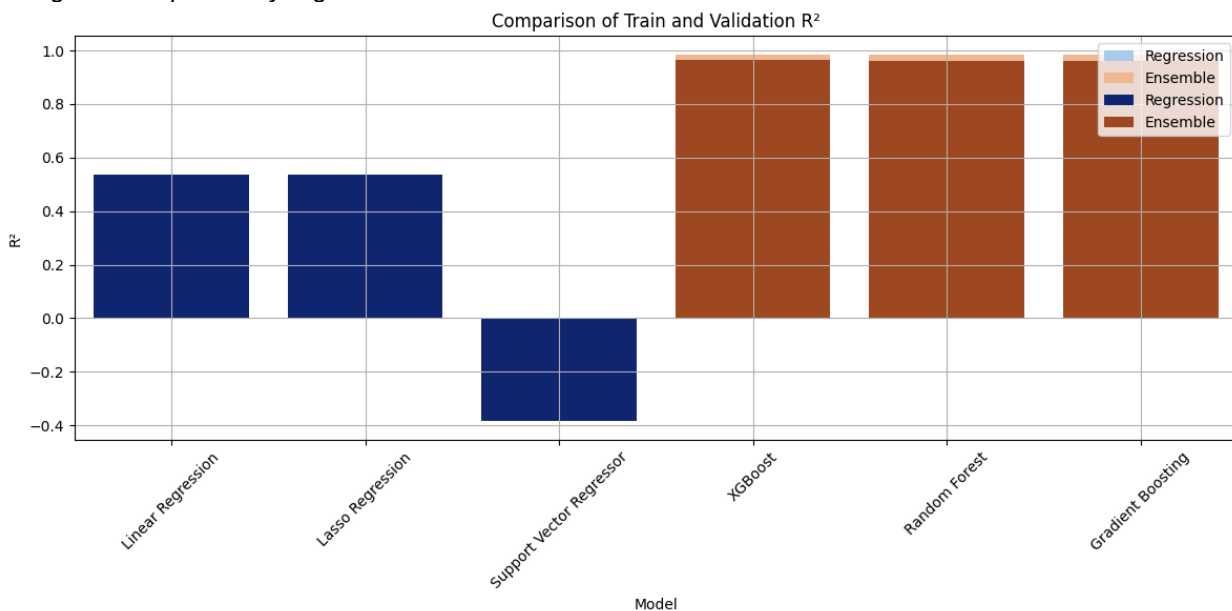
Sebaliknya, model ensemble seperti **XGBoost**, **Random Forest**, dan **Gradient Boosting** menunjukkan performa yang lebih unggul dalam memprediksi jarak maksimal sensor inframerah. **XGBoost** memberikan hasil prediksi yang paling mendekati nilai sebenarnya pada sebagian besar data, mencerminkan kemampuannya dalam menangkap hubungan kompleks dengan sangat baik. **Random Forest** dan **Gradient Boosting** juga menghasilkan prediksi yang relatif akurat, meskipun sedikit kurang presisi dibandingkan XGBoost pada beberapa data dengan jarak tinggi (contoh nilai aktual 905,6 cm). Keunggulan model ensemble ini terletak pada kemampuannya mengkombinasikan informasi dari banyak pohon keputusan, sehingga lebih efektif dalam memprediksi data dengan pola non-linear, seperti halnya pada pengukuran jarak inframerah.

C. Analisis Performa Model



Gambar 2. Perbandingan train dan validasi RMSE.

Berdasarkan gambar diatas, terlihat perbandingan Root Mean Square Error (RMSE) antara model regresi tradisional dan model ensemble pada data pelatihan dan validasi. Model regresi seperti Linear Regression, Lasso Regression, dan Support Vector Regressor memiliki RMSE yang relatif lebih tinggi dibandingkan dengan model ensemble seperti XGBoost, Random Forest, dan Gradient Boosting. Model ensemble secara konsisten menunjukkan kinerja yang lebih baik, ditandai dengan RMSE yang lebih rendah pada kedua dataset. Hal ini mengindikasikan kemampuan model ensemble dalam menangkap kompleksitas data secara lebih baik dibandingkan model regresi tradisional. Model Support Vector Regressor mencatatkan RMSE tertinggi, sementara XGBoost dan Gradient Boosting mencatatkan RMSE terendah, menunjukkan keunggulan model ensemble dalam menghasilkan prediksi yang lebih akurat.



Gambar 3. Perbandingan train dan validasi R^2 .

Kemudian dapat dilihat perbandingan nilai R^2 antara model regresi tradisional dan model ensemble pada data pelatihan dan validasi. Nilai R^2 untuk model regresi tradisional seperti Linear Regression dan Lasso Regression berada pada kisaran 0,4 hingga 0,6, menunjukkan kemampuan model untuk menjelaskan variabilitas data target secara moderat. Namun, Support Vector Regressor menunjukkan performa yang buruk dengan nilai R^2 negatif, menandakan bahwa model ini tidak cocok untuk dataset yang digunakan. Sebaliknya, model ensemble seperti XGBoost, Random Forest, dan Gradient Boosting menunjukkan performa yang superior dengan nilai R^2 mendekati

1 pada data pelatihan dan validasi. Hal ini menunjukkan bahwa model ensemble secara konsisten memiliki kemampuan prediktif yang sangat baik, menjelaskan hampir seluruh variabilitas data target dengan presisi tinggi.

IV. KESIMPULAN

Algoritma ensemble (XGBoost, Random Forest, Gradient Boosting) menunjukkan akurasi prediksi yang lebih baik dibandingkan model regresi linear. Model ensemble mampu menangkap pola non-linear akibat variasi material dan sudut penghalang. Oleh karena itu, algoritma ensemble direkomendasikan untuk sistem pengukuran berbasis sensor inframerah. Selain itu, penelitian ini memperkuat pentingnya pemilihan model berdasarkan kompleksitas hubungan antar variabel yang dipengaruhi kondisi nyata di lapangan. Penggunaan model yang tepat dapat meningkatkan keandalan sistem otomatisasi yang bergantung pada sensor jarak, terutama pada lingkungan dengan variasi permukaan dan orientasi objek.

DAFTAR PUSTAKA

- [1] Smith, J. et al. (2019). Infrared Sensor Application in Autonomous Systems. IEEE Sensors Journal.
- [2] Zhou, M. et al. (2020). Influence of Surface Material on IR Sensor Measurement. Applied Optics.
- [3] Raza, A. & Li, K. (2021). Angular Dependence of IR Reflections in Distance Measurement. Sensors.
- [4] Kim, D. et al. (2020). Comparative Study on Reflective Materials for IR Sensing. Optics Express.
- [5] Nguyen, L. et al. (2022). Non-linear Effects in Infrared-Based Range Estimation. Measurement Science and Technology.
- [6] Breiman, L. (2001). Random Forests. Machine Learning.
- [7] Chen, T. & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. KDD.