

## Perbandingan Kinerja Model Prediksi Cuaca: Random Forest, Support Vector Regression, dan XGBoost

Ahmad Syahreza<sup>1,\*</sup>, Novita Kurnia Ningrum<sup>1</sup>, Muhammad Anjas Syahrazy<sup>1</sup>

<sup>1</sup> Program Studi Informatika, Universitas Dian Nuswantoro, Indonesia

\* Correspondence: zamysyah@gmail.com

**Copyright:** © 2024 by the authors

Received: 1 Oktober 2024 | Revised: 4 Oktober 2024 | Accepted: 25 Oktober 2024 | Published: 19 Desember 2024

### Abstrak

Prediksi cuaca yang akurat sangat penting untuk memitigasi dampak perubahan cuaca dan mendukung perencanaan yang lebih baik di sektor-sektor seperti pertanian, transportasi, dan pariwisata. Indonesia sering menghadapi cuaca yang tidak terduga, seperti hujan mendadak dan kekeringan panjang, yang dapat menimbulkan kerugian besar. Penelitian ini bertujuan untuk membandingkan kinerja tiga algoritma pembelajaran mesin *Random Forest*, *Support Vector Regression (SVR)*, dan *XGBoost* dalam memprediksi cuaca menggunakan data meteorologi (suhu minimum, suhu maksimum, curah hujan, arah angin, kelembaban rata-rata) serta data IoT yang berjumlah 1650 data per variable. Variabel yang digunakan dalam penelitian ini meliputi suhu minimum, suhu maksimum, curah hujan, arah angin, dan kelembaban rata-rata. Teknik analisis data dilakukan dengan menggunakan tiga metrik evaluasi utama yaitu *Mean Absolute Error (MAE)*, *Mean Squared Error (MSE)*, dan *R-squared (R<sup>2</sup>)*. Hasil penelitian menunjukkan bahwa *XGBoost* memberikan performa terbaik dengan nilai MAE 0,3744, MSE 0,2278, dan R<sup>2</sup> 0,8183. *Random Forest* dan *SVR* juga menghasilkan prediksi yang baik, dengan nilai MAE masing-masing 0,3869 dan 0,3820, MSE 0,2422 dan 0,2524, serta R<sup>2</sup> 0,8068 dan 0,7987. Hasil penelitian menunjukkan *XGBoost* adalah model terbaik untuk prediksi cuaca, yang dapat membantu meningkatkan ketepatan dalam perencanaan pertanian dan mitigasi risiko bencana terkait cuaca.

**Kata kunci:** *machine learning; suhu; support vector regression; random forest; xgboost*

### Abstract

Accurate weather predictions are essential to mitigate the impacts of weather changes and support better planning in sectors such as agriculture, transportation, and tourism. Indonesia often faces unpredictable weather, such as sudden rains and long droughts, which can cause huge losses. This study aims to compare the performance of three machine learning algorithms *Random Forest*, *Support Vector Regression (SVR)*, and *XGBoost* in predicting weather using meteorological data (minimum temperature, maximum temperature, rainfall, wind direction, average humidity) as well as IoT data totaling 1650 data per variable. The variables used in this study include minimum temperature, maximum temperature, rainfall, wind direction, and average humidity. Data analysis techniques were performed using three main evaluation metrics, namely *Mean Absolute Error (MAE)*, *Mean Squared Error (MSE)*, and *R-squared (R<sup>2</sup>)*. The results showed that *XGBoost* gave the best performance with MAE 0.3744, MSE 0.2278, and R<sup>2</sup> 0.8183. *Random Forest* and *SVR* also produced good predictions, with MAE values of 0.3869 and 0.3820, MSE 0.2422 and 0.2524, and R<sup>2</sup> 0.8068 and 0.7987, respectively. The results show *XGBoost* is the best model for weather prediction, which can help improve accuracy in agricultural planning and weather-related disaster risk mitigation.

**Keywords:** *machine learning; temperature; support vector regression; random forest; xgboost*



## PENDAHULUAN

Perubahan iklim telah menimbulkan ketidakpastian cuaca yang berdampak signifikan pada sektor pertanian, khususnya dalam produksi pangan dan kesejahteraan petani. Cuaca yang tidak menentu mengakibatkan tantangan besar dalam perencanaan musim tanam, penggunaan air, dan mitigasi risiko kerugian ekonomi. Dengan adanya prediksi cuaca yang akurat, petani dapat mengoptimalkan pengelolaan sumber daya dan meningkatkan efisiensi pertanian (Satria et al., 2023). Ketidakpastian cuaca akibat perubahan iklim ini semakin mempertegas pentingnya teknologi prediksi cuaca yang mampu memberikan informasi yang tepat waktu dan akurat (Putri et al., 2023).

Ketidakpastian cuaca yang meningkat menuntut solusi berbasis data yang dapat mendukung prediksi cuaca secara lebih presisi. Penelitian ini menawarkan solusi dengan membandingkan tiga model pembelajaran mesin, yaitu *Random Forest*, *Support Vector Regression* (SVR), dan *XGBoost*, untuk memprediksi cuaca. Model pembelajaran mesin ini memiliki potensi besar dalam memanfaatkan data cuaca historis serta data *real-time* dari sensor IoT guna membangun prediksi yang lebih akurat (Himawan et al., 2021).

*Random forest* bekerja dengan membangun beberapa pohon keputusan dan menggabungkan hasilnya untuk meningkatkan stabilitas prediksi serta mengurangi overfitting (Priyatno et al., 2023). SVR menggunakan *hyperplane* dalam ruang dimensi tinggi untuk meminimalkan kesalahan dalam margin tertentu, membuatnya kuat dalam menangani data berdimensi tinggi (Setiawan & Suryono, 2024). Sementara itu, *XGBoost* menawarkan keunggulan dalam efisiensi dan akurasi berkat teknik ensemble learning, yang mampu menangkap pola-pola kompleks dalam dataset (Muhammady et al., 2024).

Penelitian sebelumnya yang dilakukan oleh (Sun et al., 2023) dan (Hadianto & Utomo, 2024) menunjukkan keunggulan setiap model di konteks yang berbeda. *Random Forest* dianggap efektif dalam menangani data yang berisik. Istilah data yang berisik (*noisy data*) merujuk pada data yang mengandung kesalahan atau gangguan yang dapat mengaburkan pola atau tren sebenarnya yang sedang dianalisis (Bala, 2021), sementara SVR sering unggul dalam data berdimensi tinggi (Li et al., 2023). Namun, SVR memiliki keterbatasan dalam efisiensi komputasi, terutama pada dataset yang besar (Pamungkas & Cahyono, 2024) SVR memiliki kelemahan dalam kecepatan komputasi, terutama pada dataset besar, karena memerlukan perhitungan intensif untuk menentukan *hyperplane* optimal. (Patasik & Yulianto, 2023) menyoroti efisiensi dan akurasi *XGBoost* dalam menangani dataset cuaca yang kompleks, sedangkan (Huang et al., 2022). *XGBoost* dikenal karena kemampuannya menangani dataset cuaca yang kompleks, meskipun memerlukan tuning parameter yang lebih intensif. Penelitian terbaru juga mendukung penggunaan *XGBoost* untuk memprediksi pola cuaca jangka panjang (Pebrianti et al., 2023).

Meskipun telah banyak penelitian tentang prediksi cuaca, masih sedikit yang secara langsung membandingkan kinerja *random forest*, SVR, dan *XGBoost* dalam satu penelitian. Penelitian oleh (Taqiyuddin & Bayu Sasongko, 2024) hanya menggunakan *random forest* tanpa pembanding model lain, sehingga perspektif terkait model terbaik menjadi terbatas. Selain itu, (Karim et al., 2023) hanya membandingkan *random forest* dengan SVM dan KNN, tetapi tidak mengeksplorasi model ensemble seperti *XGBoost*. Penelitian yang menerapkan *XGBoost*, seperti (Ardian et al., 2024), juga tidak menyertakan SVR sebagai pembanding. Penelitian ini mengisi celah dengan membandingkan tiga model utama, yaitu *random forest*, SVR, dan *XGBoost*, yang unggul dalam menangani data cuaca yang kompleks (Rakhmat & Mutohar, 2023). *Random forest* efektif mengelola data berisik, SVR unggul dalam menangkap pola non-linear pada data berdimensi tinggi, dan *XGBoost* efisien untuk dataset besar dengan akurasi tinggi. Meskipun setiap model memiliki keunggulan, penelitian yang membandingkan ketiganya dalam prediksi cuaca dengan evaluasi menyeluruh masih jarang dilakukan (Hutagalung et al., 2023).

Penelitian ini sangat penting karena ketidakpastian cuaca akibat perubahan iklim semakin menuntut adanya prediksi yang lebih akurat untuk membantu sektor pertanian. Pemanfaatan teknologi IoT yang semakin berkembang juga membutuhkan sistem prediksi cuaca yang dapat mengintegrasikan data historis dan *real-time* untuk membantu petani merencanakan dengan lebih baik dan mengurangi risiko akibat cuaca yang tidak menentu. Penelitian ini bertujuan untuk membandingkan kinerja tiga model pembelajaran mesin, yaitu *random forest*, SVR, dan *XGBoost*, dalam memprediksi cuaca menggunakan data cuaca historis dan data real-time dari teknologi IoT.

## METODE

Penelitian ini melalui enam tahapan utama: data *selection*, *preprocessing*, *transformation*, split data, modeling, dan evaluation. Data dikumpulkan dari situs BMKG dan sensor IoT DHT11 yang terhubung ke aplikasi Blynk, mencakup variabel cuaca yang relevan. Data yang diperoleh kemudian diolah melalui tahapan preprocessing, seperti normalisasi dan pembersihan data. Setelah itu, data diubah menjadi format yang sesuai untuk model, dibagi menjadi data latih dan uji, serta diterapkan tiga model pembelajaran mesin, yaitu *random forest*, SVR, dan *XGBoost*.

**Tabel 1.** Data variabel

Variabel		Tipe data	Jumlah data
Notasi	Nama		
X1	Temperature Min (Tn)	Numerik	1650
X2	Temperature Max (Tx)	Numerik	1650
X3	Humidity (RH_avg)	Numerik	1650
X4	Rainfall (RR)	Numerik	1650
X5	Sunlight (ss)	Numerik	1650
X6	Wind direction (ff_x)	Numerik	1650
X7	Wind velocity (ff_avg)	Numerik	1650
Y	Average temp (Tavg)	Numerik	1650

Data yang digunakan dalam penelitian ini diambil dari website resmi BMKG dan sensor IoT DHT11, data dikumpulkan dari Januari 2020 hingga Agustus 2024, menghasilkan total 1650 data untuk setiap variabel. Pada tabel 1 adalah variabel-variabel yang digunakan pada penelitian ini dan terdiri dari suhu minimum (Tn), suhu maksimum (Tx), kelembaban rata-rata (RH\_avg), curah hujan (RR), durasi sinar matahari (ss), arah angin (ff\_x), kecepatan angin (ff\_avg), dan suhu rata-rata (Tavg) sebagai variabel target.

Analisis Data Eksploratif (EDA) dilakukan untuk memahami struktur dan karakteristik dataset sebelum pemodelan dilakukan. EDA dalam penelitian ini dilakukan untuk mengidentifikasi pola musiman dan tren jangka panjang untuk memahami fluktuasi suhu berdasarkan data historis. Distribusi setiap variabel, seperti suhu maksimum, suhu minimum, kelembaban, dan lainnya, diperiksa untuk mendeteksi anomali atau *outlier*. Pendeteksian *outlier* ini bertujuan untuk mengidentifikasi data yang mungkin mengganggu kinerja model prediksi jika tidak diatasi. Selain itu, dilakukan pemeriksaan data yang hilang, dimana metode *imputasi* seperti rata-rata atau *interpolasi* digunakan untuk mengisi data yang tidak lengkap. EDA juga membantu mengidentifikasi korelasi antar variabel input dan target, memberikan wawasan lebih lanjut tentang hubungan antara variabel seperti curah hujan atau kelembaban dengan suhu rata-rata harian.

*Cross-validation* digunakan untuk membagi dataset menjadi dua bagian, bagian pertama digunakan untuk mempelajari data latih untuk melatih model, sementara yang lainnya digunakan untuk menguji dan mengvalidasi model. Teknik *K-Fold Cross-validation* dengan

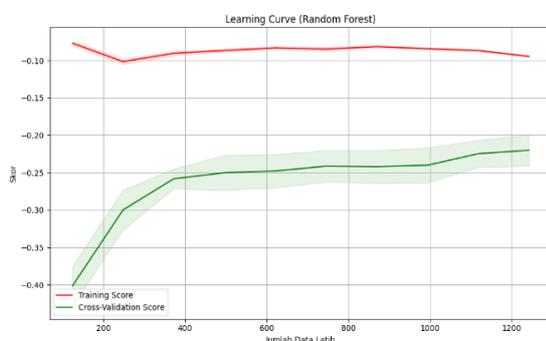
$k=10$  memastikan bahwa model tidak mengalami *overfitting* dan dapat memberikan performa yang konsisten di berbagai subset data. Teknik ini membagi data menjadi 10 subset (fold), di mana 9 subset digunakan untuk melatih model dan 1 subset digunakan untuk menguji model. Proses ini diulang 10 kali dengan setiap subset bergantian menjadi subset pengujian.

Penelitian ini menggunakan tiga algoritma pembelajaran mesin, yaitu *random forest*, SVR, dan *XGBoost*, dengan proses optimasi parameter melalui *GridSearchCV* untuk meningkatkan akurasi prediksi. Evaluasi model dilakukan menggunakan tiga metrik utama: *Mean Absolute Error* (MAE), jika hasil mendekati angka 0 maka performa model semakin baik; *Mean Squared Error* (MSE), sama dengan MAE jika hasil mendekati 0 maka performa model semakin baik, sehingga efektif mendeteksi prediksi dengan *error* besar; dan *R<sup>2</sup> Score*, yang menunjukkan seberapa baik model dapat menjelaskan variabilitas data, dimana nilai mendekati 1 model menunjukkan prediksi yang sangat akurat.

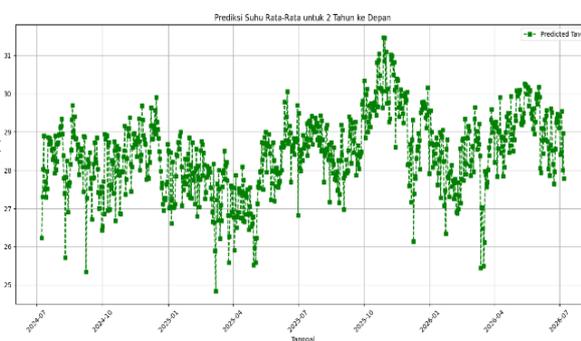
## HASIL DAN PEMBAHASAN

### Hasil

*Learning curve* pada *random forest* yang disajikan pada gambar 1 menunjukkan bahwa model memiliki kemampuan adaptasi yang cepat terhadap data. Sumbu horizontal menggambarkan jumlah data latih, sedangkan sumbu vertikal menunjukkan skor yang dicapai. Garis merah yang mewakili *training score* tetap stabil setelah jumlah data latih mencapai lebih dari 200, menunjukkan bahwa model dapat dengan baik mempelajari pola dari data tersebut. Garis hijau pada grafik adalah *cross-validation score*, yang awalnya berada di bawah skor pelatihan, tetapi terus meningkat seiring bertambahnya data. Ini menandakan bahwa model bisa bekerja lebih baik dalam generalisasi jika jumlah data diperbesar. Meskipun ada peningkatan pada skor validasi, masih ada jarak dengan skor pelatihan, menunjukkan potensi *overfitting* pada data yang lebih kecil.



Gambar 1. RF learning curve

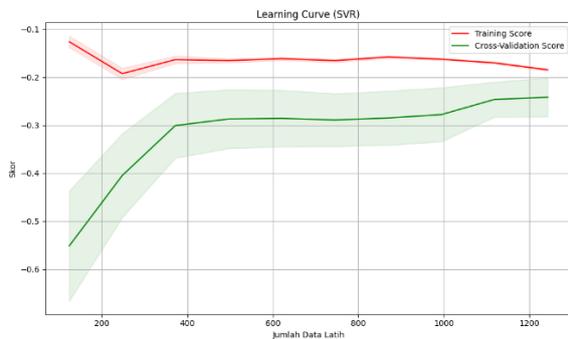


Gambar 2. Hasil prediksi random forest

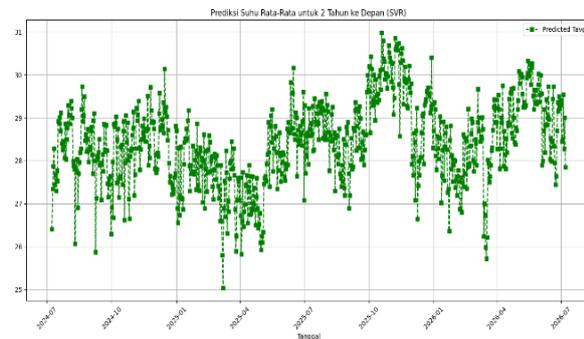
*Random forest* menghasilkan grafik prediksi pada gambar 2 yang menghasilkan nilai MAE sebesar 0,3869, MSE sebesar 0,2422, dan  $R^2$  sebesar 0,8068. Nilai MAE mengindikasikan bahwa kesalahan rata-rata prediksi model cukup rendah. MSE yang lebih kecil menunjukkan bahwa model memiliki tingkat kesalahan kuadrat yang lebih kecil, yang baik untuk meminimalkan outlier. Sementara itu, nilai  $R^2$  yang mendekati 1 menunjukkan bahwa model ini mampu menjelaskan sebagian besar variabilitas dalam data, tetapi masih ada ruang untuk perbaikan terutama dalam generalisasi.

*Learning curve* pada SVR yang disajikan pada gambar 3 menunjukkan tren serupa dengan Random Forest, tetapi performa model pada validasi silang sedikit lebih rendah. Pada awal pelatihan, skor validasi lebih rendah dari skor pelatihan dan secara bertahap naik seiring dengan bertambahnya data. Ini menunjukkan bahwa SVR membutuhkan lebih banyak data pelatihan untuk meningkatkan generalisasi. Garis merah yang mewakili *Training Score* menunjukkan bahwa model menyesuaikan data dengan baik pada fase pelatihan, tetapi garis

hijau yang menunjukkan *cross-validation score* menunjukkan bahwa generalisasi model tetap berada di bawah skor pelatihan. Sumbu horizontal di grafik menunjukkan jumlah data latih, dan sumbu vertikal menunjukkan tingkat kesalahan atau skor model pada proses pelatihan dan validasi.



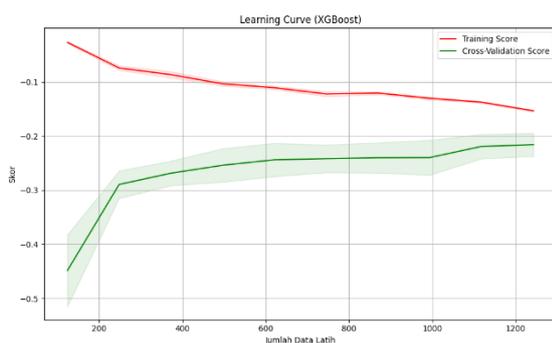
**Gambar 3.** SVR *learning curve*



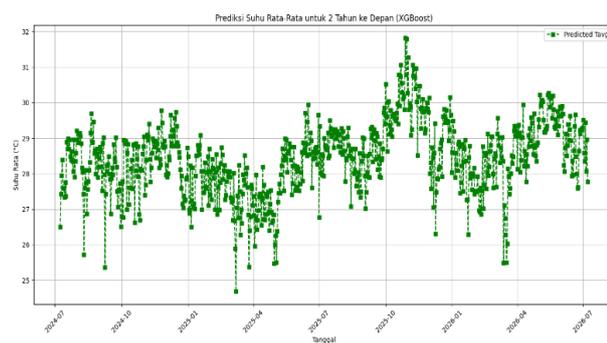
**Gambar 4.** Hasil prediksi SVR

SVR menghasilkan grafik prediksi pada gambar 5 yang menghasilkan MAE sebesar 0,3820, MSE sebesar 0,2524, dan  $R^2$  sebesar 0,7987. Meskipun nilai MAE dan MSE hampir sama dengan RF,  $R^2$  pada SVR sedikit lebih rendah, menunjukkan bahwa model ini kurang mampu menangkap variabilitas dalam data jika dibandingkan dengan Random Forest. Namun, perbedaan performa di antara keduanya tidak terlalu signifikan, mengindikasikan bahwa SVR masih menjadi model yang kompetitif dalam konteks ini.

*Learning curve* pada *XGBoost* yang disajikan pada gambar 6 menunjukkan performa terbaik dibandingkan dua model sebelumnya. Pada grafik, garis merah yang menunjukkan skor pelatihan dimulai dengan tinggi, menunjukkan bahwa model dapat dengan mudah menyesuaikan data, tetapi seiring bertambahnya jumlah data, skor pelatihan menurun sedikit. Garis hijau yang mewakili *cross-validation score* secara konsisten meningkat seiring bertambahnya jumlah data latih, menunjukkan kemampuan model untuk memprediksi dengan lebih baik pada data baru. Sumbu horizontal menunjukkan jumlah data latih, dan sumbu vertikal menunjukkan nilai kesalahan atau skor prediksi. *XGBoost* terlihat memiliki performa terbaik dalam hal generalisasi dibandingkan dua model lainnya.



**Gambar 5.** *XGBoost learning curve*



**Gambar 6.** Hasil prediksi *XGBoost*

*XGBoost* pada gambar 7 menunjukkan hasil prediksi terbaik dengan MAE sebesar 0,3744, MSE sebesar 0,2278, dan  $R^2$  sebesar 0,8183. MAE yang paling rendah di antara ketiga model menunjukkan bahwa kesalahan rata-rata prediksi sangat kecil, yang membuat model ini sangat baik dalam meminimalkan kesalahan. MSE yang lebih rendah menunjukkan bahwa *XGBoost* lebih efektif dalam menangani *outlier*, dan nilai  $R^2$  yang tertinggi menunjukkan bahwa model ini dapat menjelaskan sebagian besar variabilitas dalam data, membuatnya paling unggul di antara ketiga model dalam hal generalisasi dan akurasi.

**Tabel 2.** Performa hasil evaluasi

Model	Metrix			Best Model
	MAE	MSE	R <sup>2</sup>	
<i>Random Forest</i>	0,3869	0,2422	0,8068	<i>XGBoost</i>
SVR	0,3820	0,2524	0,7987	<i>XGBoost</i>
<i>XGBoost</i>	0,3744	0,2278	0,8183	<i>XGBoost</i>

Tabel 2, menunjukkan bahwa model *XGBoost* memiliki performa terbaik di antara *random forest*, SVR, dan *XGBoost*. *XGBoost* tidak hanya menunjukkan nilai MAE terendah (0.3744), tetapi juga mencatatkan nilai MSE paling rendah (0,2278) dan R<sup>2</sup> tertinggi (0,8183). Hal ini menegaskan bahwa *XGBoost* memiliki akurasi prediksi yang lebih baik dan lebih stabil dibandingkan model lainnya. Kemampuan ini sangat penting dalam aplikasi praktis, seperti prediksi cuaca, terutama dalam konteks sektor pertanian, di mana prediksi yang akurat dan stabil sangat diperlukan untuk pengambilan keputusan yang tepat dan berkelanjutan.

### Pembahasan

Analisis hasil dari model regresi menunjukkan bahwa masing-masing model *XGBoost*, *random forest*, dan SVR memiliki keunggulan dan kelemahan dalam memprediksi suhu rata-rata (Tavg) berdasarkan data cuaca historis. Model *XGBoost* menonjol dalam hal akurasi dan stabilitas, menghasilkan nilai MAE terendah (0,3744) dan R<sup>2</sup> tertinggi (0,8183) di antara ketiga model. Dengan kemampuan untuk menangkap pola kompleks dan interaksi non-linear antar fitur, model ini menunjukkan performa yang sangat baik dalam menghadapi data cuaca yang sering kali fluktuatif dan tidak terduga.

Hasil yang kuat dari model *XGBoost* ini sangat relevan dalam konteks pertanian, di mana akurasi prediksi suhu dapat membantu para petani mengambil keputusan yang lebih baik dalam hal waktu penanaman, pemilihan varietas tanaman, serta strategi mitigasi terhadap kondisi cuaca ekstrem. Penelitian sebelumnya, seperti yang dilakukan oleh (Sun et al., 2023), menunjukkan bahwa model berbasis pohon keputusan seperti *XGBoost* sering kali lebih efektif dalam menghadapi data cuaca dibandingkan model linier. Hal ini disebabkan oleh kemampuan *XGBoost* untuk menangkap interaksi kompleks antar variabel dan menangani non-linearitas dalam data, yang umum terjadi dalam pola cuaca. Model ini juga dilengkapi dengan teknik regularisasi yang mencegah *overfitting*, sehingga menghasilkan prediksi yang lebih stabil dan akurat. Temuan ini sejalan dengan hasil penelitian kami, yang menunjukkan bahwa *XGBoost* tidak hanya mengkonfirmasi efektivitas metode ini, tetapi juga memberikan bukti lebih lanjut mengenai kemampuannya dalam konteks prediksi suhu.

Sementara itu, model *random forest* juga menunjukkan performa yang kompetitif, dengan MAE sebesar 0,3869 dan R<sup>2</sup> sebesar 0,8068. Model ini efektif dalam mengurangi variabilitas dan memberikan hasil yang konsisten. Meskipun memiliki keunggulan dalam mengatasi masalah *overfitting* berkat agregasi dari beberapa pohon keputusan, hasilnya menunjukkan bahwa ada potensi untuk perbaikan dalam hal akurasi. Hal ini menyoroti pentingnya pemilihan hyperparameter dan teknik regularisasi yang lebih cermat. Dalam praktik, *random forest* dapat digunakan dalam pengambilan keputusan terkait manajemen risiko pertanian, dengan memberikan wawasan tentang prediksi suhu yang lebih baik dan potensi dampaknya terhadap hasil panen. Penelitian sebelumnya oleh Chen et al. (2020) mencatat bahwa *random forest* telah terbukti efektif dalam aplikasi agrikultur, terutama dalam memprediksi variabel cuaca yang mempengaruhi hasil panen. Hasil dari penelitian kami mendukung temuan ini, namun menunjukkan bahwa meskipun *random forest* efektif dalam memberikan prediksi suhu rata-rata, masih ada ruang untuk peningkatan lebih lanjut dalam hal akurasi prediksi dan stabilitas kinerja model, terutama saat menghadapi fluktuasi data cuaca

yang ekstrem. Peningkatan akurasi dan stabilitas ini penting untuk memastikan keandalan prediksi dalam konteks aplikasi praktis, seperti perencanaan dan pengelolaan pertanian yang bergantung pada informasi cuaca yang akurat.

Hasil model SVR menunjukkan performa yang lebih rendah dibandingkan dengan model *XGBoost* dan Random Forest, dengan MAE 0,3820 dan  $R^2$  0,7987. Keterbatasan ini disebabkan oleh sensitivitas SVR terhadap noise dan outlier dalam data, yang dapat mengganggu kemampuan model dalam generalisasi. Meskipun SVR efektif dalam menangkap hubungan linier, model ini kurang mampu menangani pola kompleks atau non-linear yang sering muncul dalam data cuaca. Sebaliknya, model berbasis pohon seperti *XGBoost* dan *random forest* lebih unggul dalam menangkap interaksi antar fitur, sehingga dapat menghadapi variasi yang lebih kompleks. Oleh karena itu, meskipun SVR memberikan hasil yang kompetitif, masih ada potensi untuk meningkatkan akurasinya melalui teknik pemilihan fitur dan regularisasi yang lebih baik.

Penelitian sebelumnya, seperti yang dilaporkan oleh Kumar et al. (2024), menunjukkan bahwa SVR efektif untuk data yang tidak terlalu kompleks, seperti data dengan hubungan linier yang jelas dan tingkat noise rendah. Hasil tersebut diperkuat oleh penelitian kami, yang mengonfirmasi bahwa SVR mampu menghasilkan prediksi yang baik dalam konteks data cuaca. Namun, untuk meningkatkan akurasi dan stabilitas performa SVR, diperlukan teknik tambahan, seperti pemilihan fitur yang lebih optimal dan penerapan regularisasi.

Selain itu, kami membandingkan performa SVR dengan model pembelajaran mesin lain, seperti *XGBoost* dan random forest, yang secara konsisten menunjukkan kemampuan lebih unggul dalam menangkap pola data cuaca yang kompleks. Hasil perbandingan ini menunjukkan bahwa model *XGBoost* dan *random forest* lebih adaptif terhadap kerumitan data, menjadikan kedua model ini lebih sesuai untuk aplikasi dalam prediksi cuaca ekstrem.

Implikasi praktis penelitian ini cukup signifikan. Model prediksi yang dikembangkan mampu menghasilkan akurasi tinggi dan stabilitas yang konsisten, sehingga dapat mendukung pengambilan keputusan di sektor pertanian. Sebagai contoh, model ini dapat digunakan untuk memprediksi suhu masa depan, memungkinkan implementasi tindakan preventif seperti pengaturan irigasi yang lebih efisien atau perlindungan tanaman sebelum terjadi suhu ekstrem. Hal ini sangat relevan dalam konteks perubahan iklim, yang meningkatkan ketidakpastian pola cuaca, sekaligus menjadi dasar yang kokoh bagi pengambilan keputusan dalam mitigasi risiko cuaca ekstrem.

Kontribusi penelitian ini terhadap pengembangan pengetahuan di bidang prediksi cuaca juga patut dicatat. Dengan mengintegrasikan teknologi dan metodologi terkini dalam pembelajaran mesin, penelitian ini tidak hanya memvalidasi efektivitas model-model yang telah diuji dalam kajian sebelumnya, tetapi juga menunjukkan peningkatan dalam akurasi prediksi dan pengelolaan data. Oleh karena itu, penelitian ini tidak hanya relevan secara teoritis tetapi juga memberikan dampak praktis yang substansial, membuka peluang untuk eksplorasi lebih lanjut di bidang prediksi cuaca, terutama dalam menghadapi tantangan yang ditimbulkan oleh perubahan iklim global.

## SIMPULAN

Penelitian ini menemukan bahwa model *XGBoost* unggul dalam memprediksi suhu rata-rata ( $T_{avg}$ ) dibandingkan dengan *random forest* dan SVR, terbukti melalui metrik evaluasi seperti MAE, MSE, dan  $R^2$ . Nilai MAE terendah menunjukkan kesalahan rata-rata yang minimal, yang penting untuk aplikasi yang memerlukan akurasi tinggi, sementara MSE yang lebih rendah mencerminkan kemampuan model dalam menangani outlier. Selain itu, nilai  $R^2$  yang tinggi menunjukkan proporsi variabilitas data yang berhasil dijelaskan oleh model. Temuan ini memiliki implikasi praktis yang signifikan, terutama dalam pemantauan suhu otomatis, yang dapat membantu petani mengoptimalkan waktu tanam dan panen serta

meminimalkan kerugian akibat cuaca ekstrem. Selain itu, hasil penelitian ini dapat mendukung pengembangan kebijakan yang responsif terhadap perubahan iklim, meningkatkan kesiapsiagaan masyarakat terhadap bencana terkait cuaca, sehingga relevan baik untuk pengembangan model prediksi maupun untuk kebijakan dan strategi adaptasi di berbagai sektor.

## REFERENSI

- Ardian, M., Khomsah, S., & Pandiya, R. (2024). Perbandingan Model Regresi Untuk Memprediksi Harga Jual Cabai Rawit Berdasarkan Iklim Harian. *Jurnal JUPITER*, 16(2), 549–560.
- Bala, S. (2021). COVID-19 Outbreak Prediction Analysis Using Machine Learning. *International Journal for Research in Applied Science and Engineering Technology*, 9(1), 1–7. <https://doi.org/10.22214/ijraset.2021.32690>
- Hadianto, A., & Utomo, W. H. (2024). CatBoost Optimization Using Recursive Feature Elimination. *JOIN (Jurnal Online Informatika)*, 9(2), 169–178. <https://doi.org/10.15575/join.v9i1.1324>
- Himawan, F., Pressa Perdana, & Yoedo Ageng Surya. (2021). Rancang Bangun Purwarupa Smart Garden Menggunakan Kamera, Sensor Suhu Dan Kelembaban Tanah Berbasis Internet Of Things (IOT) Dengan ESP8266. *Jurnal JEETech*, 2(2), 78–83. <https://doi.org/10.48056/jeetech.v2i2.171>
- Huang, L., Liu, Y., Huang, W., Dong, Y., Ma, H., Wu, K., & Guo, A. (2022). Combining Random Forest And XGBoost Methods In Detecting Early And Mid-Term Winter Wheat Stripe Rust Using Canopy Level Hyperspectral Measurements. *Agriculture (Switzerland)*, 12(1). <https://doi.org/10.3390/agriculture12010074>
- Hutagalung, C. A. H., Rosalind, G. A., Tuhu, D. M. S., & Agustianingsih, A. (2023). Wholesale Inventory Management Optimization: Methodological Approach With XGBoost, SVR, And Random Forest Algorithms. *Brilliance: Research of Artificial Intelligence*, 3(2), 369–377. <https://doi.org/10.47709/brilliance.v3i2.3336>
- Karim, A. A., Ary Prasetyo, M., & Saputro, M. R. (2023). Perbandingan Metode Random Forest, K-Nearest Neighbor, Dan SVM Dalam Prediksi Akurasi Pertandingan Liga Italia. *Prosiding Seminar Nasional Teknologi Dan Sains*, 2(7), 377–382.
- Kumar, M. S., Srivastava, D. M., & Prakash, D. V. (2024). Advanced Hybrid Prediction Model: Optimizing LightGBM, XGBoost, Lasso Regression, And Random Forest With Bayesian Optimization. *Journal of Theoretical and Applied Information Technology*, 15(9). 4103-4115.
- Li, H., Zhang, G., Zhong, Q., Xing, L., & Du, H. (2023). Prediction of urban forest aboveground carbon using machine learning based on landsat 8 and Sentinel-2: a case study of Shanghai, China. *Remote Sensing*, 15(1), 284. <https://doi.org/10.3390/rs15010284>
- Muhammady, D. N., Nugraha, H. A. E., Nastiti, V. R. S., Sri, C., & Aditya, K. (2024). Students Final Academic Score Prediction Using Boosting Regression Algorithms. *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)*, 10(1), 154-165. <https://doi.org/10.26555/jiteki.v10i1.28352>
- Pamungkas, A. S., & Cahyono, N. (2024). Analisis Sentimen Review ChatGPT Di Play Store menggunakan Support Vector Machine Dan K-Nearest Neighbor. *Edumatic: Jurnal Pendidikan Informatika*, 8(1), 1–10. <https://doi.org/10.29408/edumatic.v8i1.24114>
- Patasik, E. S., & Yulianto, S. (2023). Classification of Regional Languages Using Methods Gradient Boosts And Random Forest. *Jurnal Teknik Informatika (Jutif)*, 4(5), 1249–1255. <https://doi.org/10.52436/1.jutif.2023.4.5.1459>
- Pebrianti, D., Kurniawan, H., Bayuaji, L., & Rusdah, R. (2023). XgBoost Hyper-Parameter

- Tuning Using Particle Swarm Optimization For Stock Price Forecasting. *Jurnal Ilmiah Teknik Elektro Komputer Dan Informatika (JITEKI)*, 9(4), 1179–1195.
- Priyatno, A. M., Tanjung, L. S., Ramadhan, W. F., Cholidhazia, P., Jati, P. Z., & Firmananda, F. I. (2023). Comparison Random Forest Regression And Linear Regression For Forecasting BBKA Stock Price. *Jurnal Teknik Industri Terintegrasi*, 6(3), 718–732. <https://doi.org/10.31004/jutin.v6i3.16933>
- Putri, N. F., Hidayati, R., & Nirmala, I. (2023). Rancang Bangun Sistem Pemantauan Dan Kendali Budidaya Anggur Dengan Penerapan Internet Of Things (IoT) Berbasis Android. *Jurnal Riset Komputer*, 10(3), 2407–389.
- Rakhmat, G. A., & Mutohar, W. (2023). Prakiraan Hujan menggunakan Metode Random Forest dan Cross Validation. *MIND (Multimedia Artificial Intelligent Networking Database) Journal*, 8(2), 173-187.
- Satria, A., Badri, R. M., & Safitri, I. (2023). Prediksi Hasil Panen Tanaman Pangan Sumatera Dengan Metode Machine Learning. *Digital Transformation Technology*, 3(2), 389–398. <https://doi.org/10.47709/digitech.v3i2.2852>
- Setiawan, A., & Suryono, R. R. (2024). Analisis Sentimen Ibu Kota Nusantara Menggunakan Algoritma Support Vector Machine Dan Naïve Bayes. *Edumatic: Jurnal Pendidikan Informatika*, 8(1), 183–192. <https://doi.org/10.29408/edumatic.v8i1.25667>
- Sun, J., Dang, W., Wang, F., Nie, H., Wei, X., Li, P., Zhang, S., Feng, Y., & Li, F. (2023). Prediction Of TOC Content In Organic-Rich Shale Using Machine Learning Algorithms: Comparative Study Of Random Forest, Support Vector Machine, And XGBoost. *Energies*, 16(10). <https://doi.org/10.3390/en16104159>
- Taqiyuddin, M., & Bayu Sasongko, T. (2024). Prediksi Cuaca Kabupaten Sleman Menggunakan Algoritma Random Forest. *Jurnal Media Informatika Budidarma*, 8(3), 1683. <https://doi.org/10.30865/mib.v8i3.7897>