

Automated Machine Learning Berbasis Pengetahuan Fisika untuk Model Prediksi Suhu dalam Kondisi Data Terbatas

Ade Hendini

Universitas Bina Sarana Informatika

e-mail: *ade.aee@bsi.ac.id

Abstract – Temperature prediction is an important part of various sectors, especially in the agricultural sector. However, limited data and complexity between variables are the main challenges in developing temperature predictive models. To overcome these problems, this study proposes a new approach by combining Automated Machine Learning (AutoML) and physics knowledge in one efficient and domain-aware framework. By comparing two AutoML platforms, namely FLAML and H2O AutoML. AutoML plays a role in automating model selection, feature selection, and adaptive parameter tuning, while physics-based components are used to limit the space to remain consistent with scientific principles. In this study, daily meteorological data such as temperature, humidity, rainfall, and solar radiation were obtained from the NASA POWER service. The model was developed using the FLAML and H2O AutoML libraries which were then evaluated using RMSE and R^2 metrics. The results show that H2O AutoML provides better prediction performance with an RMSE value of 0.6329 and an R^2 of 0.6582, compared to FLAML which produces an RMSE of 0.7001 and an R^2 of 0.4762. The visualization shows that H2O predictions are closer to the actual values and follow the temperature trend more accurately.

Keywords - AutoML, Physics-Informed Machine Learning, Temperature Prediction, Limited Data, FLAML, H2O AutoML, NASA POWER

Abstrak – Prediksi suhu merupakan bagian penting dalam berbagai sektor terutama pada sektor pertanian. Namun, keterbatasan data dan kompleksitas antar variabel menjadi tantangan utama pada pengembangan model prediktif suhu. Untuk mengatasi masalah tersebut, penelitian ini mengusulkan pendekatan baru dengan menggabungkan Automated Machine Learning (AutoML) dan pengetahuan fisika dalam satu kerangka kerja yang efisien dan domain-aware. Dengan membandingkan dua platform AutoML yaitu FLAML dan H2O AutoML. AutoML berperan mengotomatisasi pemilihan model, seleksi fitur, dan penyetelan parameter secara adaptif, sementara komponen berbasis pengetahuan fisika digunakan untuk membatasi ruang agar tetap konsisten dengan prinsip ilmiah. Dalam studi ini, data meteorologi harian seperti suhu, kelembapan, curah hujan, dan radiasi matahari diperoleh dari layanan NASA POWER. Model dikembangkan menggunakan pustaka FLAML dan H2O AutoML yang kemudian dievaluasi menggunakan metrik RMSE dan R^2 . Hasil menunjukkan bahwa H2O AutoML memberikan performa prediksi yang lebih baik dengan nilai RMSE sebesar 0.6329 dan R^2 sebesar 0.6582, dibanding FLAML yang menghasilkan RMSE sebesar 0.7001 dan R^2 sebesar 0.4762. Visualisasi menunjukkan bahwa prediksi H2O lebih dekat terhadap nilai aktual dan mengikuti tren suhu dengan lebih akurat.

Kata Kunci – AutoML, Physics-Informed Machine Learning, Prediksi Suhu, Data Terbatas, FLAML, H2O AutoML, NASA POWER

I. PENDAHULUAN

Dalam beberapa dekade terakhir, kemajuan teknologi kecerdasan buatan (AI) telah mendorong penggunaan machine learning (ML) di berbagai sektor, termasuk lingkungan pertanian. Salah satu aplikasi penting adalah prediksi suhu, baik suhu tanah, udara, maupun air, yang memiliki implikasi signifikan terhadap sistem irigasi, pengelolaan

tanaman, dan pengendalian iklim mikro [1]. Namun, pengembangan model ML konvensional memerlukan keahlian teknis tinggi serta proses eksplorasi fitur, seleksi algoritma, dan tuning hiperparameter yang memakan waktu.

Automated Machine Learning (AutoML) muncul sebagai solusi untuk mengotomatisasi proses pembangunan model ML dari awal hingga akhir [2]. Dengan AutoML, proses pemodelan menjadi lebih mudah diakses oleh pengguna non-ahli serta mempercepat pengembangan sistem prediktif. Namun demikian, sebagian besar pendekatan AutoML yang ada masih murni data-driven dan mengabaikan aspek pengetahuan domain, termasuk prinsip-prinsip fisis yang mendasari fenomena yang dimodelkan [3].

Dalam konteks prediksi suhu, prinsip-prinsip fisis seperti hukum termodinamika, keseimbangan energi, konduktivitas termal, serta interaksi antara kelembapan, radiasi matahari, dan jenis tanah memainkan peran penting. Mengabaikan aspek ini dapat menyebabkan model ML kehilangan akurasi dan tidak mampu melakukan generalisasi dalam kondisi ekstrem atau tidak teramati sebelumnya [4].

Tantangan utama dalam pemodelan suhu dengan ML adalah kelangkaan dan ketidakteraturan data. Di banyak wilayah, khususnya pedesaan dan negara berkembang, data sensor suhu sering tidak tersedia secara konsisten atau memiliki kualitas rendah [5]. Kondisi ini membuat model data-driven murni sulit dikembangkan tanpa risiko overfitting atau bias prediksi.

Pendekatan AutoML muncul untuk mengotomatisasi pipeline ML, mulai dari preprocessing hingga memilih model terbaik [6].

FLAML merupakan salah satu framework AutoML ringan yang dikembangkan oleh Microsoft Research. Framework ini dirancang untuk efisiensi serta memungkinkan pelatihan model prediksi tanpa banyak intervensi pengguna dan menggunakan sumber daya yang rendah, sedangkan H2O AutoML memiliki kemampuan menghasilkan model dengan akurasi tinggi [7].

AutoML dapat memudahkan proses pelatihan model, walaupun demikian pendekatan ini cenderung menghasilkan model *black-box* yang kurang memperhatikan validitas pada hukum fisika. Sehingga pada konteks suhu, model dapat menghasilkan nilai prediksi statistik sah namun secara ilmiah tidak masuk akal [8].

Untuk mengatasi hal tersebut pendekatan *Physics-Informed Machine Learning* (PIML) mulai diperkenalkan untuk mengintegrasikan hukum fisika ke dalam proses pelatihan ML, sehingga model dapat menjaga konsistensi fisika walaupun dengan data yang terbatas [9].

Dalam kenyataannya penerapan PIML dalam AutoML masih terbatas. Keterbatasan struktur internal AutoML seperti FLAML dan H2O menyulitkan modifikasi langsung terhadap *loss function* agar sesuai dengan hukum fisika, dan salah satu solusinya adalah dengan menambahkan *post-processing* untuk meminimalkan pelanggaran hukum fisika [10]. Pada kondisi dataset yang terbatas pendekatan data-driven AutoML mudah terjadi overfitting karena tidak ada pengetahuan tambahan yang dapat membimbing dalam proses pembelajaran.

Banyak sistem fisika hanya memiliki sedikit data dan data sensor seringkali *noisy* dan memiliki spasial yang rendah, ini mengharuskan PIML dapat belajar dengan efisien dari keterbatasan data dengan mengambil pengetahuan fisis sebagai regularisasi [11].

Tabel I
Perbandingan AutoML Konvensional dengan PIML

Aspek	AutoML Konvensional	PIML
Prinsip Dasar	Optimisasi otomatis pipeline ML (preprocessing, model, hyperparameter) berdasarkan metrik statistik	Integrasi hukum fisika ke dalam proses pelatihan model
Ketergantungan pada data	Sangat bergantung pada jumlah dan kualitas data	Dapat bekerja dengan data terbatas karena sebelumnya diproses dengan hukum fisika
Kesesuaian untuk sistem fisis	Prediksi dapat melanggar hukum fisika	Cocok untuk sistem dinamis, rekayasa, fluida, energi dan lingkungan

Dari pemaparan tersebut peneliti bertujuan untuk membandingkan dua pendekatan AutoML yaitu FLAML dan H2O AutoML dalam membangun model prediksi suhu lingkungan dengan data yang terbatas.

II. PENELITIAN YANG TERKAIT

Penelitian sebelumnya telah mengungkapkan berbagai percobaan terkait AutoML dan PIML dalam prediksi sistem fisis. Berikut adalah tabel perbandingan:

Tabel II
Perbandingan Penelitian Sebelumnya

No	Penulis	Judul	Pendekatan	Hasil	Gap Penelitian
1.	Wu et al. (2022)	AutoPINN: When AutoML Meets Physics-Informed Neural Networks	AutoML + PINNS	Mendesain AutoML untuk mencari arsitektur dan hyperparameter terbaik pada PINNs [12]	Belum mendukung loss fisis
2	Nghiem et al. (2023)	Physics-Informed ML for Modeling and Control of Dynamical Systems	PIML + Control (ODE system)	Tutorial dan kerangka pemodelan + kontrol berbasis fisika, membahas integrasi ke sistem riil dan digital twin	Tidak menggabungkan AutoML
3	Guo et al. (2023)	A Physics-Informed Automatic Neural Network Generation Framework for Emerging Device Modeling	AutoPINN (AutoNN + PINN)	AutoPINN secara otomatis dapat menghasilkan model neural network yang akurat dan memenuhi karakteristik fisika [13]	Model hanya diuji pada satu jenis perangkat berbasis data simulasi yang baik, sehingga belum terbukti keandalannya pada data nyata atau perangkat dengan kompleksitas fisika yang lebih tinggi

Pengembangan sistem prediksi suhu telah banyak dilakukan menggunakan pendekatan berbasis machine learning seperti Random Forest, Support Vector Regression (SVR), dan Artificial Neural Networks (ANN). Beberapa studi menunjukkan keberhasilan model ini dalam melakukan prediksi suhu berdasarkan data historis dan data lingkungan, seperti kelembapan, curah hujan, serta intensitas radiasi matahari. Di sisi lain, perkembangan AutoML seperti Auto-sklearn, Google AutoML, dan FLAML[14] memungkinkan proses otomatisasi pipeline ML, mulai dari seleksi fitur, pemilihan model, hingga tuning hiperparameter.

Namun, pendekatan AutoML konvensional tersebut sebagian besar masih berfokus pada optimasi metrik akurasi dengan pendekatan black-box dan data-driven murni. Dalam konteks sistem fisis seperti suhu tanah atau udara, hal ini menjadi keterbatasan, karena model yang dikembangkan seringkali gagal merepresentasikan dinamika sistem yang sebenarnya dan cenderung tidak robust terhadap kondisi di luar distribusi data pelatihan [8]. Oleh karena itu, dalam beberapa tahun terakhir, pendekatan Physics-Informed Machine Learning (PIML) mulai dikembangkan untuk mengatasi keterbatasan ini.

PIML adalah pendekatan yang menggabungkan pembelajaran dari data dengan informasi eksplisit dari hukum-hukum fisika, seperti persamaan diferensial parsial (PDE), hukum termodinamika, atau batasan-batasan fisik lainnya [9]. Berbagai studi menunjukkan bahwa dengan mengintegrasikan pengetahuan fisika, model ML dapat belajar lebih stabil, lebih cepat, dan menghasilkan prediksi yang lebih konsisten secara ilmiah, terutama dalam kondisi keterbatasan data [8].

III. METODE PENELITIAN

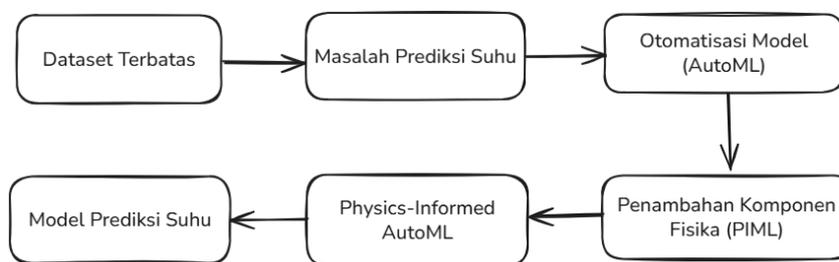
Prediksi suhu merupakan masalah regresi yang kompleks, terutama ketika data yang tersedia terbatas, tidak lengkap, atau bersifat spasial-temporal. Model machine learning konvensional membutuhkan jumlah data yang besar agar mampu belajar secara efektif dan menghasilkan prediksi yang akurat. Dalam situasi kekurangan data, model

seperti Random Forest atau Neural Networks cenderung mengalami overfitting dan kehilangan kemampuan generalisasi.

Automated Machine Learning (AutoML) hadir sebagai solusi yang mampu mengotomatisasi pemilihan model, seleksi fitur, dan tuning parameter, sehingga mempercepat pengembangan sistem prediktif. Namun, AutoML konvensional bersifat data-driven dan tidak mempertimbangkan karakteristik sistem fisika dari fenomena suhu, seperti konservasi energi atau pengaruh radiasi matahari. Hal ini menjadikan AutoML kurang optimal untuk fenomena ilmiah dengan struktur fisis yang kompleks.

Physics-Informed Machine Learning (PIML) menawarkan solusi dengan memasukkan pengetahuan domain ke dalam proses pembelajaran model. Misalnya, dengan menambahkan loss berbasis hukum termodinamika atau membatasi ruang pencarian parameter agar sesuai dengan perilaku fisika. Dengan mengintegrasikan PIML ke dalam pipeline AutoML, diharapkan akan dihasilkan model yang tidak hanya akurat, tetapi juga konsisten secara ilmiah, meskipun data yang tersedia terbatas.

penelitian ini diawali dengan mengidentifikasi bagaimana membangun model prediksi suhu yang efisien dan andal dalam kondisi data terbatas. Kemudian, ditetapkan bahwa AutoML menjadi solusi dasar yang dipilih, namun perlu ditingkatkan dengan pengetahuan fisika agar menghasilkan model yang lebih baik. Oleh karena itu, penelitian ini menggabungkan AutoML dan PIML dalam satu kerangka kerja terpadu.



Gambar 1. Metode Penelitian

Penelitian ini merupakan penelitian kuantitatif eksperimental yang bertujuan untuk mengembangkan dan menguji model prediksi suhu berbasis AutoML yang diinformasikan oleh pengetahuan fisika.

Objek dalam penelitian ini adalah model prediktif suhu, sedangkan data yang digunakan berasal dari dataset publik seperti NASA POWER dengan membandingkan dua model yaitu FLAML dan H2O AutoML. Dataset akan mencakup variabel suhu, kelembapan, radiasi matahari, penguapan.

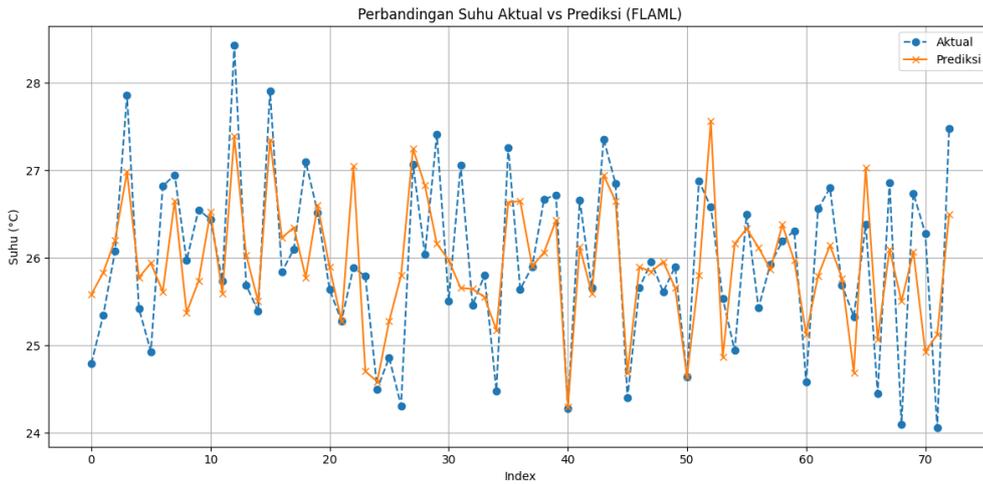
IV. HASIL DAN PEMBAHASAN

Hasil evaluasi model prediksi suhu menunjukkan bahwa terdapat perbedaan kinerja yang cukup signifikan antara model yang dibangun menggunakan FLAML dan H2O AutoML. Berdasarkan metrik evaluasi pada data uji, model H2O AutoML menghasilkan nilai RMSE sebesar 0.6329 dan R² sebesar 0.6582, sedangkan model FLAML memberikan nilai RMSE sebesar 0.7001 dan R² sebesar 0.4762. Hal ini mengindikasikan bahwa H2O AutoML mampu menghasilkan prediksi suhu yang lebih mendekati nilai aktual serta dapat menjelaskan variabilitas data yang lebih tinggi dibanding FLAML.

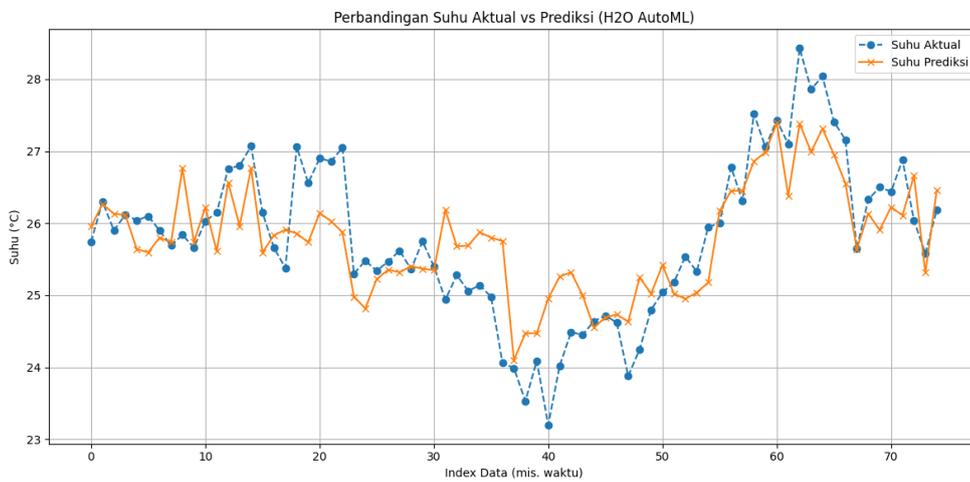
Tabel III
Hasil Evaluasi FLAML dan H2O AutoML

Metrik	FLAML	H2O AutoML
RMSE	0.7001373538326101	0.6328821093654221
R ²	0.4762245903959522	0.6581913893477669

Nilai RMSE yang lebih rendah pada H2O AutoML menunjukkan bahwa kesalahan rata-rata prediksi terhadap suhu aktual lebih kecil dibanding FLAML. Dengan kata lain, model yang dihasilkan oleh H2O AutoML lebih akurat dalam meramalkan suhu berdasarkan fitur-fitur seperti kelembapan udara, radiasi matahari, dan curah hujan yang digunakan sebagai input. Sementara itu, nilai R² sebesar 0.6582 pada H2O AutoML mengindikasikan bahwa sekitar 65.82% variasi pada suhu aktual dapat dijelaskan oleh model, yang berarti model memiliki daya prediksi yang lebih baik. Sebaliknya, model FLAML hanya mampu menjelaskan sekitar 47.62% variasi, menunjukkan keterbatasan dalam menangkap hubungan kompleks antara fitur input dan target suhu. Ini menunjukkan bahwa prediksi H2O AutoML lebih mendekati nilai aktual dan lebih handal dalam menangkap hubungan antara variabel input dan target suhu.

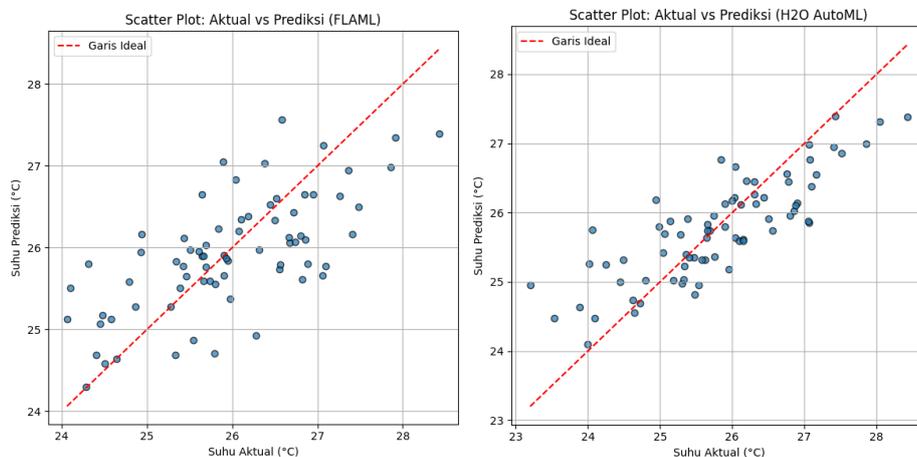


Gambar 2. Perbandingan Suhu Aktual vs Prediksi (FLAML)



Gambar 3. Perbandingan Suhu Aktual vs Prediksi (H2O AutoML)

Pada grafik Perbandingan Suhu Aktual vs Prediksi tersebut menunjukkan pola prediksi FLAML lebih fluktuatif dan tidak mengikuti tren aktual dengan baik. Terjadi banyak deviasi dan perpotongan antara garis aktual dan prediksi yang menunjukkan bahwa model FLAML kurang stabil dan tidak konsisten dalam mempelajari pola data. Sedangkan pada H2O AutoML menunjukkan korelasi tren yang lebih akurat antara suhu aktual dan prediksi. Garis prediksi cenderung mengikuti garis aktual yang menunjukkan sedikitnya deviasi terhadap garis aktual.



Gambar 4.a Scatter Plot : Aktual vs Prediksi (FLAML) b. Scatter Plot : Aktual vs Prediksi (H2O AutoML)

Pada grafik Scatter FLAML titik-titik prediksi tersebar cukup jauh dari garis ideal dibandingkan dengan H2O AutoML yang cenderung lebih dekat dengan garis ideal. Pada FLAML terjadi *underestimation* dan *overestimation* yang cukup banyak sedangkan H2O AutoML Sebagian besar prediksi berada disepanjang garis diagonal yang menunjukkan bahwa adanya korelasi yang kuat antara suhu prediksi dan suhu aktual.

V. KESIMPULAN

Penelitian ini bertujuan untuk mengevaluasi kinerja dua pendekatan Automated Machine Learning (AutoML), yaitu FLAML dan H2O AutoML, dalam membangun model prediksi suhu dengan data yang terbatas. Berdasarkan hasil eksperimen dan visualisasi, disimpulkan bahwa kedua model memiliki kemampuan prediktif yang berbeda, baik dari sisi akurasi, stabilitas, maupun kemampuan generalisasi.

Secara kuantitatif, model H2O AutoML menunjukkan kinerja yang lebih unggul dibanding FLAML. H2O menghasilkan nilai Root Mean Squared Error (RMSE) sebesar 0.6329 dan koefisien determinasi (R^2) sebesar 0.6582, sementara FLAML hanya mencapai RMSE sebesar 0.7001 dan R^2 sebesar 0.4762. Hal ini menunjukkan bahwa H2O lebih akurat dalam memprediksi suhu dan lebih baik dalam menjelaskan keanekaragaman data. Perbedaan kinerja ini konsisten dengan hasil scatter plot dan visualisasi prediksi terhadap data aktual.

Dari segi visualisasi, H2O AutoML mampu menghasilkan prediksi suhu yang lebih stabil dan mendekati nilai aktual. Garis prediksi pada grafik cenderung sejajar dengan garis aktual, dan titik-titik scatter plot tersebar lebih dekat ke garis ideal yang menunjukkan error yang rendah. Sebaliknya, prediksi FLAML menunjukkan pola fluktuatif, dengan banyak titik yang menyimpang dari garis ideal. Hal ini menunjukkan bahwa FLAML kurang mampu dalam menangkap pola nonlinear dalam data suhu secara optimal.

Meskipun demikian, FLAML tetap memiliki keunggulan dalam hal efisiensi dan kecepatan proses pelatihan. Model FLAML lebih ringan dan cocok digunakan dalam lingkungan komputasi terbatas, seperti perangkat IoT atau *edge computing*. FLAML juga lebih fleksibel untuk dikustomisasi dengan pendekatan physics-informed, sehingga tetap relevan dalam skenario tertentu, terutama jika integrasi dengan pengetahuan domain menjadi prioritas utama.

Secara keseluruhan, H2O AutoML lebih direkomendasikan untuk skenario prediksi suhu yang memerlukan akurasi tinggi dan data kompleks. Namun, pemilihan metode AutoML sebaiknya mempertimbangkan konteks aplikasi, sumber daya komputasi yang tersedia, serta kemungkinan integrasi dengan pengetahuan fisika. Kombinasi antara efisiensi FLAML dan kekuatan prediktif H2O AutoML dapat menjadi dasar untuk eksplorasi model hybrid pada untuk studi lanjutan berikutnya.

DAFTAR PUSTAKA

- [1] R. K. Goel, C. S. Yadav, S. Vishnoi, and R. Rastogi, "Smart agriculture – Urgent need of the day in developing countries," *Sustainable Computing: Informatics and Systems*, vol. 30, p. 100512, 2021, doi: <https://doi.org/10.1016/j.suscom.2021.100512>.
- [2] X. He, K. Zhao, and X. Chu, "AutoML: A survey of the state-of-the-art," *Knowl Based Syst*, vol. 212, p. 106622, 2021, doi: <https://doi.org/10.1016/j.knosys.2020.106622>.
- [3] F. Hutter, L. Kotthoff, and J. Vanschoren, "The Springer Series on Challenges in Machine Learning Automated Machine Learning Methods, Systems, Challenges." [Online]. Available: <http://www.springer.com/series/15602>
- [4] V. K. Quy *et al.*, "IoT-enabled smart agriculture: architecture, applications, and challenges," *Applied Sciences*, vol. 12, no. 7, p. 3396, 2022.
- [5] L. G, R. C, and G. P, "An automated low cost IoT based Fertilizer Intimation System for smart agriculture," *Sustainable Computing: Informatics and Systems*, vol. 28, p. 100300, 2020, doi: <https://doi.org/10.1016/j.suscom.2019.01.002>.
- [6] M.-A. Zöller and M. F. Huber, "Benchmark and Survey of Automated Machine Learning Frameworks," *J. Artif. Int. Res.*, vol. 70, pp. 409–472, May 2021, doi: 10.1613/jair.1.11854.
- [7] S. A. Mezzasalma, "On the Extension of a Physical Body in Classical Motion. An Analogy for a Pseudo-Velocity Concept and Wiener's Process in (Ideal) Polymer Solutions," 2020. [Online]. Available: <https://arxiv.org/abs/2009.06105>
- [8] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, "Physics-informed machine learning," *Nature Reviews Physics*, vol. 3, no. 6, pp. 422–440, 2021, doi: 10.1038/s42254-021-00314-5.
- [9] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *J Comput Phys*, vol. 378, pp. 686–707, 2019, doi: <https://doi.org/10.1016/j.jcp.2018.10.045>.
- [10] S. Liao, T. Xue, J. Jeong, S. Webster, K. Ehmann, and J. Cao, "Hybrid thermal modeling of additive manufacturing processes using physics-informed neural networks for temperature prediction and parameter identification," *Comput Mech*, vol. 72, no. 3, pp. 499–512, Jan. 2023, doi: 10.1007/s00466-022-02257-9.

-
- [11] T. X. Nghiem *et al.*, “Physics-Informed Machine Learning for Modeling and Control of Dynamical Systems,” 2023. [Online]. Available: <https://arxiv.org/abs/2306.13867>
- [12] X. Wu *et al.*, “AutoPINN: When AutoML Meets Physics-Informed Neural Networks,” 2022. [Online]. Available: <https://arxiv.org/abs/2212.04058>
- [13] G. Guo, H. You, C. Li, Z. Tang, and O. Li, “A Physics-Informed Automatic Neural Network Generation Framework for Emerging Device Modeling,” *Micromachines (Basel)*, vol. 14, no. 6, 2023, doi: 10.3390/mi14061150.
- [14] C. Wang, Q. Wu, X. Liu, and L. Quintanilla, “Automated Machine Learning & Tuning with FLAML,” in *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, in KDD '22. New York, NY, USA: Association for Computing Machinery, 2022, pp. 4828–4829. doi: 10.1145/3534678.3542636.