

Penerapan Algoritma Regresi Random Forest Untuk Prediksi Produksi Jagung Menggunakan Data Statistik Sistem Pertanian Cerdas Smart City

Ahmad Bagus Prakoso^{1*}, Muhammad Anugrah Putra², Muhammad Hanif Hilmi³, Satya Yoma Patria Risky⁴,
Joni Maulindar⁵

¹Ilmu Komputer
Universitas Duta Bangsa
^{1*}230103124@mhs.udb.ac.id

²Ilmu Komputer
Universitas Duta Bangsa
²230103142@mhs.udb.ac.id

³Ilmu Komputer
Universitas Duta Bangsa
³230103143@mhs.udb.ac.id

⁴Ilmu Komputer
Universitas Duta Bangsa
⁴230103151@mhs.udb.ac.id

⁵Ilmu Komputer
Universitas Duta Bangsa
⁵joni_maulindar@udb.ac.id

Abstrak— Produksi jagung di Indonesia masih dihadapkan pada berbagai permasalahan seperti ketidakstabilan hasil panen serta kurangnya data prediktif yang mendukung pengambilan keputusan berbasis informasi. Penelitian ini bertujuan untuk mengembangkan model prediksi produksi jagung yang akurat menggunakan algoritma Random Forest Regression berbasis data statistik pertanian. Metode yang digunakan bersifat kuantitatif dengan memanfaatkan data luas panen dan produktivitas jagung dari 38 provinsi di Indonesia tahun 2024, yang dianalisis menggunakan Python dan pustaka machine learning. Hasil evaluasi menunjukkan bahwa model mampu menjelaskan 98% variasi produksi jagung, dengan nilai R^2 sebesar 0,98 pada data uji. Nilai Mean Squared Error (MSE) dan Mean Absolute Error (MAE) yang rendah menandakan tingkat akurasi yang tinggi. Visualisasi hasil prediksi menunjukkan bahwa nilai-nilai prediksi sangat mendekati nilai aktual. Fitur luas panen memiliki kontribusi paling besar dalam prediksi, diikuti oleh produktivitas. Model disimpan dalam format .pkl dan diintegrasikan ke dalam aplikasi web untuk memudahkan penggunaannya oleh petani maupun dinas pertanian. Penelitian ini mendukung penerapan pertanian cerdas dalam konsep Smart City serta dapat digunakan sebagai alat bantu pengambilan keputusan berbasis data.

Kata kunci— Jagung, Produksi Pertanian, Prediksi, Random Forest Regression, Smart City, Machine Learning.

Abstract— Corn production in Indonesia continues to face challenges such as yield instability and a lack of predictive data to support informed decision-making. This study aims to develop an accurate corn production prediction model using the Random Forest Regression algorithm based on agricultural statistical data. A quantitative approach was applied using data on harvest area and productivity from 38 provinces in Indonesia in 2024, processed with Python and machine learning libraries. The evaluation results show that the model explains 98% of the variance in corn production, with an R^2 score of 0.98 on the test data. Low Mean Squared Error (MSE) and Mean Absolute Error (MAE) values indicate high predictive accuracy. The prediction visualization reveals that the predicted values closely align with actual data. Harvest area emerged as the most influential feature, followed by productivity. The model was saved in .pkl format and integrated into a web-based application to facilitate its use by farmers and agricultural agencies. This study supports the implementation of smart agriculture within the Smart City framework and provides a data-driven decision-support tool.

Keywords— Corn, Agricultural Production, Prediction, Random Forest Regression, Smart City, Machine Learning, Agriculture.

I. PENDAHULUAN

Pertanian merupakan sektor strategis yang memainkan peran penting dalam ketahanan pangan dan pembangunan ekonomi, terutama di negara agraris seperti Indonesia[1]. Salah satu komoditas utama yang memiliki nilai ekonomis tinggi adalah jagung. Jagung tidak hanya digunakan sebagai bahan pangan pokok kedua setelah beras, tetapi juga sebagai bahan baku

industri pakan ternak dan bioenergi[2]. Oleh karena itu, peningkatan produktivitas dan efisiensi dalam produksi jagung menjadi kebutuhan mendesak, khususnya di tengah tantangan perubahan iklim, keterbatasan lahan, serta pertumbuhan populasi yang semakin tinggi. Namun, proses produksi jagung di berbagai daerah masih menghadapi sejumlah kendala, seperti fluktuasi hasil panen, ketergantungan terhadap musim, serta kurangnya data dan

informasi yang akurat untuk mendukung pengambilan keputusan yang tepat oleh petani maupun pemangku kebijakan.

Dengan berkembangnya konsep Smart City, berbagai sektor kehidupan mulai memanfaatkan teknologi untuk menciptakan sistem yang lebih efisien, responsif, dan terintegrasi. Pertanian cerdas atau smart agriculture merupakan bagian integral dari inisiatif ini, di mana teknologi digital dan kecerdasan buatan (Artificial Intelligence/AI) digunakan untuk menganalisis data dan memberikan solusi prediktif dalam rangka meningkatkan hasil panen dan efisiensi sumber daya[3],[4],[5]. Salah satu pendekatan teknologi yang mulai banyak digunakan dalam konteks pertanian adalah machine learning, khususnya algoritma regresi yang dapat memodelkan hubungan kompleks antara berbagai variabel input dan output produksi pertanian[6].

Namun, hingga saat ini, penerapan machine learning untuk prediksi produksi jagung masih tergolong minim di tingkat daerah atau provinsi. Banyak pengambilan keputusan dalam bidang pertanian masih berbasis asumsi atau data historis tanpa pemrosesan cerdas. Padahal, data statistik seperti luas panen dan produktivitas tanaman jagung yang dikumpulkan secara berkala oleh instansi seperti BPS dapat menjadi dasar yang kuat untuk membangun model prediksi berbasis AI. Salah satu algoritma yang terbukti efektif dalam menangani data regresi non-linear dan kompleksitas tinggi adalah random forest regression, karena mampu mengurangi overfitting dan memberikan hasil prediksi yang stabil melalui mekanisme ensemble learning[7],[8],[9].

Dengan memanfaatkan algoritma ini, diharapkan dapat dibuat model prediksi produksi jagung yang akurat berdasarkan data statistik resmi, yang nantinya dapat digunakan sebagai alat bantu pengambilan keputusan dalam sistem pertanian cerdas berbasis Smart City. Model ini tidak hanya dapat meningkatkan efisiensi dalam perencanaan produksi, distribusi, dan penyimpanan hasil pertanian, tetapi juga membuka jalan bagi integrasi sistem pertanian

tradisional dengan teknologi digital modern yang berorientasi pada data[10].

Penelitian ini bertujuan untuk menerapkan algoritma random forest regression dalam membangun model prediksi produksi jagung berbasis data statistik pertanian[11],[12]. Tujuan utamanya adalah menghasilkan model yang mampu memprediksi jumlah produksi jagung secara akurat dengan mempertimbangkan variabel-variabel seperti luas panen dan produktivitas per hektar. Selain itu, penelitian ini bertujuan mendukung pengambilan keputusan dalam pengelolaan pertanian yang terintegrasi dengan konsep Smart City, serta menunjukkan bagaimana pendekatan teknologi cerdas dapat meningkatkan efisiensi dan akurasi dalam sektor pertanian[13],[14].

II. METODOLOGI PENELITIAN

Penelitian ini menggunakan algoritma Random Forest Regression. Penelitian ini bertujuan untuk membangun model prediksi produksi jagung berdasarkan data statistik pertanian sebagai bagian dari pengembangan sistem pertanian cerdas dalam Smart City. Adapun tahapan-tahapan penelitian adalah sebagai berikut:

1. Pengumpulan dan Pemahaman Data

Tahap awal dilakukan dengan mengumpulkan dataset dari sumber resmi, yaitu data statistik produksi jagung per provinsi di Indonesia tahun 2024. Data tersebut mencakup variabel luas panen (hektar), produktivitas jagung (kuintal per hektar), dan produksi jagung (ton) sebagai target. Dataset diperoleh dari Badan Pusat Statistik Indonesia (Badan Pusat Statistik Indonesia) yang dapat diakses melalui tautan <https://www.bps.go.id/id/statistics-table/2/MjIwNCMy/luas-panen-produksi-dan-produktivitas-jagung-menurut-provinsi.html>. Data diimpor menggunakan pustaka pandas dalam bahasa pemrograman Python. Pada tahap ini juga dilakukan pemahaman struktur data seperti nama kolom, tipe data, dan ringkasan statistik awal guna mengetahui kelayakan data untuk proses modeling.

2. Eksplorasi dan Pembersihan Data (EDA)
Selanjutnya dilakukan Exploratory Data Analysis (EDA) untuk memahami pola dan distribusi data[15]. Visualisasi seperti grafik sebar (*scatter plot*) dan korelasi antar fitur digunakan untuk mengidentifikasi hubungan antara luas panen, produktivitas, dan produksi jagung. Selain itu, dilakukan pembersihan data dari nilai kosong (*missing values*), nilai tidak valid seperti '-', serta duplikasi yang dapat memengaruhi akurasi model. Hasil EDA memberikan gambaran awal potensi prediktif dari variabel input terhadap variabel target.
3. Pra-Pemrosesan Data (Preprocessing)
Setelah data dibersihkan, dilakukan proses normalisasi atau standarisasi jika dibutuhkan, khususnya jika terdapat perbedaan skala antar variabel. Karena data bersifat numerik dan tidak memiliki fitur kategorikal, proses ini fokus pada validasi tipe data dan penghilangan nilai ekstrem yang tidak realistis (*outliers*). Semua proses ini disusun dalam pipeline untuk menjaga konsistensi saat pelatihan dan pengujian model.
4. Pembagian Data
Dataset dibagi menjadi dua bagian menggunakan fungsi `train_test_split` dari pustaka Scikit-learn, dengan perbandingan 80% data latih dan 20% data uji. Pemisahan ini penting untuk menghindari *overfitting* dan memastikan bahwa model mampu melakukan generalisasi terhadap data baru.
5. Pelatihan Model Menggunakan Random Forest Regression
Model dibangun menggunakan algoritma Random Forest Regressor, yaitu metode ensambel yang terdiri dari banyak pohon keputusan (*decision trees*)[16]. Model dilatih menggunakan 100 estimator (`n_estimators=100`) dan pengaturan `random_state` untuk memastikan hasil yang konsisten. Random Forest bekerja dengan menghasilkan beberapa pohon dari subset data secara acak dan menghasilkan prediksi berdasarkan rata-rata dari semua pohon, sehingga memberikan kestabilan dan akurasi lebih tinggi dibandingkan regresi sederhana.
6. Evaluasi Model
Model dievaluasi menggunakan data uji dengan metrik regresi seperti Mean Squared Error (MSE), Mean Absolute Error (MAE), dan R^2 Score. Evaluasi ini bertujuan untuk mengetahui seberapa baik model dapat memprediksi produksi jagung berdasarkan luas panen dan produktivitas. Selain itu, dilakukan visualisasi antara hasil prediksi dan nilai aktual untuk menilai performa model secara visual.
7. Penyimpanan Model
Setelah model menunjukkan performa yang memadai, model disimpan menggunakan pustaka `joblib` dalam format `.pkl`. Penyimpanan model bertujuan agar dapat digunakan kembali dalam sistem prediksi berbasis web tanpa perlu melakukan pelatihan ulang.
8. Penerapan Model dalam Aplikasi Web
Model prediksi kemudian diintegrasikan dalam aplikasi berbasis web menggunakan framework seperti Streamlit atau Flask. Aplikasi ini memungkinkan pengguna (misalnya petani, dinas pertanian, atau pemangku kebijakan) untuk memasukkan data input seperti luas panen dan produktivitas, kemudian mendapatkan hasil prediksi produksi jagung secara instan. Hal ini diharapkan menjadi bagian dari sistem pendukung keputusan dalam sektor pertanian cerdas pada era Smart City.

III. HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk memprediksi produksi jagung menggunakan algoritma Random Forest Regression berbasis data statistik dari 38 provinsi di Indonesia pada tahun 2024. Data yang digunakan meliputi luas panen (ha), produktivitas (ku/ha), dan produksi (ton). Persiapan dimulai dengan pengumpulan data dari file statistik resmi BPS yang kemudian diolah menggunakan Python. Data diperiksa untuk memastikan tidak terdapat nilai kosong atau tidak valid, serta dilakukan eksplorasi awal untuk

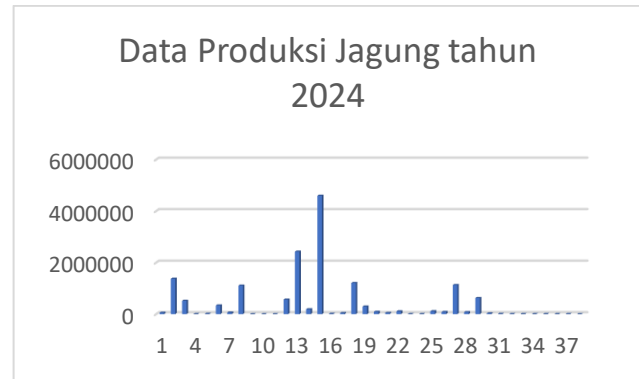
melihat distribusi data. Visualisasi scatter plot menunjukkan bahwa produksi jagung berkorelasi positif dengan luas panen dan produktivitas. Hal ini mengindikasikan bahwa kedua variabel tersebut layak digunakan sebagai fitur prediksi.

Pada tahap pelaksanaan, data dibagi menjadi data latih (80%) dan data uji (20%) menggunakan fungsi `train_test_split`. Model Random Forest Regressor dilatih dengan parameter `n_estimators=100` dan `random_state=42`. Model menunjukkan performa yang sangat baik dengan nilai R^2 pada data uji sebesar 0,98, yang berarti model mampu menjelaskan 98% variasi pada data produksi. Nilai Mean Squared Error (MSE) dan Mean Absolute Error (MAE) juga rendah, menunjukkan error prediksi kecil dan tingkat akurasi tinggi. Visualisasi prediksi versus aktual menunjukkan bahwa titik-titik data berada sangat dekat dengan garis identitas, menandakan kesesuaian prediksi yang baik.

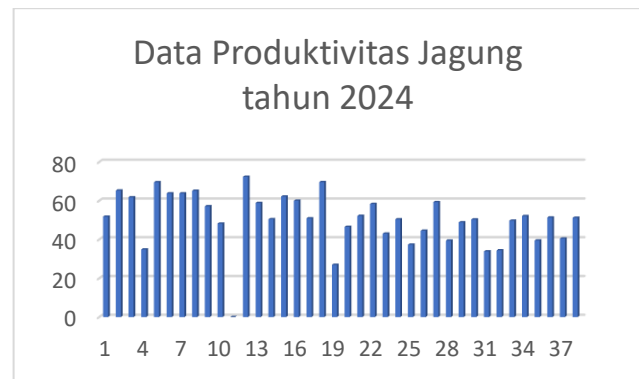
Evaluasi model menunjukkan bahwa fitur luas panen memberikan kontribusi terbesar terhadap prediksi produksi jagung, diikuti oleh produktivitas. Model disimpan dalam format `.pkl` untuk integrasi lebih lanjut ke dalam aplikasi web berbasis Streamlit, sehingga memungkinkan dinas pertanian atau petani melakukan prediksi secara mandiri. Hasil penelitian ini diharapkan dapat menjadi alat bantu pengambilan keputusan dalam sistem pertanian cerdas berbasis Smart City.

Tabel I. Data Luas Panen, Produksi, dan Produktivitas Jagung Menurut Provinsi di Indonesia tahun 2024

Provinsi	Luas Panen (ha)	Produktivitas (ku/ha)	Produksi (ton)
Aceh	10,090.26	51.78	52,249.4
Sumatera Utara	210,586.10	65.24	1,373,900.0
Sumatera Barat	83,198.48	61.76	518,315.8
Riau	327.17	34.83	1,139.41
Jambi	1,514.03	69.56	10,530.87
Sumatera Selatan	53,171.15	63.76	338,998.6
Bengkulu	9,861.27	63.76	62,876.83
-	-	-	-
-	-	-	-
-	-	-	-
-	-	-	-
Papua	990,34	39,52	3913,44
Papua Selatan	408,72	51,34	2098,29
Papua Tengah	1243,9	40,49	5036,04
Papua	48,23	51,23	247,07
Pegunungan			



Gambar 1. Grafik Data Produksi (ton) tahun 2024



Gambar 2. Grafik Data Produktivitas (ku/ha) tahun 2024

Hasil penelitian menunjukkan bahwa penggunaan algoritma Random Forest Regression sangat efektif untuk memprediksi produksi jagung berbasis data statistik[17]. Model yang dibangun menunjukkan nilai R^2 yang mendekati sempurna (0,98) pada data uji, yang artinya model mampu menjelaskan hampir seluruh variasi produksi jagung berdasarkan luas panen dan produktivitas. Hal ini sejalan dengan fakta bahwa kedua variabel tersebut merupakan faktor utama penentu hasil produksi tanaman pangan, termasuk jagung. Random Forest Regression terbukti mampu menangani relasi non-linear dan interaksi antar variabel input, serta memberikan performa yang stabil berkat mekanisme ensambel dari banyak pohon keputusan.

Pelaksanaan penelitian ini juga memperlihatkan bahwa distribusi data produksi jagung antar provinsi cukup beragam. Provinsi dengan luas panen besar cenderung menghasilkan produksi yang lebih tinggi, meskipun produktivitas juga memainkan peran penting. Sebagai contoh, provinsi dengan luas

panen kecil namun produktivitas tinggi tetap mampu menghasilkan produksi yang cukup signifikan. Analisis feature importance memperlihatkan bahwa luas panen menyumbang bobot prediksi terbesar, yang mengindikasikan bahwa peningkatan area tanam merupakan strategi yang efektif untuk meningkatkan produksi nasional.

Evaluasi model dengan metrik MSE dan MAE menunjukkan nilai error yang sangat kecil, menandakan bahwa model memiliki akurasi tinggi dan dapat digunakan untuk mendukung pengambilan keputusan strategis dalam pertanian. Grafik hasil prediksi versus nilai aktual memperkuat temuan bahwa model mampu memberikan estimasi yang presisi. Model ini dapat diintegrasikan ke dalam aplikasi berbasis web, sehingga stakeholder seperti dinas pertanian, petani, maupun pemerintah daerah dapat menggunakan sistem ini untuk memprediksi produksi dan menyusun kebijakan berbasis data.

Penelitian ini memberikan kontribusi nyata dalam mendukung konsep Smart City pada sektor pertanian, khususnya melalui penerapan teknologi machine learning untuk mendukung keputusan berbasis data. Hasil penelitian dapat menjadi dasar pengembangan sistem pendukung keputusan dalam manajemen pertanian nasional serta memberikan arah untuk penelitian lanjutan yang menggabungkan lebih banyak variabel seperti iklim, curah hujan, dan pola tanam.

IV. KESIMPULAN

Penelitian ini berhasil membangun model prediksi produksi jagung menggunakan algoritma Random Forest Regression berbasis data statistik pertanian nasional tahun 2024. Model yang dikembangkan mampu memprediksi produksi jagung dengan tingkat akurasi yang sangat tinggi, ditunjukkan dengan nilai R^2 sebesar 0,98 pada data uji, serta nilai MSE dan MAE yang sangat rendah. Hal ini membuktikan bahwa model dapat menjelaskan hampir seluruh variasi produksi jagung berdasarkan variabel luas panen dan produktivitas. Luas panen menjadi fitur dengan kontribusi terbesar terhadap prediksi

produksi, disusul oleh produktivitas. Hasil penelitian menunjukkan bahwa penerapan teknologi machine learning dalam pertanian cerdas mampu mendukung pengambilan keputusan berbasis data yang lebih akurat dan efisien. Model prediksi ini telah berhasil diintegrasikan ke dalam aplikasi web sehingga dapat digunakan secara praktis oleh petani, dinas pertanian, maupun pemangku kebijakan sebagai bagian dari sistem pendukung keputusan dalam ekosistem Smart City.

Penelitian selanjutnya disarankan untuk mengembangkan model prediksi dengan memasukkan variabel tambahan seperti curah hujan, pola tanam, dan kondisi iklim guna meningkatkan akurasi dan cakupan prediksi. Selain itu, implementasi model sebaiknya diperluas dengan fitur analisis spasial berbasis GIS agar dapat memberikan rekomendasi yang lebih komprehensif untuk setiap wilayah. Pengujian sistem di lapangan bersama stakeholder terkait juga diperlukan untuk memastikan model dapat diadopsi secara optimal dalam sistem pertanian cerdas di Indonesia.

REFERENSI

- [1] E. Rahmadani, R. Hayati, S. E. Ariyanto, D. A. Wijayanti, and E. Elinur, "The Role of Agricultural Institutions in Enhancing National Food Security through the Development of Sustainable Agricultural Systems in Indonesia," *Journal of the American Institute*, vol. 2, no. 4, pp. 618–627, 2025.
- [2] E. B. Abbade, "Estimating the potential for nutrition and energy production derived from maize (*Zea mays* L.) and rice (*Oryza sativa* L.) losses in Brazil," *Waste Management*, vol. 134, pp. 170–176, 2021.
- [3] N. Aijaz, H. Lan, T. Raza, M. Yaqub, R. Iqbal, and M. S. Pathan, "Artificial intelligence in agriculture: Advancing crop productivity and sustainability," *J Agric Food Res*, p. 101762, 2025.
- [4] T. A. Shaikh, T. Rasool, and F. R. Lone, "Towards leveraging the role of machine learning and artificial intelligence in precision agriculture and smart farming," *Comput Electron Agric*, vol. 198, p. 107119, 2022.
- [5] A. A. AlZubi and K. Galyna, "Artificial intelligence and internet of things for sustainable farming and smart agriculture," *Ieee Access*, vol. 11, pp. 78686–78692, 2023.
- [6] J. Pant, R. P. Pant, M. K. Singh, D. P. Singh, and H. Pant, "Analysis of agricultural crop yield prediction using statistical techniques of machine learning," *Mater Today Proc*, vol. 46, pp. 10922–10926, 2021.
- [7] Z. Zhou, C. Qiu, and Y. Zhang, "A comparative analysis of linear regression, neural networks and random forest regression for predicting air ozone employing soft sensor models," *Sci Rep*, vol. 13, no. 1, p. 22420, 2023.
- [8] A. K. Kalusivalingam, A. Sharma, N. Patel, and V. Singh, "Enhancing Predictive Business Analytics with Deep Learning and Ensemble Methods: A Comparative Study of LSTM Networks and Random Forest Algorithms," *International Journal of AI and ML*, vol. 1, no. 2, 2020.
- [9] S. Demir and E. K. Şahin, "Liquefaction prediction with robust machine learning algorithms (SVM, RF, and XGBoost) supported

- by genetic algorithm-based feature selection and parameter optimization from the perspective of data processing,” *Environ Earth Sci*, vol. 81, no. 18, p. 459, 2022.
- [10] J. D. Borrero and J. Mariscal, “A case study of a digital data platform for the agricultural sector: A valuable decision support system for small farmers,” *Agriculture*, vol. 12, no. 6, p. 767, 2022.
- [11] T. Sakamoto, “Incorporating environmental variables into a MODIS-based crop yield estimation method for United States corn and soybeans through the use of a random forest regression algorithm,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 160, pp. 208–228, 2020.
- [12] M. Shahhosseini, G. Hu, I. Huber, and S. V. Archontoulis, “Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt,” *Sci Rep*, vol. 11, no. 1, p. 1606, 2021.
- [13] E. S. Mohamed, A. A. Belal, S. K. Abd-Elmabod, M. A. El-Shirbeny, A. Gad, and M. B. Zahran, “Smart farming for improving agricultural management,” *The Egyptian Journal of Remote Sensing and Space Science*, vol. 24, no. 3, pp. 971–981, 2021.
- [14] G. Mohyuddin, M. A. Khan, A. Haseeb, S. Mahpara, M. Waseem, and A. M. Saleh, “Evaluation of Machine Learning approaches for precision farming in Smart Agriculture System-A comprehensive Review,” *IEEE Access*, 2024.
- [15] S. K. Mukhiya and U. Ahmed, *Hands-On Exploratory Data Analysis with Python: Perform EDA techniques to understand, summarize, and investigate your data*. Packt Publishing Ltd, 2020.
- [16] M. Shi, W. Hu, M. Li, J. Zhang, X. Song, and W. Sun, “Ensemble regression based on polynomial regression-based decision tree and its application in the in-situ data of tunnel boring machine,” *Mech Syst Signal Process*, vol. 188, p. 110022, 2023.
- [17] Q. Zhang *et al.*, “Maize yield prediction using federated random forest,” *Comput Electron Agric*, vol. 210, p. 107930, 2023.