

Prediksi Konsumsi Energi Termal Bangunan Menggunakan EPO-CatBoost

Mutiara Fadhilatuzzahro

Informatika, Universitas Pembangunan Nasional “Veteran” Jawa Timur

21081010205@student.upnjatim.ac.id

Abstrak— Sektor bangunan saat ini menjadi salah satu penyumbang terbesar konsumsi energi di dunia. Bangunan menyerap 49% dari total energi global dengan 60% diantaranya digunakan untuk kebutuhan pemanasan dan pendinginan. Semakin meningkatnya kebutuhan energi termal pada bangunan menuntut adanya sistem prediksi yang akurat untuk memperkirakan konsumsi energi sejak tahap perancangan. Model prediksi yang akurat dapat mendukung kebijakan efisiensi energi. Penelitian ini bertujuan untuk mengembangkan model prediksi konsumsi energi termal bangunan menggunakan model *CatBoost* yang dioptimasi dengan algoritma *Emperor Penguin Optimizer* (EPO). Evaluasi dilakukan menggunakan dua skenario pembagian data, yaitu 70:30 dan 80:20 untuk menilai konsistensi performa model pada berbagai proporsi data latih dan uji. Hasil penelitian menunjukkan bahwa penerapan algoritma EPO secara konsisten meningkatkan kinerja model *CatBoost*. Pada pembagian data 70:30, nilai RMSE untuk target *Heating Load* menurun dari 0,3522 menjadi 0,3306 dan *Cooling Load* dari 0,7007 menjadi 0,6239. Sedangkan, pada pembagian data 80:20 nilai RMSE menurun dari 0,3352 menjadi 0,3100 untuk *Heating Load* dan dari 0,6747 menjadi 0,5589 untuk *Cooling Load*. Dapat disimpulkan bahwa optimasi *hyperparameter* menggunakan *Emperor Penguin Optimizer* (EPO) mampu meningkatkan performa prediksi model *CatBoost* secara signifikan dan efektif untuk memperbaiki akurasi serta kestabilan model.

Kata Kunci— *CatBoost*, *Emperor Penguin Optimizer* (EPO), Optimasi *Hyperparameter*, Prediksi Konsumsi Energi Bangunan, Efisiensi Energi Bangunan.

I. PENDAHULUAN

Kesadaran global akan pentingnya efisiensi energi semakin meningkat dalam beberapa tahun terakhir. Studi yang dilakukan oleh para ahli global menunjukkan bahwa persentase konsumsi energi tertinggi berasal dari sektor bangunan [1]. Bangunan menyumbang sekitar 40% dari total energi yang dikonsumsi secara global. Bangunan membutuhkan energi dalam jumlah besar untuk mengoperasikan berbagai sistem atau peralatan seperti pemanas atau pendingin ruangan, pencahayaan, dan beragam keperluan lainnya sehingga bangunan memiliki dampak yang signifikan terhadap penggunaan energi.

Sistem HVAC (*Heating, Ventilation, and Air Conditioning*) merupakan konsumen energi terbesar di dalam bangunan, dengan kontribusi sekitar 40–50% dari total konsumsi energi [2]. Sistem ini beroperasi untuk mengatur suhu dan kelembapan udara guna mempertahankan kondisi kenyamanan termal di dalam ruangan. Sistem ini bertugas untuk memenuhi beban pemanasan (*heating load*) dan beban pendinginan (*cooling*

load) bangunan, dimana keduanya berkaitan erat dengan beban termal (*thermal load*) bangunan. Kinerja sistem HVAC sangat dipengaruhi oleh berbagai faktor seperti kondisi cuaca, karakteristik termal bangunan, pola okupansi, serta pengaturan sistem kontrol. Variasi pada faktor-faktor tersebut dapat menyebabkan fluktuasi yang signifikan terhadap total konsumsi energi. Oleh karena itu diperlukan upaya dalam mengurangi konsumsi energi pada bangunan, terutama energi termal guna mengurangi dampak perubahan iklim dan mendukung keberlanjutan [3].

Upaya penghematan energi pada bangunan memerlukan perumusan berbagai strategi untuk mengelola penggunaan energi secara efektif. Salah satu aspek penting dalam merancang strategi tersebut adalah mengembangkan model prediksi konsumsi energi yang akurat. Model prediksi yang akurat memungkinkan pengambilan keputusan yang lebih tepat untuk mendukung penghematan energi. Model *Machine Learning* seperti *CatBoost* menunjukkan performa kompetitif pada tugas prediksi yang melibatkan fitur heterogen, namun performanya sangat bergantung pada pemilihan *hyperparameter* yang sesuai. Oleh karena itu, diperlukan proses *hyperparameter tuning* untuk menyesuaikan *hyperparameter* model agar performa prediksi menjadi optimal dan lebih andal. Peningkatan akurasi prediksi, meskipun relatif kecil nyatanya berpotensi menghasilkan pengurangan konsumsi energi yang signifikan [4].

Dalam penelitian ini, penulis mengusulkan EPO-CatBoost, yakni integrasi *Emperor Penguin Optimizer* (EPO) untuk optimasi *hyperparameter* pada model *CatBoost* dalam memprediksi konsumsi energi termal. Algoritma EPO pada diterapkan untuk meningkatkan akurasi prediksi beban termal guna mendukung penerapan strategi penghematan energi yang lebih efektif.

II. LANDASAN TEORI

A. Konsumsi Energi Bangunan

Konsumsi energi bangunan terdiri dari energi yang dikonsumsi oleh peralatan untuk pendinginan, pemanasan, ventilasi, pencahayaan, kipas, dan beberapa peralatan listrik lainnya [5]. Konsumsi energi pada bangunan berkaitan dengan pemakaian energi untuk beragam keperluan, antara lain:

1. Energi termal yang dikeluarkan untuk mengkondisikan suhu dalam bangunan.
2. Energi mekanik yang dikeluarkan untuk menjalankan utilitas bangunan.

3. Energi cahaya yang digunakan untuk menciptakan pencahayaan buatan.

Keperluan energi tersebut dilakukan untuk mendukung berjalannya sistem dalam bangunan seperti sistem ventilasi, pencahayaan, peralatan pendukung, proses produksi, serta peralatan yang memanfaatkan energi utama. Berdasarkan sumber energinya, faktor-faktor yang mempengaruhi konsumsi energi terbagi menjadi dua kategori, yakni faktor eksternal (luar bangunan) yang bergantung pada alam seperti iklim atau cuaca dan faktor internal (dalam bangunan) yang bergantung pada kebutuhan atau aktivitas penghuni bangunan [6]. Efisiensi energi pada bangunan merujuk pada usaha untuk mengurangi jumlah energi yang diperlukan dalam menyediakan barang dan layanan untuk bangunan. misalnya, teknologi isolasi rumah memungkinkan penggunaan lebih sedikit untuk pemanasan dan pendinginan guna mencapai dan mempertahankan suhu yang nyaman [7]. Dalam mencapai efisiensi energi pada bangunan tidak hanya berfokus pada pembangunan struktur ramah lingkungan baru atau merenovasi bangunan yang sudah ada dengan sistem yang lebih efisien, melainkan juga melibatkan pemikiran mendalam secara menyeluruh mengenai seluruh siklus hidup bangunan, mulai dari perancangan dan pembangunan hingga penggunaan, perbaikan, dan penghentian operasional [8].

B. CatBoost

CatBoost adalah algoritma ML yang dikembangkan dengan memanfaatkan teknik *gradient boosting* pada pohon keputusan (GBDT). Algoritma ini pertama kali diperkenalkan oleh tim insinyur Yandex pada tahun 2017. CatBoost menggunakan teknik peringkat promosi untuk mengoptimalkan kinerja GBDT, memungkinkan algoritma ini untuk memanfaatkan seluruh dataset dalam proses pelatihan. Selain itu, CatBoost juga dirancang untuk mengurangi risiko *overfitting* terjadi selama pelatihan model [9]. Algoritma CatBoost memiliki kemampuan bawaan untuk mengonversi nilai data non-numerik menjadi nilai numerik secara efisien dan otomatis, tanpa memerlukan pra-pemrosesan manual yang kompleks sehingga dapat menghasilkan hasil yang baik dalam satu kali eksekusi untuk pengambilan keputusan. Pada kasus regresi, CatBoost berfokus pada mengurangi kesalahan prediksi dengan memanfaatkan decision trees yang ditingkatkan menggunakan *gradient boosting* [10].

C. Emperor Penguin Optimizer (EPO)

Emperor Penguin Optimizer terinspirasi dari perilaku berkumpul penguin kaisar di Antartika. Dhiman pertama kali mengusulkannya pada tahun 2018. Penguin kaisar adalah hewan sosial yang mencari makan dan berburu dalam kelompok [11]. Konsep utama dari algoritma EPO adalah meniru pergerakan dan strategi bertahan hidup penguin kaisar untuk melakukan pencarian solusi optimal. Spesies ini terkenal dengan perilakunya dalam membentuk kelompok padat yang disebut *huddle*, di mana setiap individu saling merapat untuk mempertahankan kehangatan tubuh di tengah suhu yang ekstrem. Dalam formasi *huddle*, penguin secara dinamis bergerak, bergantian menempati posisi di bagian dalam dan luar kelompok untuk mendistribusikan panas secara merata.

EPO dikembangkan sebagai respons terhadap perilaku berkumpul penguin kaisar dalam menemukan solusi yang paling tepat dengan mencoba banyak pilihan (sumber kehangatan) dan fokus pada pilihan yang menjanjikan secara seimbang [12].

III. METODOLOGI

A. Pengumpulan Data

Dataset yang digunakan bersumber dari Web UCI *Machine Learning Repository*. Dataset ini dapat diakses melalui <https://archive.ics.uci.edu/dataset/242/energy+efficiency>. Dataset terdiri atas 8 variabel input dan 2 variabel output (target), yaitu *Heating Load* (Y1) dan *Cooling Load* (Y2).

B. Pembagian Data

Dataset dibagi menjadi dua bagian dengan dua skenario, yaitu 70% data latih (*training set*) dan 30% data uji (*testing set*) serta 80% data latih dan 20% data uji menggunakan fungsi *train_test_split* dari *Scikit-learn* dengan parameter *random_state=42* agar replikasi hasil tetap konsisten.

C. Pengembangan Model

Model dasar yang digunakan dalam penelitian ini adalah CatBoost. Proses optimasi dilakukan terhadap beberapa *hyperparameter* utama, yaitu *learning_rate*, *depth*, *l2_leaf_reg*, *bagging_temperature*, dan *border_count*. Dalam mencari kombinasi *hyperparameter* terbaik, digunakan algoritma *Emperor Penguin Optimizer* (EPO) dengan fungsi fitness berupa *Root Mean Square Error* (RMSE) pada data uji. Optimasi dilakukan dengan ukuran populasi sebanyak 20 agen dan 50 iterasi. Setiap agen (penguin) merepresentasikan kombinasi *hyperparameter* CatBoost yang berbeda, dan pergerakannya mengikuti mekanisme optimasi EPO.

D. Evaluasi Model

Setelah diperoleh kombinasi *hyperparameter* terbaik, model CatBoost diuji menggunakan data uji untuk menghitung performa prediksi menggunakan metrik *Root Mean Square Error* (RMSE), *Mean Absolute Error* (MAE), dan *Coefficient of Determination* (R^2).

E. Analisis Hasil dan Evaluasi

Hasil RMSE, MAE, dan R^2 dari model CatBoost standar dan CatBoost yang dioptimasi dengan EPO dibandingkan untuk menilai pengaruh penerapan algoritma EPO terhadap peningkatan akurasi prediksi konsumsi energi termal bangunan. Selain itu, perbedaan hasil antara skenario 70:30 dan 80:20 dianalisis untuk melihat pengaruh proporsi data pelatihan terhadap kemampuan generalisasi model.

IV. HASIL DAN PEMBAHASAN

Penelitian ini mengimplementasikan model CatBoost untuk memprediksi konsumsi energi termal bangunan dengan dua variabel target, yaitu *Heating Load* (HL) dan *Cooling Load* (CL). Dalam menilai pengaruh proporsi pembagian data terhadap kinerja model, diterapkan dua skenario pembagian data meliputi rasio 70:30 dan 80:20. Kinerja model dievaluasi menggunakan tiga metrik utama diantaranya yaitu RMSE, MAE, R^2 . Semakin kecil nilai RMSE dan MAE menunjukkan

kesalahan prediksi yang rendah. Sementara itu, nilai R^2 yang mendekati 1 menandakan kemampuan model menjelaskan variasi data aktual dengan baik. Tabel 1 menunjukkan hasil evaluasi model untuk target *Heating Load* (HL) pada pembagian data 70:30.

TABEL 1. HASIL UNTUK TARGET HEATING LOAD PADA PEMBAGIAN 70:30

Model	Metrik Evaluasi		
	RMSE	MAE	R^2
CatBoost	0.3522	0.2461	0.9988
EPO-CatBoost	0.3306	0.2196	0.9989

Dari hasil Tabel 1, terlihat bahwa model EPO-CatBoost mampu menurunkan nilai RMSE dari 0.3522 menjadi 0.3306 dan MAE dari 0.2461 menjadi 0.2196. Selain itu, nilai R^2 mengalami peningkatan dari 0.9988 menjadi 0.9989 yang menandakan adanya peningkatan kemampuan model dalam menjelaskan hubungan antara variabel input dan target.

TABEL 2. HASIL UNTUK TARGET COOLING LOAD PADA PEMBAGIAN 70:30

Model	Metrik Evaluasi		
	RMSE	MAE	R^2
CatBoost	0.7007	0.4770	0.9946
EPO-CatBoost	0.6239	0.3988	0.9957

Tabel 3 menampilkan hasil pengujian untuk target *Cooling Load* (CL) pada pembagian data 70:30. Pada prediksi target CL dengan rasio 70:30, Model EPO-CatBoost juga menunjukkan performa yang lebih baik dengan penurunan RMSE sekitar 10,9% dan MAE sekitar 16,4% dibandingkan model CatBoost tanpa optimasi. Nilai R^2 juga mengalami peningkatan dari 0.9946 menjadi 0.9957.

TABEL 3. HASIL UNTUK TARGET HEATING LOAD PADA PEMBAGIAN 80:20

Model	Metrik Evaluasi		
	RMSE	MAE	R^2
CatBoost	0.3352	0.2398	0.9989
EPO-CatBoost	0.3100	0.2052	0.9991

Tabel 3 menampilkan hasil pengujian untuk target HL pada pembagian data 80:20. Pada prediksi target HL dengan rasio 80:20, terlihat CatBoost-EPO menurunkan nilai RMSE dari 0.3352 menjadi 0.3100 dan MAE dari 0.2398 menjadi 0.2052. Nilai R^2 mengalami peningkatan menjadi 0.9991. Hasil untuk target *Cooling Load* (CL) pada skenario pembagian data 80:20 ditunjukkan pada Tabel 4.

TABEL 4. HASIL UNTUK TARGET COOLING LOAD PADA PEMBAGIAN 80:20

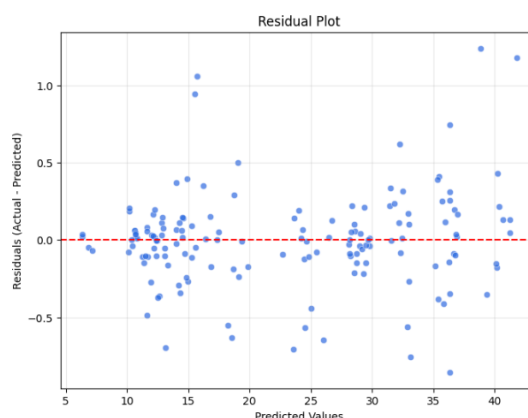
Model	Metrik Evaluasi		
	RMSE	MAE	R^2
CatBoost	0.6747	0.4437	0.9951
EPO-CatBoost	0.5589	0.3414	0.9966

Pada target *Cooling Load*, model hasil optimasi (EPO–CatBoost) menghasilkan penurunan RMSE sebesar 17,1% dan MAE sebesar 23,0% dibandingkan model standar. Peningkatan nilai R^2 dari 0.9951 menjadi 0.9966 juga memperkuat bukti bahwa optimasi *hyperparameter* dengan algoritma EPO efektif meningkatkan kemampuan generalisasi model terhadap data baru.

Secara keseluruhan, dapat dilihat bahwa algoritma EPO memberikan peningkatan kinerja yang konsisten pada semua skenario, terutama pada prediksi *Cooling Load* dengan pembagian data 80:20 yang menunjukkan peningkatan paling signifikan. Hasil penelitian ini membuktikan bahwa optimasi *hyperparameter* menggunakan *Emperor Penguin Optimizer* (EPO) secara signifikan meningkatkan performa model CatBoost dalam memprediksi konsumsi energi termal bangunan. EPO mampu menyesuaikan kombinasi parameter penting seperti *learning rate*, *depth*, *l2_leaf_reg*, *bagging temperature*, dan *border count* secara adaptif sehingga model lebih optimal dalam menghasilkan prediksi.

Selanjutnya, visualisasi residual plot dan distribusi error digunakan untuk menilai karakteristik kesalahan prediksi model CatBoost setelah dilakukan optimasi *hyperparameter* menggunakan *Emperor Penguin Optimizer* (EPO). Analisis ini difokuskan pada hasil pada target HL dengan rasio data 80:20 karena skenario tersebut memberikan hasil metrik evaluasi terbaik pada penelitian ini. Gambar 1 menampilkan visualisasi residual plot model CatBoost hasil optimasi EPO pada target HL dengan rasio data 80:20.

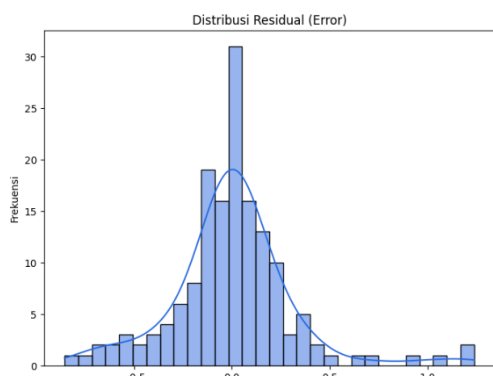
GAMBAR 1. RESIDUAL PLOT



Pada residual plot, titik-titik residual terlihat tersebar secara acak di sekitar garis horizontal nol tanpa pola yang jelas. Ini menandakan bahwa model CatBoost yang telah dioptimasi

dengan *Emperor Penguin Optimizer* (EPO) berhasil menangkap hubungan antara variabel input dan target dengan baik, sehingga kesalahan prediksi bersifat acak dan tidak sistematis. Sebaran residual yang cukup merata di seluruh rentang nilai prediksi juga mengindikasikan bahwa model tidak mengalami masalah heteroskedastisitas sehingga performa prediksi cukup stabil pada berbagai nilai output. Meski terdapat beberapa residual yang cukup besar (*outlier*), jumlahnya relatif sedikit dan tidak mengganggu pola umum error. Gambar 2 menampilkan visualisasi histogram residual model CatBoost hasil optimasi EPO pada target HL dengan rasio data 80:20.

GAMBAR 2. HISTOGRAM RESIDUAL



Distribusi residual model menunjukkan pola yang mendekati distribusi normal dengan puncak di sekitar nol. Mayoritas residual berkisar di antara -0.5 hingga 0.5, yang menandakan bahwa sebagian besar prediksi model memiliki kesalahan yang relatif kecil dan terpusat di sekitar nilai aktual. Kurva *kernel density estimation* (KDE) yang mengikuti pola histogram memperkuat kesan bahwa error model bersifat simetris dan tidak condong ke satu arah tertentu. Kondisi ini menunjukkan bahwa model tidak mengalami bias prediksi yang signifikan, baik overprediction maupun underprediction. Meski terdapat sedikit ekor distribusi yang melebar ke arah nilai residual yang lebih ekstrem, hal ini masih dalam batas wajar.

Secara keseluruhan, hasil visualisasi ini memperkuat hasil metrik evaluasi kuantitatif (RMSE, MAE, dan R^2), yang menunjukkan bahwa optimasi *hyperparameter* menggunakan EPO berhasil meningkatkan akurasi dan stabilitas model CatBoost. Dengan kombinasi parameter optimal yang diperoleh (*learning rate*, *depth*, *l2_leaf_reg*, *bagging temperature*, dan *border count*), model mampu menghasilkan prediksi yang lebih akurat dan andal dibandingkan dengan model tanpa optimasi.

V. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa model CatBoost mampu memberikan performa prediksi yang sangat baik terhadap konsumsi energi

termal bangunan, baik pada target *Heating Load* maupun *Cooling Load*. Model ini secara efektif mempelajari hubungan kompleks antara faktor-faktor yang memengaruhi konsumsi energi bangunan. Nilai koefisien determinasi (R^2) yang tinggi, yaitu di atas 0,99, menunjukkan bahwa sebagian besar variasi data dapat dijelaskan oleh model CatBoost. Optimasi *hyperparameter* menggunakan algoritma EPO terbukti meningkatkan kinerja model secara signifikan. Nilai RMSE dan MAE mengalami penurunan, sementara nilai R^2 meningkat pada kedua skenario pembagian data (70:30 dan 80:20). Hasil tersebut menunjukkan bahwa kombinasi EPO-CatBoost efektif dalam memperbaiki akurasi serta kestabilan model.

REFERENSI

- [1] A. H. Alharbi, D. S. Khafaga, A. M. Zaki, E.-S. M. El-Kenawy, A. Ibrahim, A. A. Abdelhamid, M. M. Eid, M. El-Said, N. Khodadadi, L. Abualigah, and M. A. Saeed, "Forecasting of energy efficiency in buildings using multilayer perceptron regressor with waterwheel plant algorithm hyperparameter," *Frontiers in Energy Research*, vol. 12, 2024. doi:10.3389/fenrg.2024.1393794.
- [2] B. Mehdizadeh Khorrami, A. Soleimani, A. Pinnarelli, G. Brusco, & P. Vizza, "Forecasting heating and cooling loads in residential buildings using machine learning: a comparative study of techniques and influential indicators," *Asian Journal of Civil Engineering*, vol. 25, no. 2, pp. 1163–1177, 2024. doi:10.1007/s42107-023-00834-8.
- [3] R. A. Ibrahim, N. Belouaggadia, and L. Zalewski, "Application of Ventilated Solar Façades to enhance the energy efficiency of buildings: A comprehensive review," *Energy Reports*, vol. 13, pp. 1266–1292, 2025. doi:10.1016/j.egy.2024.12.051.
- [4] S. Afzal, B. M. Ziapour, A. Shokri, H. Shakibi, and B. Sobhani, "Building energy consumption prediction using multilayer perceptron neural network-assisted models; comparison of different optimization algorithms," *Energy*, vol. 282, 2023, Art. no. 128446, doi:10.1016/j.energy.2023.128446.
- [5] F. Asdrubali and U. Desideri, "Introduction," in *Handbook of Energy Efficiency in Buildings: A Life Cycle Approach*, Elsevier, 2019, ch. 1, pp. 1-3, doi:10.1016/B978-0-12-812817-6.00001-2.
- [6] V. S. C. Widardji, *Objektivitas Target Nilai Intensitas Konsumsi Energi (IKE) dari GBCI dan ASEAN-USAID pada Bangunan Hotel di Indonesia*, S2 thesis, Universitas Atma Jaya Yogyakarta, Yogyakarta, Indonesia, 2017. [Online]. Available: <https://repository.uajy.ac.id/eprint/11306/>.
- [7] M. Yang and X. Yu, *Energy Efficiency: Benefits for Environment and Society*, Green Energy and Technology. London: Springer-Verlag, 2015. doi:10.1007/978-1-4471-6666-5.
- [8] B. N. Jørgensen and Z. Ma, "Towards energy efficient buildings by digital transformation of the building lifecycle," *Energy Informatics*, vol. 7, Article no. 81, 2024. doi:10.1186/s42162-024-00405-4.
- [9] J. Zheng, M. Hu, C. Wang, S. Wang, B. Han, and H. Wang, "Spatial patterns of residents' daily activity space and its influencing factors based on the CatBoost model: A case study of Nanjing, China," *Frontiers of Architectural Research*, vol. 11, no. 6, pp. 1193-1204, 2022. doi:10.1016/j.foar.2022.04.003.
- [10] M. Abdullah and S. Said, "Performance evaluation of machine learning regression models for rainfall prediction," *Preprint*, 2023. doi:10.21203/rs.3.rs-3258529/v1.
- [11] H. Jia, K. Sun, W. Song, X. Peng, C. Lang, and Y. Li, "Multi-Strategy Emperor Penguin Optimizer for RGB Histogram-Based Color Satellite Image Segmentation Using Masi Entropy," *IEEE Access*, vol. 7, pp. 134448-134474, 2019. doi:10.1109/ACCESS.2019.2942064.
- [12] M. A. Angel and T. Jaya, "An Enhanced Emperor Penguin Optimization Algorithm for Secure Energy Efficient Load Balancing in Wireless Sensor Networks," *Wireless Personal Communications*, vol. 125, no. 3, pp. 2101-2127, 2022. doi:10.1007/s11277-022-09647-5.