

Optimasi Model XGBoost dengan Optuna untuk Prediksi Wisatawan Mancanegara Bandara Kualanamu

Maisie Yunita Malva¹, Anggraini Puspita Sari^{2*}, Eva Yulia Puspaningrum³

^{1,2,3} Informatika, Universitas Pembangunan Nasional “Veteran” Jawa Timur

¹121081010038@student.upnjatim.ac.id

³evapuspaningrum.if@upnjatim.ac.id

*Corresponding author email: anggraini.puspita.if@upnjatim.ac.id

Abstrak— Pariwisata merupakan salah satu sektor penting dalam perekonomian Indonesia yang memberikan kontribusi signifikan terhadap pertumbuhan ekonomi melalui peningkatan devisa dan perluasan lapangan kerja. Bandara Internasional Kualanamu sebagai gerbang utama wisatawan mancanegara (wisman) di Sumatera Utara mengalami fluktuasi jumlah kunjungan yang dipengaruhi oleh berbagai faktor ekonomi, sosial, dan global. Penelitian ini bertujuan mengembangkan model prediksi jumlah kunjungan wisman menggunakan algoritma XGBoost yang dioptimasi dengan Optuna untuk meningkatkan akurasi peramalan. Data yang digunakan merupakan data deret waktu bulanan dari Badan Pusat Statistik periode Januari 2008 hingga Desember 2024 dengan total 204 observasi. Proses penelitian meliputi tahap *preprocessing* data (*data cleaning*, konversi tipe data, penanganan *outlier*, dan normalisasi menggunakan *Min-Max Scaler*), pembagian data dengan rasio 80:20, pemodelan XGBoost, optimasi *hyperparameter* menggunakan Optuna berbasis TPE, dan evaluasi model menggunakan metrik MAE, RMSE, dan MAPE. Hasil penelitian menunjukkan nilai MAE sebesar 0,1347, RMSE sebesar 0,1749, dan MAPE sebesar 13,47%, yang mengindikasikan performa yang dapat diterima untuk data dengan volatilitas tinggi. Model mampu menangkap tren umum dengan baik, namun memiliki keterbatasan dalam memprediksi fluktuasi tajam dan lonjakan ekstrem pada periode 2023-2024. Penerapan Optuna terbukti efektif dalam meningkatkan efisiensi optimasi *hyperparameter*. Model ini dapat digunakan untuk mendukung perencanaan strategis pariwisata jangka menengah di Sumatera Utara.

Kata Kunci— XGBoost, Optuna, prediksi wisman, *time series forecasting*, Bandara Kualanamu, *machine learning*, optimasi *hyperparameter*.

I. PENDAHULUAN

Indonesia memiliki potensi pariwisata yang sangat besar dengan keanekaragaman alam, budaya, dan keunikan daerah yang menjadi daya tarik bagi wisatawan mancanegara (wisman). Sebagai salah satu sektor utama, pariwisata berperan penting dalam mendorong pertumbuhan ekonomi melalui peningkatan devisa, perluasan lapangan kerja, dan peningkatan kesejahteraan masyarakat. Menurut Badan Pusat Statistik (BPS), sebagian wisman masuk melalui jalur udara, dengan Bandara Internasional Kualanamu sebagai salah satu gerbang utama di wilayah Sumatera Utara. Fluktuasi jumlah kunjungan wisman yang dipengaruhi oleh berbagai faktor ekonomi, sosial, dan global menunjukkan perlunya model prediksi yang akurat

untuk mendukung perencanaan kebijakan pariwisata yang efektif [1], [2].

Metode peramalan (*forecasting*) data deret waktu (*time series*) telah banyak digunakan dalam analisis pariwisata, antara lain model *Seasonal Autoregressive Integrated Moving Average* (SARIMA) dan SARIMAX. Model-model klasik ini telah terbukti mampu memberikan prediksi yang cukup baik untuk data dengan pola musiman yang jelas. Namun, model-model tersebut memiliki keterbatasan dalam menangkap hubungan *non-linear* yang kompleks antar variabel serta memerlukan asumsi stasioneritas data [2]. Selain itu, proses identifikasi model dan pemilihan parameter pada metode statistik tradisional sering kali memerlukan pengetahuan statistik yang mendalam dan proses iteratif yang cukup kompleks.

Pendekatan berbasis jaringan syaraf tiruan seperti *Multilayer Perceptron* (MLP) juga telah diterapkan dalam peramalan wisatawan dan menunjukkan kemampuan dalam menangkap pola *non-linear*. Namun, metode ini memiliki beberapa kelemahan, yaitu membutuhkan data dalam jumlah besar, rentan terhadap *overfitting*, dan sulit diinterpretasikan karena bersifat *black box*. Oleh karena itu, diperlukan pendekatan alternatif yang lebih efisien dalam hal komputasi dan dapat menghasilkan akurasi yang lebih tinggi. [3].

Perkembangan teknologi *machine learning* menghadirkan alternatif yang lebih adaptif terhadap karakteristik data pariwisata yang kompleks dan *non-stasioner*. Salah satu algoritma yang banyak digunakan adalah *Extreme Gradient Boosting* (XGBoost), yaitu algoritma *ensemble boosting* berbasis pohon keputusan yang mampu menangkap hubungan *non-linear* dan interaksi kompleks antar fitur [4]. XGBoost telah terbukti efektif dalam berbagai aplikasi *time series forecasting* dengan performa yang kompetitif [5]. Dibandingkan dengan model tradisional seperti ARIMA atau SARIMA, XGBoost mampu menghasilkan prediksi yang lebih stabil pada data dengan pola yang kompleks. Namun, performa XGBoost sangat bergantung pada pemilihan *hyperparameter* yang optimal. Proses penyetelan parameter secara manual atau menggunakan *grid search* sering kali memakan waktu dan kurang efisien. Untuk mengatasi hal tersebut, Optuna digunakan sebagai kerangka kerja otomatisasi optimasi *hyperparameter* berbasis *Tree-structured Parzen Estimator* (TPE) yang dapat menemukan kombinasi parameter terbaik secara efisien [6].

Dalam penelitian ini, dikembangkan model prediksi jumlah kunjungan wisatawan mancanegara di Bandara Internasional

Kualanamu menggunakan algoritma XGBoost dengan optimasi Optuna. Model ini dirancang untuk meningkatkan efisiensi peramalan dengan memanfaatkan proses optimasi *hyperparameter* otomatis yang adaptif. Hasil penelitian diharapkan dapat memberikan kontribusi dalam pengembangan metode peramalan berbasis *machine learning*, khususnya dalam penerapan algoritma *gradient boosting* yang lebih optimal pada analisis data pariwisata.

II. METODOLOGI PENELITIAN

Beberapa tahapan dilakukan dalam penelitian ini sebagai proses utama dalam membangun model peramalan, meliputi tahap pengumpulan data hingga evaluasi model.

A. Pengumpulan Data

Pada penelitian ini, data yang digunakan untuk memprediksi jumlah kunjungan wisatawan mancanegara diperoleh dari situs resmi Badan Pusat Statistik (BPS) melalui laman <https://www.bps.go.id>. Data tersebut berisi jumlah kunjungan wisman melalui Bandara Kualanamu yang disajikan dalam bentuk deret waktu bulanan mulai dari Januari 2008 hingga Desember 2024, dengan total 204 data observasi.

B. Preprocessing Data

Tahap *preprocessing* dilakukan untuk memastikan kualitas dan konsistensi data sebelum digunakan pada proses pemodelan. Beberapa langkah yang dilakukan meliputi:

- *Data Cleaning*

Pada tahap ini dilakukan pemeriksaan terhadap nilai kosong (*missing values*), duplikasi, serta data yang tidak valid. Nilai kosong ditangani menggunakan metode interpolasi agar data tetap utuh dan representatif.

- *Konversi Tipe Data*

Dalam tahap ini seluruh data waktu diubah ke dalam format *datetime* agar dapat diproses dengan benar dalam analisis deret waktu dan pembagian data.

- *Penanganan Outlier*

Tahap ini dilakukan identifikasi terhadap nilai-nilai ekstrem (*outlier*) yang berpotensi memengaruhi hasil analisis dan kinerja model. Deteksi *outlier* dilakukan menggunakan pendekatan Interquartile Range (IQR), yaitu dengan menghitung selisih antara kuartil ketiga (Q3) dan kuartil pertama (Q1).

- *Normalisasi data*

Tahap normalisasi dilakukan untuk menyeragamkan skala nilai pada setiap fitur agar model dapat mempelajari pola data secara lebih stabil dan efisien. Proses ini menggunakan metode *Min-Max Scaler*, yang mengubah rentang nilai setiap fitur ke dalam skala 0 hingga 1.

C. Pembagian Data

Data dibagi menjadi dua bagian, yaitu data latih (*training set*) sebesar 80% dan data uji (*testing set*) sebesar 20%. Pembagian dilakukan secara berurutan berdasarkan waktu untuk mencegah terjadinya *data leakage*.

D. Pemodelan XGBoost

Tahap ini melibatkan pembangunan model prediksi menggunakan algoritma XGBoost. Model dilatih menggunakan data latih yang telah diproses untuk mempelajari pola hubungan antara waktu dan jumlah kunjungan wisman. XGBoost merupakan algoritma berbasis *ensemble* yang menggabungkan sejumlah pohon keputusan (*decision trees*) untuk meminimalkan kesalahan prediksi secara bertahap [4]. Secara umum, prediksi model XGBoost dapat dinyatakan dalam persamaan (1).

$$\hat{y}_i = \phi(x_i) = \sum_{k=1}^K f_k(x_i), \quad f_k \in F \quad (1)$$

Dengan \hat{y}_i adalah prediksi untuk data ke- i , $\phi(x_i)$ adalah fungsi *ensemble*, $f_k(x_i)$ adalah nilai prediksi dari pohon ke- k untuk sampel ke- i . Fungsi objektif XGBoost bertujuan untuk meminimalkan *loss function* dengan menambahkan *regularization term* guna mengontrol kompleksitas model. Secara matematis, fungsi objektif dirumuskan pada (2).

$$L(\phi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \quad (2)$$

Dengan $l(\hat{y}_i, y_i)$ adalah *loss function* yang mengukur selisih antara prediksi \hat{y}_i dan nilai aktual y_i , sedangkan $\Omega(f_k)$ adalah fungsi regularisasi untuk menghindari *overfitting*. Fungsi regularisasi tersebut dinyatakan dalam (3).

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2 \quad (3)$$

Dengan T adalah jumlah daun dalam pohon keputusan, γ adalah parameter penalti untuk jumlah daun, dan λ parameter regularisasi bobot. Untuk mempercepat proses optimasi, XGBoost menggunakan ekspansi Taylor orde kedua pada *loss function*, sebagaimana dinyatakan dalam (4).

$$L^{(t)} \approx \sum_{i=1}^n \left[l(y_i, \hat{y}_i^{(t-1)}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t) \quad (4)$$

Dengan g_i adalah turunan pertama dan h_i adalah turunan kedua dari *loss function* terhadap prediksi sebelumnya. Bobot optimal pada setiap daun dihitung dengan menurunkan fungsi objektif terhadap bobot dan menyamakannya dengan nol, menghasilkan (5).

$$w_j^* = -\frac{G_j}{H_j + \lambda} \quad (5)$$

Dengan $G_j = \sum_{i \in I_j} g_i$ dan $H_j = \sum_{i \in I_j} h_i$, yaitu jumlah gradien dan Hessian untuk daun ke- j .

Setelah bobot dihitung, XGBoost mengevaluasi setiap kemungkinan pembagian (*split*) untuk menentukan titik pemisahan terbaik berdasarkan *gain*, sebagaimana pada (6):

$$L_{split} = \frac{1}{2} \left[\frac{(\sum_{i \in I_L} g_i)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{(\sum_{i \in I_R} g_i)^2}{\sum_{i \in I_R} h_i + \lambda} - \frac{(\sum_{i \in I} g_i)^2}{\sum_{i \in I} h_i + \lambda} \right] - \gamma \quad (6)$$

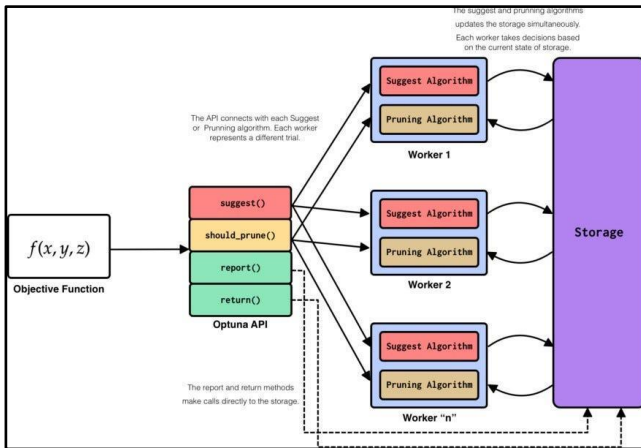
Dengan I_L dan I_R masing-masing adalah himpunan data pada node kiri dan kanan setelah pemisahan, serta $I = I_L \cup I_R$ adalah himpunan data sebelum pemisahan. *Split* dipilih hanya jika $L_{split} > 0$, artinya pembagian tersebut memberikan pengurangan *loss* yang lebih besar daripada penalti kompleksitas γ .

Setelah pohon optimal terbentuk, hasil prediksi diperbarui secara iteratif menggunakan *learning rate* (η) untuk mengontrol kontribusi setiap pohon terhadap model keseluruhan sebagaimana pada (7).

$$\hat{y}_i^{(t)} = \hat{y}_i^{(0)} + \eta \cdot f_t(x_i) \quad (7)$$

Dengan $\hat{y}_i^{(t)}$ adalah prediksi pada iterasi ke- t dan $f_t(x_i)$ hasil dari pohon baru pada iterasi tersebut. Melalui kombinasi mekanisme regularisasi, pembelajaran aditif, dan pembaruan berbasis gradien, XGBoost mampu menghasilkan model prediksi yang akurat, efisien, serta tahan terhadap *overfitting*.

E. Optimasi Hyperparameter dengan Optuna



Gbr. 1 Arsitektur proses optimasi Optuna

Proses optimasi *hyperparameter* dilakukan menggunakan Optuna, yaitu *framework* otomatisasi pencarian parameter terbaik berbasis metode TPE [7]. Optuna digunakan untuk meningkatkan kinerja model XGBoost dengan mencari kombinasi parameter paling optimal, seperti *learning_rate*, *max_depth*, *n_estimators*, dan *subsample*. Melalui pendekatan adaptif, Optuna secara efisien mengevaluasi berbagai konfigurasi model dan memilih hasil dengan tingkat kesalahan

prediksi terendah berdasarkan metrik evaluasi yang digunakan. Arsitektur proses Optimasi Optuna dapat dilihat pada Gbr. 1.

F. Evaluasi

Tahap evaluasi dilakukan untuk menilai kinerja model dalam memprediksi jumlah kunjungan wisatawan mancanegara berdasarkan data aktual. Proses ini bertujuan untuk mengukur tingkat kesalahan dan ketepatan hasil prediksi yang dihasilkan oleh model. Evaluasi dilakukan dengan menggunakan beberapa metrik statistik, yaitu *Mean Absolute Error* (MAE) untuk menghitung rata-rata selisih absolut antara nilai prediksi dan nilai aktual [8], sebagaimana pada (8).

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (8)$$

Dengan y_i adalah nilai aktual, \hat{y}_i adalah nilai yang diprediksi, dan n adalah jumlah data. Semakin kecil nilai MAE, semakin tinggi akurasi prediksi model [9]. *Root Mean Square Error* (RMSE) untuk mengukur besar kesalahan prediksi secara keseluruhan [10], sebagaimana pada (9).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (9)$$

Dengan y_i , \hat{y}_i , dan n memiliki definisi yang sama seperti pada (8). Nilai RMSE yang lebih kecil menunjukkan bahwa hasil prediksi model mendekati nilai aktual secara keseluruhan.

Mean Absolute Percentage Error (MAPE) yang menunjukkan tingkat kesalahan dalam bentuk persentase [11], [12].

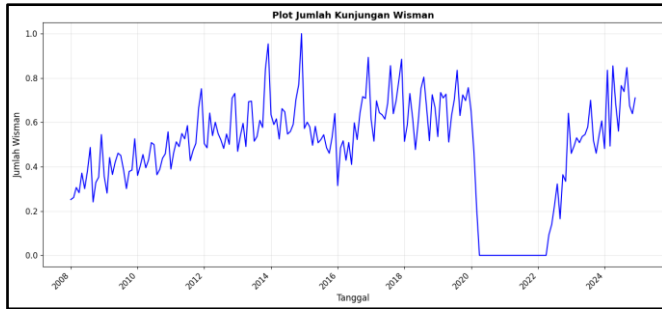
$$MAPE = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{x_t - y_t}{x_t} \right| \quad (10)$$

Dengan y_i , \hat{y}_i , dan n memiliki definisi yang sama seperti pada (8). Nilai MAPE yang kecil menunjukkan bahwa model memiliki tingkat kesalahan relatif yang rendah terhadap data aktual.

III. HASIL DAN PEMBAHASAN

A. Analisis Data

Tren jumlah kunjungan wisatawan mancanegara melalui Bandara Kualanamu pada periode 2008 hingga 2024 menunjukkan pola peningkatan yang stabil hingga mencapai puncak pada tahun 2018 hingga 2019. Namun, terjadi penurunan tajam pada tahun 2020 hingga 2021 akibat dampak pandemi COVID-19 yang menyebabkan pembatasan perjalanan internasional. Setelah tahun 2022, jumlah kunjungan kembali menunjukkan pemulihan bertahap dengan pola fluktuatif yang mencerminkan aktivitas pariwisata musiman. Pola ini mengindikasikan adanya potensi pertumbuhan positif sektor pariwisata pasca-pandemi di wilayah Sumatera Utara sebagaimana ditunjukkan pada Gbr 2.



Gbr. 2 Tren jumlah kunjungan wisman melalui Bandara Kualanamu (2008–2024)

B. Pembagian Data

Dataset yang digunakan dalam penelitian ini terdiri dari 204 data observasi bulanan jumlah kunjungan wisman melalui Bandara Kualanamu pada periode Januari 2008 hingga Desember 2024. Data tersebut kemudian dibagi menjadi dua bagian, yaitu data latih (*training set*) dan data uji (*testing set*).

Total data: 204
Training : 163 (2008-01 - 2021-07)
Testing : 41 (2021-08 - 2024-12)

Gbr. 3 Pembagian data *train* dan *test*

Berdasarkan Gbr. 3, sebanyak 163 data (80%) digunakan sebagai data latih yang mencakup periode Januari 2008 hingga Juli 2021, sedangkan 41 data (20%) sisanya digunakan sebagai data uji untuk periode Agustus 2021 hingga Desember 2024. Pembagian ini dilakukan secara berurutan berdasarkan waktu (*time-based splitting*) agar tidak terjadi kebocoran data (*data leakage*) antara data pelatihan dan pengujian.

Pemisahan ini bertujuan agar model XGBoost dapat mempelajari pola historis dari data latih dan kemudian diuji kemampuannya dalam memprediksi periode waktu yang belum pernah dilihat sebelumnya, sehingga evaluasi performa model menjadi lebih objektif.

C. Hasil evaluasi

Evaluasi model dilakukan untuk menilai kinerja model XGBoost yang telah dioptimasi menggunakan Optuna dalam memprediksi jumlah kunjungan wisatawan mancanegara. Berdasarkan hasil pengujian pada data uji, diperoleh nilai MAE sebesar 0,1347, RMSE sebesar 0,1749, dan MAPE sebesar 13,47%, sebagaimana ditunjukkan pada Gbr. 4.

Hasil Evaluasi pada Data Test:	
MAE	: 0.1347327136
RMSE	: 0.1749693750
MAPE(%)	: 13.4732713571

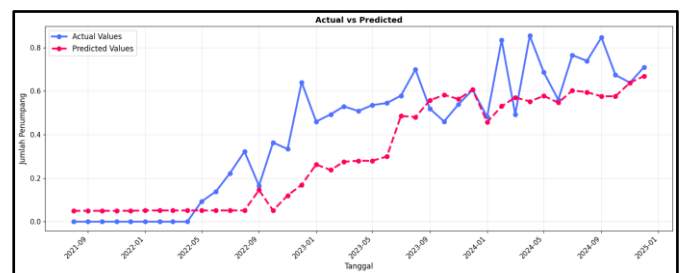
Gbr. 4 Hasil metrik evaluasi

Nilai-nilai tersebut menunjukkan bahwa model memiliki tingkat kesalahan yang relatif rendah dalam skala normalisasi dan mampu memprediksi dengan akurasi yang cukup baik.

MAPE di bawah 15% mengindikasikan performa yang memadai untuk data deret waktu dengan karakteristik musiman seperti kunjungan wisman.

D. Visualisasi Hasil Prediksi

Berdasarkan Gambar 5, perbandingan antara nilai aktual dan hasil prediksi menunjukkan bahwa model mampu menangkap pola umum data dengan tingkat kesesuaian yang cukup baik, meskipun masih terdapat perbedaan pada beberapa periode. Pada fase awal (November 2021 hingga April 2022), nilai aktual berada pada level mendekati nol akibat dampak pembatasan perjalanan pasca-pandemi COVID-19, sedangkan nilai prediksi cenderung berada sedikit di atasnya. Hal ini mengindikasikan adanya bias awal ketika data aktual berada pada fase sangat rendah.



Gbr. 5 Visualisasi *actual vs predicted*

Mulai Mei 2022, nilai aktual menunjukkan tren peningkatan bertahap dengan beberapa lonjakan pada pertengahan hingga akhir 2022. Model mulai mengikuti arah tren ini, meskipun terdapat sedikit *lag* waktu dan prediksi cenderung lebih stabil dibandingkan fluktuasi aktual. Pada periode 2023 hingga akhir 2024, terlihat konvergensi yang lebih baik antara kedua garis. Meskipun masih terdapat perbedaan pada beberapa titik puncak, model secara umum berhasil mengikuti tren dan level nilai aktual dengan deviasi yang lebih kecil.

Model cenderung menghasilkan prediksi yang lebih stabil dan konservatif, terutama pada periode dengan volatilitas tinggi. Karakteristik ini umum terjadi pada model *time series univariate* yang memprioritaskan pola historis rata-rata. Nilai MAPE sebesar 13,47% mengindikasikan bahwa model memiliki tingkat akurasi yang dapat diterima untuk perencanaan operasional, dengan pemahaman bahwa prediksi mungkin kurang tepat pada periode dengan perubahan pola mendadak atau kejadian tidak terduga.

IV. KESIMPULAN

Penelitian ini berhasil mengembangkan model prediksi jumlah kunjungan wisman melalui Bandara Kualanamu menggunakan algoritma XGBoost yang dioptimasi dengan Optuna. Hasil evaluasi menunjukkan nilai MAE sebesar 0,1347, RMSE sebesar 0,1749, dan MAPE sebesar 13,47%, yang mengindikasikan performa yang dapat diterima untuk data *time series* dengan volatilitas tinggi.

Model mampu menangkap tren jangka panjang dengan baik, namun memiliki keterbatasan dalam memprediksi fluktuasi tajam dan lonjakan ekstrem, terutama pada periode 2023 hingga 2024. Model cenderung menghasilkan prediksi konservatif berdasarkan pola rata-rata historis, yang merupakan karakteristik umum dari *time series forecasting univariate* tanpa variabel eksogen. Penerapan Optuna terbukti efektif dalam meningkatkan efisiensi optimasi *hyperparameter* pada model XGBoost.

Secara keseluruhan, model ini dapat digunakan untuk perencanaan strategis jangka menengah dengan mempertimbangkan adanya deviasi pada periode dengan volatilitas tinggi. Untuk pengembangan lebih lanjut, disarankan mengintegrasikan variabel eksogen seperti indeks ekonomi, kebijakan visa, dan kalender *event*, serta mengeksplorasi pendekatan *hybrid* dengan metode *deep learning* guna meningkatkan akurasi prediksi pada pola data yang lebih dinamis.

REFERENSI

- [1] I. Rochayati, U. D. Syafitri, I. M. Sumertajaya, and I. J. of S. and I. A. IJSA, "KAJIAN MODEL PERAMALAN KUNJUNGAN WISATAWAN MANCANEGARA DI BANDARA KUALANAMU MEDAN TANPA DAN DENGAN KOVARIAT," *Indonesian Journal of Statistics and Its Applications*, vol. 3, no. 1, pp. 18–32, Feb. 2019, doi: 10.29244/ijsa.v3i1.171.
- [2] M. Y. Febrian and A. W. Wijayanto, "Prediksi Jumlah Wisatawan Mancanegara Yang Masuk Melalui Bandara Kualanamu Menggunakan Big Data Google Trends," *Seminar Nasional Official Statistics*, vol. 2024, no. 1, pp. 851–862, Nov. 2024, doi: 10.34123/semnasoffstat.v2024i1.2273.
- [3] H. Mukhtar, R. Muhammad, T. Reny Medikawati, and Yoze Rizki, "Peramalan Kedatangan Wisatawan Mancanegara Ke Indonesia Menurut Kebangsaan Perbulannya Menggunakan Metode Multilayer Perceptron," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 2, no. 2, pp. 113–119, Dec. 2021, doi: 10.37859/coscitech.v2i2.3324.
- [4] T. Chen and C. Guestrin, "XGBoost," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA: ACM, Aug. 2016, pp. 785–794. doi: 10.1145/2939672.2939785.
- [5] A. Prayuda and I. Pratama, "PREDIKSI JUMLAH KEDATANGAN WISATAWAN MANCANEGARA DI INDONESIA BERDASARKAN PINTU MASUK KEDATANGAN UDARA," *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 9, no. 2, pp. 232–241, Jul. 2024, doi: 10.36341/rabit.v9i2.4787.
- [6] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, New York, NY, USA: ACM, Jul. 2019, pp. 2623–2631. doi: 10.1145/3292500.3330701.
- [7] Y. Cai, J. Feng, Y. Wang, Y. Ding, Y. Hu, and H. Fang, "The Optuna-LightGBM-XGBoost Model: A Novel Approach for Estimating Carbon Emissions Based on the Electricity-Carbon Nexus," *Applied Sciences*, vol. 14, no. 11, p. 4632, May 2024, doi: 10.3390/app14114632.
- [8] T. O. Hodson, "Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not," *Geosci Model Dev*, vol. 15, no. 14, pp. 5481–5487, Jul. 2022, doi: 10.5194/gmd-15-5481-2022.
- [9] Nurdin, N. Suarna, and W. Prihartono, "ALGORITMA REGRESI LINIER SEDERHANA UNTUK PREDIKSI PENGGUNAAN VOLUME AIR BERDASARKAN JENIS PELANGGAN PDAM," *Jurnal Kecerdasan Buatan dan Teknologi Informasi*, vol. 4, no. 1, pp. 43–52, Jan. 2025, doi: 10.69916/jkbt.v4i1.187.
- [10] A. V. Tatachar, "Comparative Assessment of Regression Models Based On Model Evaluation Metrics," *International Research Journal of Engineering and Technology*, 2021, [Online]. Available: www.irjet.net
- [11] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Comput Sci*, vol. 7, p. e623, Jul. 2021, doi: 10.7717/peerj-cs.623.
- [12] P. Goodwin and R. Lawton, "On the asymmetry of the symmetric MAPE," *Int J Forecast*, vol. 15, no. 4, pp. 405–408, Oct. 1999, doi: 10.1016/S0169-2070(99)00007-2.