Journal of Electrical Engineering and Computer (JEECOM)
Vol. 8, No. 3 (2026), DOI: 10.33650/jeecom.v8i3.16937
p-ISSN: 2715-0410; e-ISSN: 2715-6427

Predicting Daily Online Motorcycle Ride-Hailing Orders Using Online Hours, Day Type, And Weather

Purnomo Hadi Susilo [1*], Mochammad Sihabudin Firdaus Rifani [2], Affan Bachri [3], Mustain [4], Mochammad Sholikhin [5]

[1, 2, 4, 5] Department of Informatics Engineering, Universitas Islam Lamongan, Lamongan, Indonesia
[3] Department of Electrical Engineering, Universitas Islam Lamongan, Lamongan, Indonesia

Article Info

Article history:

Received August 12, 2026
Revised August 17, 2026
Accepted October 15, 2026

Keywords:
demand forecasting; driver-day prediction; multiple linear regression; motorcycle ride-hailing; online transportation; web application

ABSTRACT

Daily order fluctuations complicate workload estimation for online motorcycle ride-hailing drivers. This study develops an interpretable multiple linear regression model to predict daily completed orders using online hours, day type, and weather. The dataset contains 100 driver-day observations collected from eight partner drivers operating in central Tulungagung. To preserve observation order, the first 70 records were used for training and the last 30 for testing. Day type and weather were expressed as binary indicators, yielding the operational equation $\hat{y} = -1.9887 + 1.0036X_1 + 1.4187D_{weekend} + 2.3968D_{clear}$. The model obtained $R^2 = 0.8440$ and MAPE = 15.62% on training data. On the held-out test set, it achieved $R^2 = 0.8341$, MAPE = 13.48%, MAE = 0.989 orders, and RMSE = 1.322 orders. Twenty of 30 predictions (66.7%) were within one order of the actual value, and 26 of 30 (86.7%) were within two orders. However, predictions were positively biased by 0.784 orders on average, with 23 of 30 observations overpredicted and a maximum absolute error of 4.458 orders. The model was integrated into a Flask-MySQL web application for data management and interactive prediction. The results show that a compact linear model can provide explainable driver-level workload estimates, but its reliability remains conditional on the small, locally collected sample. Rolling validation, driver-aware evaluation, richer temporal variables, and count-regression benchmarks are required before operational deployment at scale.

This is an open access article under the CC BY-SA license.

Corresponding Author:
Purnomo Hadi Susilo
Department of Informatics Engineering, Universitas Islam Lamongan, Lamongan, Indonesia
Email: purnomo@unisla.ac.id; Orcid ID : https://orcid.org/0009-0001-8573-7460 :

1. INTRODUCTION

Ride-hailing has changed urban mobility by connecting passengers and drivers through real-time digital platforms. Its operational benefits are accompanied by system-level concerns such as additional vehicle travel, uneven supply, and driver idle time. At the individual-driver level, daily order volumes fluctuate, which makes near-term workload difficult to anticipate from the limited operational information available. This study estimates daily order counts; it does not optimize how long a driver should remain online or prescribe a working schedule. Large-scale evidence also shows that ride-hailing activity varies systematically across time and space, including clear differences between weekdays and weekends. [1], [2]

ISSN: 2715-6427

Demand is shaped by interacting operational and environmental factors. Built environment, prior ridership, seasonality, and weather have measurable relationships with transportation-network-company trips. Weather effects are not uniform: rainfall and wind can alter both passenger demand and the available supply of drivers, while their effects may differ by day type and time of day. These findings support the use of online duration, day type, and weather as a compact set of driver-observable predictors, while also cautioning that the resulting coefficients describe local associations rather than universal causal effects. [3], [4]

Recent demand-forecasting research has emphasized high-capacity models. Fusion convolutional long short-term memory networks capture spatial, temporal, and exogenous dependencies; multi-dimensional machine-learning frameworks combine spatial, temporal, meteorological, and point-of-interest features; and GRU-based models explicitly incorporate rainfall and rest-day factors. Attention-enhanced BiLSTM, Transformer, and multi-task graph models further improve the ability to represent complex urban demand surfaces. [5]-[10] These approaches are valuable when large, high-frequency spatiotemporal datasets are available, but that data requirement is not met by the present sample of 100 driver-day records. Linear regression is therefore not claimed to be universally superior; it was selected as a transparent, low-complexity baseline suited to three predictors, a small sample, coefficient-level interpretation, and direct inspection of error direction. [16]

Explanatory studies show that ride-hailing associations vary with spatial scale, road and transit accessibility, land-use mix, and local walkability. However, four gaps remain relevant to this setting. First, most studies predict aggregate demand at platform, regional, or grid level rather than completed orders per driver-day. Second, they depend on large order traces, coordinates, and contextual features that small driver groups cannot readily obtain. Third, the closest Indonesian comparison evaluates transaction prediction with linear regression and random forest but does not report ordered holdout behavior, directional bias, or driver-level tolerance bands. Fourth, prior work rarely separates predictive accuracy from the runtime performance of a lightweight local web implementation. Consequently, evidence is still lacking on whether a three-input, explainable model can produce credible driver-day estimates and how its errors behave in a small local sample. [11]-[16]

This study therefore evaluates an interpretable multiple linear regression model for predicting daily online motorcycle ride-hailing orders in central Tulungagung. Its contribution is threefold. First, it expresses the original numerical category coding as equivalent binary indicators so that the day and weather coefficients have direct operational meaning. Second, it expands performance reporting beyond R-squared and MAPE to include MAE, RMSE, prediction bias, error tolerances, and residual direction. Third, it connects the fitted model to a Flask-MySQL application while distinguishing predictive accuracy from software latency, which was not measured. The objective is not to compete with platform-scale deep learning or optimize driver schedules, but to determine whether a low-data, low-cost, explainable model can provide credible preliminary workload estimates.

2. METHOD

2.1. Research design and data

The study used a quantitative predictive design. One hundred historical driver-day records were collected from eight partner drivers operating on an online motorcycle ride-hailing platform in central Tulungagung, Indonesia. Data were gathered using Google Forms, WhatsApp messages, and assisted entry when respondents needed help. Each record represented one driver's daily activity. Personal identifiers were not included as model predictors. The relatively small sample motivated a deliberately compact model with three inputs, following the principle that model complexity should be commensurate with the available evidence. [16]

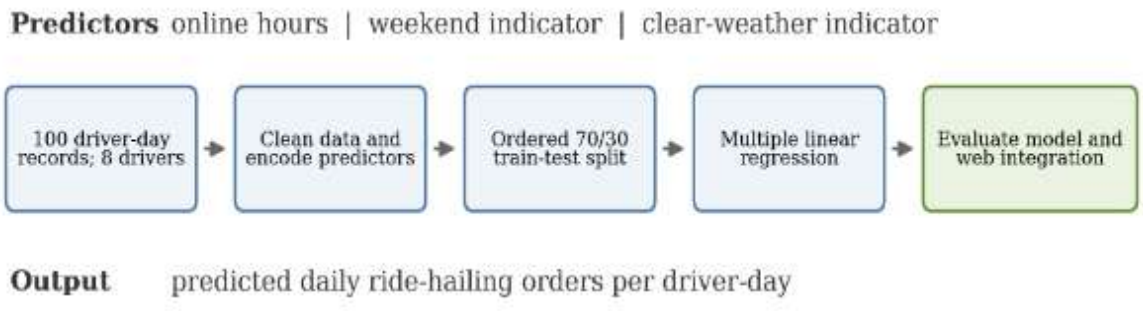


Figure 1. Research and implementation workflow.

2.2. Variables and category coding

The response Y was the observed number of completed online motorcycle ride-hailing orders per driver-day. X1 represented daily online hours. The source dataset encoded weekday as 1 and weekend as 1.5, and rain as 1 and clear weather as 1.5. Although this coding is mathematically valid for two-level categories, the raw coefficients correspond to a full one-unit change and can be misread because the observed contrast is only 0.5. For interpretable reporting, the same fitted predictions were rewritten with Dweekend = 1 for weekends (0 otherwise) and Dclear = 1 for clear weather (0 for rain). This linear reparameterization changes only the intercept and category coefficients; fitted values, residuals, R-squared, and MAPE remain identical. [30]

Table 1. Model variables and operational coding

Symbol	Variable	Operational definition	Coding/unit
Y	Daily orders	Completed ride-hailing orders per driver-day	Count
X_1	Online hours	Total hours the driver was online that day	Hours
Dweekend	Day type	Weekend indicator	1 = weekend; 0 = weekday
Dclear	Weather	Clear-weather indicator	1 = clear; 0 = rain

2.3. Regression model and implementation

Multiple linear regression estimates the conditional mean of a numeric response as an additive function of predictors. For driver-day i, the fitted model was specified as follows: [16]

$$\hat{y}_i = \beta_0 + \beta_1 X_{1i} + \beta_2 Dweekend_{,i} + \beta_3 Dclear_{,i} \quad (1)$$

Parameters were estimated by ordinary least squares using LinearRegression in scikit-learn. The application was implemented in Python with Flask for the web layer and MySQL for storing driver-day data. The interface accepts online hours, day type, and weather, applies the fitted model, and displays the predicted order count together with evaluation information. Scikit-learn provides a consistent implementation of supervised regression and metric functions used in the analysis. [17]

2.4. Train–test split and evaluation

To reproduce the source spreadsheet and preserve record order, observations 1–70 were assigned to training and observations 71–100 to testing. No random shuffle was applied. An ordered holdout is more defensible than a random split when records may contain temporal dependence, because random resampling can leak nearby patterns across sets. Nevertheless, a single ordered split is only one realization of future performance and does not replace rolling-origin or blocked validation. [18], [19]

Performance was assessed with complementary metrics. R^2 compares squared prediction error with the error of predicting the observed mean; MAE reports the average absolute deviation in orders; RMSE gives additional weight to large misses; MAPE gives a scale-free percentage error; and mean prediction bias indicates systematic over- or underprediction. Forecast-evaluation literature recommends reporting multiple measures because each emphasizes a different loss structure. [20]–[23]

$$R^2 = 1 - [\Sigma(y_i - \hat{y}_i)^2 / \Sigma(y_i - \bar{y})^2] \quad (2)$$

$$MAE = (1/n)\Sigma|y_i - \hat{y}_i| \quad (3)$$

$$RMSE = \sqrt{[(1/n)\Sigma(y_i - \hat{y}_i)^2]} \quad (4)$$

$$MAPE = (100\%/n)\Sigma|(y_i - \hat{y}_i)/y_i| \quad (5)$$

MAPE is easy to communicate but becomes unstable when actual values are zero or close to zero and weights low outcomes more heavily. All test outcomes in this study were positive, so MAPE was defined, yet low-order days still exerted disproportionate influence. MAE, RMSE, bias, and tolerance-based accuracy were therefore calculated from the 30 published actual–prediction pairs to provide an operational interpretation. [24], [25]

3. RESULTS AND DISCUSSION

3.1. Fitted equation and coefficient interpretation

Using the original category values, the fitted equation was $\hat{y} = -9.6197 + 1.0036X_1 + 2.8373X_2 + 4.7937X_3$. Re-expressing the two categories as indicators gives the equivalent and more interpretable model:

$$\hat{y} = -1.9887 + 1.0036X_1 + 1.4187Dweekend + 2.3968Dclear \quad (6)$$

ISSN: 2715-6427

Table 2. Regression coefficients and operational effects

Term	Coefficient	Operational interpretation, holding other inputs constant
Intercept	−1.9887	Baseline extrapolation for a rainy weekday at zero online hours
Online hours	+1.0036	Each additional online hour is associated with about 1.004 additional orders
Weekend	+1.4187	A weekend is associated with about 1.419 more orders than a weekday
Clear weather	+2.3968	Clear weather is associated with about 2.397 more orders than rain

Table 2 summarizes the fitted coefficients and their operational interpretations under the binary-indicator coding. Online hours had the largest directly controllable effect: within the observed operating range, one additional hour corresponded to approximately one additional order. The weekend and clear-weather increments were smaller than the coefficients suggested by the original 1/1.5 coding because the actual category contrast was 0.5, not 1. The negative intercept should not be interpreted as a realistic zero-hour demand forecast; it is an extrapolated baseline outside the practical range. If the model is exposed to very low online-hour inputs, the application should constrain predictions to nonnegative counts.

The direction of the weekend effect is consistent with evidence that ride-hailing trips can be higher on weekends, while the weather effect reflects this local sample rather than a universal rule. Other cities have reported weather impacts that depend on rainfall intensity, daytime, trip distance, and whether demand or completed ridership is modeled. Similarly, built-environment studies show spatially heterogeneous associations. Consequently, the coefficients should be used for local prediction, not as causal estimates transferable to another platform or city. [2]–[4], [11]–[14]

3.2. Predictive performance

Table 3. Training and held-out predictive performance

Dataset	n	R^2	MAPE	MAE (orders)	RMSE (orders)	Bias (pred. − actual)
Training	70	0.8440	15.62%	Not reported	Not reported	Not reported
Testing	30	0.8341	13.48%	0.989	1.322	+0.784

Table 3 compares predictive performance on the training set and the ordered held-out test set. The training and test R^2 values differed by only 0.0099, while test MAPE was 2.14 percentage points lower than training MAPE. This absence of a deterioration on the held-out block is encouraging and provides no immediate sign of severe overfitting. However, the conclusion must remain limited: the test set contains only 30 observations, the split was evaluated once, and the records came from eight drivers in one locality. $R^2 = 0.8341$ means that the model reduced squared error substantially relative to the test-set mean benchmark; it does not mean that 83.41% of individual predictions were correct. [23]

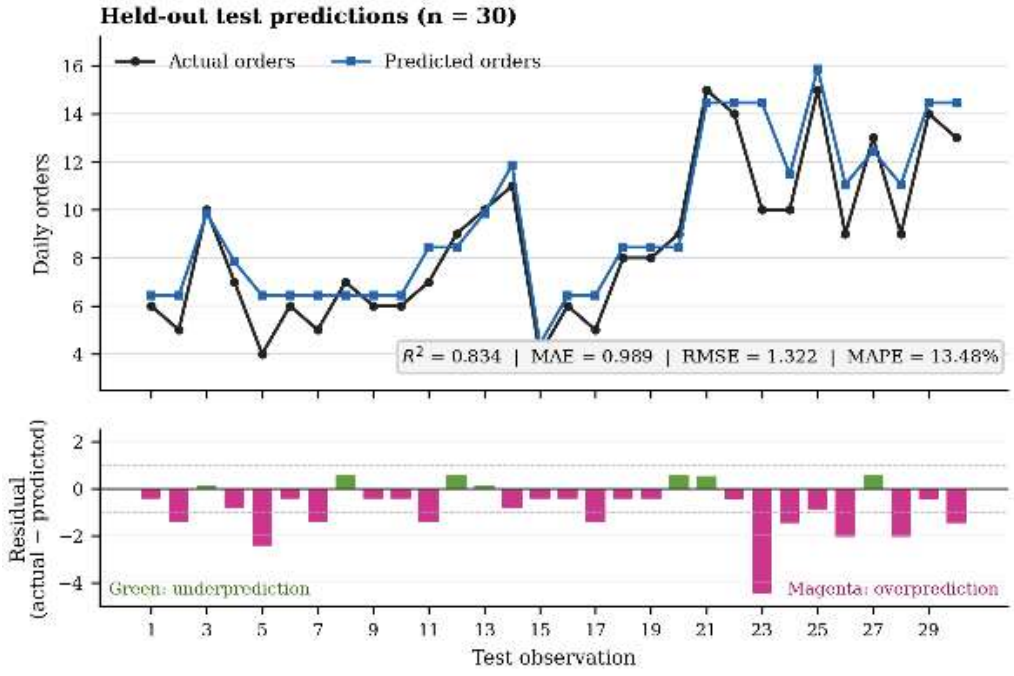


Figure 2. Actual values, predictions, and residuals on the held-out test set.

Table 4. Operational error profile on 30 test observations

Diagnostic	Result	Interpretation
Mean actual / predicted	8.700 / 9.484	Predictions were higher by 0.784 orders on average
Median absolute error	0.563 orders	Half of the predictions missed by no more than 0.563 orders
Within ±1 order	20/30 (66.7%)	Two-thirds were operationally close to the observed count
Within ±2 orders	26/30 (86.7%)	Most predictions were within a two-order error band
Within ±3 orders	29/30 (96.7%)	Only one observation missed by more than three orders
Over / under predictions	23 / 7	Errors were directionally imbalanced toward overprediction
Largest absolute error	4.458 orders	Actual 10; predicted 14.458 at test observation 23
Largest percentage error	60.74%	Actual 4; predicted 6.430 at test observation 5

MAE provides the clearest driver-level interpretation: a typical test prediction missed by about one order. RMSE was 0.333 orders higher than MAE, indicating that a small number of larger misses increased squared loss. Figure 2 and Table 4 show that most errors were modest, but the model overpredicted 23 of 30 cases. The average predicted count was 9.484 against an actual mean of 8.700, producing a positive bias of 0.784 orders. Thus, acceptable aggregate fit coexisted with directional miscalibration.

The error pattern matters when interpreting possible use. If the estimates were used for planning, overprediction could encourage longer online time than realized demand would justify, whereas underprediction could understate a high-demand day. The model itself does not optimize either decision. The largest absolute miss occurred when the observed value was 10 orders and the model predicted 14.458; the largest percentage error occurred on a four-order day. These cases reinforce the need for uncertainty ranges and explicit calibration checks rather than relying only on average accuracy.

3.3. Comparative positioning

Table 5. Position of the proposed model relative to selected ride-hailing studies

Study	Data/model emphasis	Relevance to this study
Ke et al. [5]	Platform-scale spatiotemporal deep learning with exogenous variables	Shows gains from rich historical and spatial signals unavailable in the small driver sample
Guo et al. [6]	Multi-city demand forecasting and vehicle dispatch using multidimensional features	Connects prediction to supply allocation, but operates at platform/region scale
Qi et al. [7]	Multi-factor GRU with rainfall and rest-day information	Supports day/weather inputs and multi-metric evaluation
Zhao et al. [8]	Random-forest feature processing plus attention BiLSTM	Represents a nonlinear benchmark for larger datasets
Bi et al. [9]	Transformer-based high-resolution spatiotemporal prediction	Prioritizes spatial accuracy and training speed on millions of orders
Pramesti and Baihaqi [15]	Indonesian online-motorcycle-taxi transactions; linear regression vs. random forest	Provides the closest local methodological context
This study	100 driver-days; three interpretable inputs; linear regression	Targets low-data driver-level estimation and exposes bias/tolerance behavior

Direct accuracy ranking across Table 5 would be invalid because the studies use different cities, sample sizes, prediction horizons, aggregation levels, and target definitions. The proposed R^2 and MAPE cannot be treated as superior to a GRU or Transformer score reported on a different dataset. Its comparative advantage is instead interpretability and low deployment cost: three manually observable inputs produce a result without a GPU, spatial grid, or long demand history. That advantage is relevant for a small driver-facing prototype, but it comes with a lower ceiling for capturing nonlinear peaks, events, traffic congestion, prices, and spatiotemporal interactions.

3.4. Count-data suitability and validation limitations

Daily orders are nonnegative integer counts, whereas ordinary least squares assumes a continuous conditional mean with unconstrained errors. Linear regression remains useful as a transparent baseline when counts are not extremely sparse and predictions stay in a reasonable range, but it can generate negative values and may not model a variance that grows with the mean. Poisson and negative binomial generalized linear models are natural benchmarks for count responses; the negative binomial form is especially relevant when variability exceeds the Poisson mean–variance restriction. [26]–[28]

The evaluation design introduces three additional risks. First, the records are nested within eight drivers, but a driver-grouped split was not available; the model may partially learn a driver mix that is also present in testing. Second, a single 70/30 ordered split gives no uncertainty interval around performance. Third, omitted predictors—including hour-of-day distribution, public holidays, rainfall intensity, traffic, promotions, surge pricing, location, and prior-day demand—can create residual structure. A stronger study would use rolling-origin evaluation, leave-one-driver-out validation, coefficient confidence intervals, residual diagnostics, and comparisons with a mean/seasonal naïve baseline, Poisson regression, negative binomial regression, random forest, and gradient boosting. [18], [19], [30]

3.5. Web-application performance

The Flask–MySQL implementation completed the intended functional flow: historical data could be stored, the regression model could be fit from the training subset, user inputs could be transformed with the same coding, and a daily order estimate could be returned with evaluation values. This demonstrates functional integration, not measured software performance. Response time, throughput, memory use, database latency, concurrent-user behavior, and model-refresh time were not instrumented in the thesis dataset and therefore are not reported. Claiming that the application is computationally fast would require reproducible timing tests, ideally median and 95th-percentile latency under specified hardware and concurrent loads. Until those tests are performed, the verified performance claim is limited to the predictive metrics and successful end-to-end functionality.

4. CONCLUSION

A multiple linear regression model using online hours, day type, and weather predicted daily online motorcycle ride-hailing orders with held-out R-squared = 0.8341, MAPE = 13.48%, MAE = 0.989 orders, and RMSE = 1.322 orders. Twenty of 30 test predictions (66.7%) were within one order of the observed value, and 26 of 30 (86.7%) were within two orders. In indicator form, one additional online hour was associated with 1.004 additional orders, a weekend with 1.419 more orders than a weekday, and clear weather with 2.397 more orders than rain. Despite the strong aggregate fit, 23 of 30 cases were overpredicted and the mean bias was +0.784 orders, showing that accuracy and directional calibration must be interpreted together. The model should therefore be treated as an explainable prototype for preliminary workload estimation rather than an optimizer or production decision system. Its evidence base is limited to 100 observations from eight drivers in central Tulungagung, and software latency was not measured. Future studies should collect longer multi-driver series, add temporal, traffic, event, location, and fare variables, use rolling and driver-grouped validation, quantify uncertainty, evaluate web latency and concurrent use, and compare linear regression with Poisson, negative-binomial, and nonlinear baselines before wider deployment.

REFERENCES

[1] A. Henao and W. E. Marshall, “The impact of ride-hailing on vehicle miles traveled,” Transportation, vol. 46, pp. 2173–2194, 2019, doi: 10.1007/s11116-018-9923-2.

[2] A. Du, H. A. Rakha, and J. Breuer, “An in-depth spatiotemporal analysis of ride-hailing travel: The Chicago case study,” Case Stud. Transp. Policy, vol. 10, no. 1, pp. 118–129, 2022, doi: 10.1016/j.cstp.2021.11.010.

[3] M. S. Hasnine, J. Hawkins, and K. N. Habib, “Effects of built environment and weather on demands for transportation network company trips,” Transp. Res. Part A Policy Pract., vol. 150, pp. 171–185, 2021, doi: 10.1016/j.tra.2021.06.011.

[4] S. Liu, H. Jiang, and Z. Chen, “Quantifying the impact of weather on ride-hailing ridership: Evidence from Haikou, China,” Travel Behav. Soc., vol. 24, pp. 257–269, 2021, doi: 10.1016/j.tbs.2021.04.002.

[5] J. Ke, H. Zheng, H. Yang, and X. Chen, “Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach,” Transp. Res. Part C Emerg. Technol., vol. 85, pp. 591–608, 2017, doi: 10.1016/j.trc.2017.10.016.

[6] Y. Guo, Y. Zhang, Y. Boulaksil, and N. Tian, “Multi-dimensional spatiotemporal demand forecasting and service vehicle dispatching for online car-hailing platforms,” Int. J. Prod. Res., vol. 60, no. 6, pp. 1832–1853, 2022, doi: 10.1080/00207543.2021.1871675.

[7] Q. Qi, R. Cheng, and H. Ge, “Short-term travel demand prediction of online ride-hailing based on multi-factor GRU model,” Sustainability, vol. 14, no. 7, Art. no. 4083, 2022, doi: 10.3390/su14074083.

[8] X. Zhao, K. Sun, S. Gong, and X. Wu, “RF-BiLSTM neural network incorporating attention mechanism for online ride-hailing demand forecasting,” Symmetry, vol. 15, no. 3, Art. no. 670, 2023, doi: 10.3390/sym15030670.

[9] S. Bi, C. Yuan, S. Liu, L. Wang, and L. Zhang, "Spatiotemporal prediction of urban online car-hailing travel demand based on Transformer network," Sustainability, vol. 14, no. 20, Art. no. 13568, 2022, doi: 10.3390/su142013568.

[10] Y. Guo, Y. Chen, and Y. Zhang, "Enhancing demand prediction: A multi-task learning approach for taxis and TNCs," Sustainability, vol. 16, no. 5, Art. no. 2065, 2024, doi: 10.3390/su16052065.

[11] Z. Wang, X. Gong, Y. Zhang, S. Liu, and N. Chen, "Multi-scale geographically weighted elasticity regression model to explore the elastic effects of the built environment on ride-hailing ridership," Sustainability, vol. 15, no. 6, Art. no. 4966, 2023, doi: 10.3390/su15064966.

[12] G. Zhao, Z. Li, Y. Shang, and M. Yang, "How does the urban built environment affect online car-hailing ridership intensity among different scales?" Int. J. Environ. Res. Public Health, vol. 19, no. 9, Art. no. 5325, 2022, doi: 10.3390/ijerph19095325.

[13] F. Zhao, J. Ma, C. Yin, W. Tang, X. Wang, and J. Yin, "Spatiotemporal heterogeneous effects of built environment and taxi demand on ride-hailing ridership," Appl. Sci., vol. 14, no. 1, Art. no. 142, 2024, doi: 10.3390/app14010142.

[14] R. Si and Y. Lin, "Exploring the spatiotemporal associations between ride-hailing demand, visual walkability, and the built environment: Evidence from Chengdu, China," Sustainability, vol. 17, no. 12, Art. no. 5441, 2025, doi: 10.3390/su17125441.

[15] D. F. Pramesti and I. Baihaqi, "Perbandingan prediksi jumlah transaksi ojek online menggunakan regresi linier dan random forest," Generation J., vol. 7, no. 3, pp. 21–30, 2023, doi: 10.29407/gj.v7i3.20676.

[16] G. James, D. Witten, T. Hastie, and R. Tibshirani, An Introduction to Statistical Learning: With Applications in R, 2nd ed. New York, NY, USA: Springer, 2021, doi: 10.1007/978-1-0716-1418-1.

[17] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," J. Mach. Learn. Res., vol. 12, pp. 2825–2830, 2011.

[18] C. Bergmeir and J. M. Benítez, "On the use of cross-validation for time series predictor evaluation," Inf. Sci., vol. 191, pp. 192–213, 2012, doi: 10.1016/j.ins.2011.12.028.

[19] D. R. Roberts et al., "Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure," Ecography, vol. 40, no. 8, pp. 913–929, 2017, doi: 10.1111/ecog.02881.

[20] F. Petropoulos et al., "Forecasting: Theory and practice," Int. J. Forecast., vol. 38, no. 3, pp. 705–871, 2022, doi: 10.1016/j.ijforecast.2021.11.001.

[21] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," Int. J. Forecast., vol. 22, no. 4, pp. 679–688, 2006, doi: 10.1016/j.ijforecast.2006.03.001.

[22] D. Koutsandreas, E. Spiliotis, F. Petropoulos, and V. Assimakopoulos, "On the selection of forecasting accuracy measures," J. Oper. Res. Soc., vol. 73, no. 5, pp. 937–954, 2022, doi: 10.1080/01605682.2021.1892464.

[23] 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, Art. no. e623, 2021, doi: 10.7717/peerj-cs.623.

[24] A. de Myttenaere, B. Golden, B. Le Grand, and F. Rossi, "Mean absolute percentage error for regression models," Neurocomputing, vol. 192, pp. 38–48, 2016, doi: 10.1016/j.neucom.2015.12.114.

[25] S. Kim and H. Kim, "A new metric of absolute percentage error for intermittent demand forecasts," Int. J. Forecast., vol. 32, no. 3, pp. 669–679, 2016, doi: 10.1016/j.ijforecast.2015.12.003.

[26] R. B. O'Hara and D. J. Kotze, "Do not log-transform count data," Methods Ecol. Evol., vol. 1, no. 2, pp. 118–122, 2010, doi: 10.1111/j.2041-210X.2010.00021.x.

[27] D. Lord and F. Mannering, "The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives," Transp. Res. Part A Policy Pract., vol. 44, no. 5, pp. 291–305, 2010, doi: 10.1016/j.tra.2010.02.001.

[28] J. A. Nelder and R. W. M. Wedderburn, "Generalized linear models," J. R. Stat. Soc. Ser. A, vol. 135, no. 3, pp. 370–384, 1972, doi: 10.2307/2344614.

[29] T. Gneiting and J. Resin, "Regression diagnostics meets forecast evaluation: Conditional calibration, reliability diagrams, and coefficient of determination," Electron. J. Stat., vol. 17, no. 2, pp. 3226–3286, 2023, doi: 10.1214/23-EJS2180.

[30] M. Kuhn and K. Johnson, Feature Engineering and Selection: A Practical Approach for Predictive Models. Boca Raton, FL, USA: Chapman & Hall/CRC, 2019, doi: 10.1201/9781315108230.