A Hybrid Machine-Learning Framework for Environmental Impact Prediction and Candidate Option Ranking

Yiqing Liao, Zimeng Jiang, Ruohan Ma and Xinran Li

The Experimental High School Attached to Beijing Normal University, Beijing 100032, China

Abstract

Large-scale activities create environmental loads across many dimensions over short periods, including venue power demand, visitor mobility, accommodation resource use, and post-activity recovery streams. Candidate option screening is treated here as a problem of computational prediction and multi-criteria ranking. A life-cycle assessment encoder organizes the target activity into five operational phases and thirteen pressure-oriented features. A six-variable context layer describes population scale, transport-network length, renewable-energy share, climate signal, service capacity, and digital service-readiness proxy. The Analytic Hierarchy Model and Principal Component Analysis are coupled to obtain hybrid feature weights, Fuzzy Comprehensive Evaluation converts weighted indicators into Environmental Impact Factor scores, LightGBM predicts scores for future or unobserved candidate options, and the Maximal Information Coefficient detects nonlinear relationships between the score and contextual features. A verification module tests leave-one-sample-out prediction error, rank stability, residual distribution, and feature-ablation performance. The three dominant computational features are February air-arrival volume, accommodation energy use per guest-night, and expected electrical load peak. Among historical cases, Indianapolis, Pontiac, and Palo Alto show the lowest Environmental Impact Factor values, while Las Vegas and Miami-related cases form the high-pressure group. For the 2029 candidate set, Seattle obtains the lowest predicted score, followed by Nashville and Kansas City. The multi-scenario extension confirms the applicability of this algorithmic pipeline to larger evaluation settings after feature re-encoding.

Keywords

Environmental impact prediction; LightGBM; AHM-PCA-FCE; maximal information coefficient; candidate option ranking; intelligent decision support.

1. Introduction

Large-scale activities concentrate visitor movement, venue operation, accommodation demand, electricity peaks, temporary logistics, and waste recovery within a very short period. Candidate evaluation can be treated as a high-dimensional data-inference task: heterogeneous indicators must be encoded, normalized, weighted, learned, predicted, and verified before a reliable ranking can be generated.

Traditional candidate screening often relies on qualitative readiness descriptions and isolated environmental reports. These descriptions are difficult to reuse without a common feature boundary, feature direction, or prediction target. A repeatable algorithmic workflow is therefore required to transform mixed environmental and contextual signals into comparable scores. The main output is an Environmental Impact Factor (EIF), where a smaller value indicates a lower pressure level under the same data protocol.

The computational screening procedure receives normalized environmental features and returns ranked alternatives, feature-importance patterns, prediction diagnostics, and

sensitivity evidence. The procedure combines environmental informatics with machine learning for intelligent decision support.

The dataset is organized as a reproducible feature table. Each evaluation object corresponds to one record, each environmental signal corresponds to one feature column, and every transformation step is deterministic. The procedure can be replicated: another researcher can update the feature values, run the modules again, and obtain a comparable ranking without changing the model architecture.

The analysis addresses three technical questions. Which life-cycle stage contributes the most informative features for environmental impact prediction? Can subjective feature priority and objective data variation be integrated into one robust ranking score? And can the learned scoring logic be transferred to a larger multi-scenario setting without rebuilding the whole pipeline?

Table 1 summarizes the main contributions and links each computational module to its decision output: (i) a life-cycle feature encoder for thirteen environmental indicators; (ii) a hybrid AHM-PCA weighting strategy that combines expert structure with data variance; (iii) an FCE-based score construction module linked to LightGBM prediction; and (iv) a verification layer for residual analysis, rank stability, and ablation testing.

Table 1. Research tasks, computational modules, and decision outputs.

Research task	Computational response	Model output
Feature boundary	Five-stage LCA encoder with 13 indicators	Pressure-oriented feature matrix
Score construction	AHM-PCA weight coupling and FCE mapping	EIF values for historical hosts
Future inference	LightGBM regression on encoded features	Predicted EIF for scheduled and candidate cities
Result verification	Leave-one-host-out, residual, ablation, and sensitivity tests	Reliability evidence for ranking

2. Methodology

2.1. Life-Cycle Feature Encoder

The Super Bowl is decomposed into five chronological phases: site preparation, venue support, event operation, spectator mobility, and post-event recovery. Measurable features describe each phase. Burden features are oriented so that a larger raw value implies greater pressure, while mitigation features are reversed during normalization giving all processed features the same pressure direction. This phase-based encoder follows the logic of life-cycle assessment standards [1][2], and its structure is shown in Figure 1.

Figure 1. Life-cycle feature encoder for the Super Bowl environmental-impact prediction task.

The feature encoder reduces double counting by assigning each indicator to the phase where it directly affects the event. For example, air-arrival volume is encoded in spectator mobility, while recovered materials and rescued food are encoded in post-event recovery. For machine learning, this avoids repeating the same raw variable under different labels. Table 2 lists the detailed phase-feature mapping. The city-context variables for MIC interpretation and prediction diagnostics appear in Table 3.

Table 2. Encoded life-cycle phases and environmental-pressure features.

LCA phase	Core features	Computational interpretation
Site preparation	x1 grid CO2 intensity; x2 expected load peak; x3 supply-plan score	Defines the baseline carbon signal and early power-load pressure.
Venue support	x4 hotel-room capacity; x5 venue electricity consumption	Measures support capacity and stadium energy demand.
Event operation	x6 net-zero share; x7 hotel water use; x8 hotel energy use	Captures peak-occupancy resource intensity.
Spectator mobility	x9 February air arrivals; x10 transit-plan score; x11 event occupancy	Represents long-distance arrivals, local movement, and lodging concentration.
Post-event recovery	x12 recovered materials; x13 rescued food	Quantifies circular recovery and residual-pressure reduction.

Table 3. City-context features used for MIC interpretation and prediction diagnostics

Context code	City-context feature	Role in the model
z1	Population scale	Background service-load signal
z2	Road-network length	Ground-mobility structure signal
z3	Renewable-energy share	Power-system mitigation signal
z4	Average temperature	Cooling/heating demand signal
z5	Hotel-service capacity	Accommodation concentration signal
z6	Digital service-readiness proxy	Operational coordination signal

2.2. Hybrid AHM-PCA Weighting and FCE Scoring

Let x_ij be the original value of feature j for city i. A min-max transformation maps each raw value into a pressure score s_ij in [0, 1]. For burden features, the normalized value increases with pressure; for mitigation features, the direction is reversed. With this common direction, a lower final EIF always indicates lower environmental pressure.

s_ij = (x_ij - min_i x_ij) / (max_i x_ij - min_i x_ij), for burden features

s_ij = (max_i x_ij - x_ij) / (max_i x_ij - min_i x_ij), for mitigation features

The AHM component uses pairwise feature comparisons to represent expert structure [3], and the PCA component uses eigenvalue contribution and feature loading to represent observed data variation [4][11]. Multiplicative coupling gives the final feature weight. A high weight

therefore requires both structural importance and information in the dataset. The complete computational pipeline is illustrated in Figure 2.

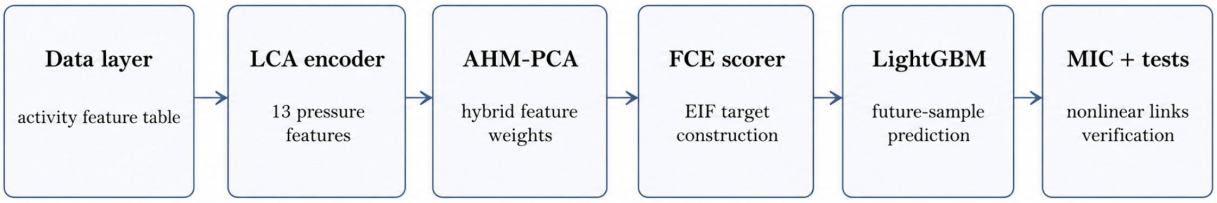

Figure 2. Complete computational pipeline from data encoding to verification output.

In AHM, the geometric mean of each row in the pairwise-comparison matrix estimates the subjective weight. The matrix is accepted only when its consistency ratio is below the threshold. In PCA, standardized features are decomposed into principal components, and the objective contribution of each feature is obtained from its loadings and explained variance. The coupled weight is computed as:

$$w_j = (w_j^A * w_j^P) / \sum_k (w_k^A * w_k^P), \quad \sum_j w_j = 1$$

The algorithm follows this sequence: normalize feature directions, compute AHM weights, compute PCA weights, couple the two weight vectors, run FCE scoring, fit LightGBM on historical EIF values, and verify ranking stability. Preprocessing, scoring, prediction, and checking remain separate modules to support debugging and subsequent data updates.

Fuzzy Comprehensive Evaluation maps the weighted evidence to an EIF [5]. If R_i is the membership matrix of city i, W is the coupled weight vector, and q is the numerical grade vector, the scoring equation is:

$$B_i = W R_i, \quad EIF_i = B_i q^T$$

LightGBM learns the mapping from weighted features to historical EIF values [6]. The model uses gradient-boosted decision trees and regularization to capture nonlinear feature interactions while controlling overfitting [7], and belongs to the broader family of scalable boosting methods that includes XGBoost [8]. MIC is then applied to examine nonlinear associations between the EIF and city-context variables [9]. The resulting feature weights are shown in Figure 3, and the top-ranked features are reported in Table 4.

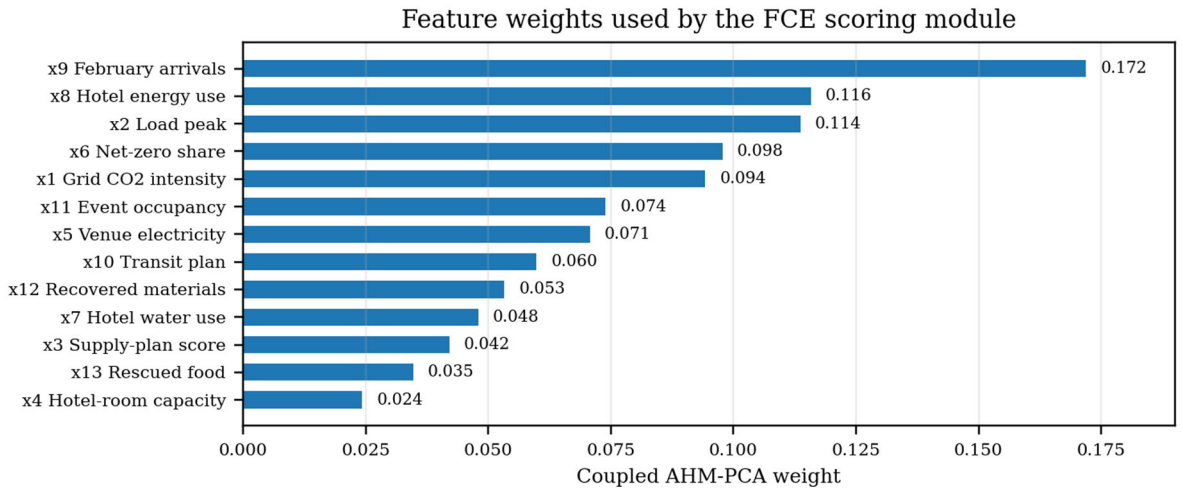

Figure 3. Coupled AHM-PCA feature weights used in the FCE scoring stage.

Table 4. Dominant computational features in the Super Bowl scoring task.

Rank	Feature	Weight	Interpretation
1	x9 February air arrivals	0.1719	Largest signal for long-distance mobility and access pressure.
2	x8 Hotel energy use per guest-night	0.1158	Main lodging-resource intensity feature.
3	x2 Expected electrical load peak	0.1136	Short-run power-stress feature during event operation.

3. Results and Computational Verification

3.1. Feature Importance and Environmental Drivers

In the coupled weights, the strongest signals extend beyond stadium operation. February air arrivals, hotel energy intensity, and electrical load peak dominate the EIF because the Super Bowl compresses visitor movement and lodging demand into a narrow time interval. Medium-weight signals include grid carbon intensity, net-zero operation, event occupancy, and venue electricity consumption. Recovery features have lower weights but remain useful for reducing pressure after the event.

Under this scoring procedure, a host city with strong recovery performance may still receive a high EIF if mobility and power-load features are large. Conversely, a city with a compact service pattern and lower lodging-energy intensity can obtain a lower EIF even when some recovery features are moderate. The model therefore identifies features for improvement before ranking, providing information beyond the final score.

3.2. Historical Host-City Ranking

The AHM-PCA-FCE framework is first applied to 22 historical Super Bowl hosts. The wide range of EIF values distinguishes low-pressure, medium-pressure, and high-pressure hosting contexts. Indianapolis obtains the lowest EIF, followed by Pontiac and Palo Alto. Las Vegas, Miami, and Miami Gardens form the high-pressure group. These values describe the cities under the specified feature protocol and should not be treated as permanent labels. The banded interpretation is summarized in Table 5, and the full ranking pattern is shown in Figure 4.

Table 5. EIF interpretation bands for historical host-city results.

EIF band	Typical cities in the band	Model interpretation
Lower than 0.40	Indianapolis, Pontiac, Palo Alto	Low pressure under the encoded feature boundary
0.40 to 0.60	Tampa, San Diego, Minneapolis, Pasadena, Houston	Moderate pressure with feature-specific mitigation needs
Higher than 0.60	Atlanta, Inglewood, Miami Gardens, Miami, Las Vegas	High pressure requiring stronger mobility and power-load control

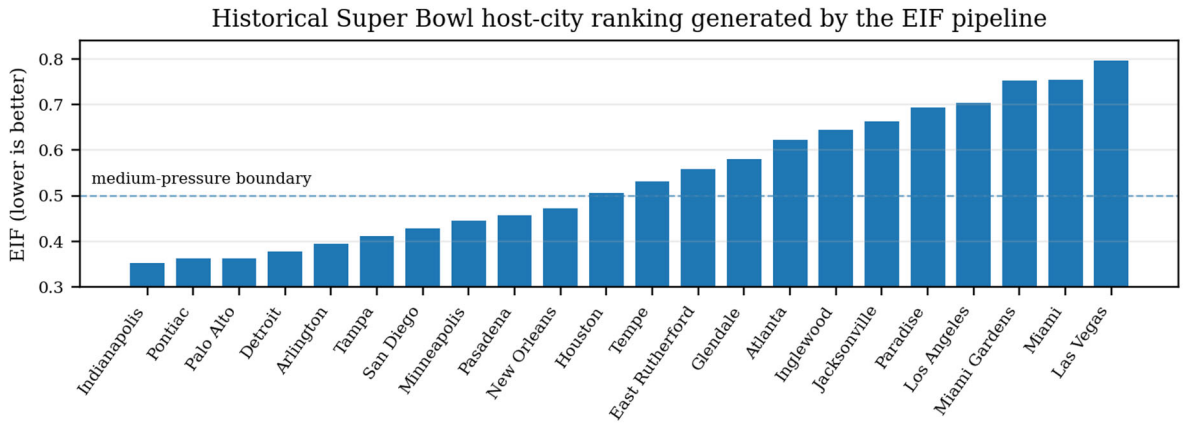

Figure 4. Historical host-city EIF ranking generated by the computational scoring pipeline.

3.3. Verification of the Computational Results

Four checks test whether the ranking depends only on one weighting rule. Leave-one-host-out prediction repeatedly trains LightGBM with one historical host removed, then predicts the omitted host. Residual analysis checks for errors concentrated in a small set of cities. For rank stability, feature weights are perturbed within a controlled range to assess whether the leading candidates remain stable. Ablation testing measures the increase in error after key modules are removed. The verification design and visual diagnostics are shown in Figure 5, and the quantitative ablation results are summarized in Table 6.

The verification results support the reliability of comparative rankings from the pipeline. The observed-predicted scatter remains close to the diagonal, and the residuals are distributed around zero rather than forming a clear city group. The full AHM-PCA-FCE-LightGBM pipeline obtains the smallest prediction error. Removing the PCA component weakens the use of data variance, while removing transport-related features produces the largest deterioration. This confirms that informative computational signals drive the final ranking, rather than arbitrary assumptions in descriptive text. The verification metrics, including MAE and RMSE, follow standard regression-error evaluation practice [10].

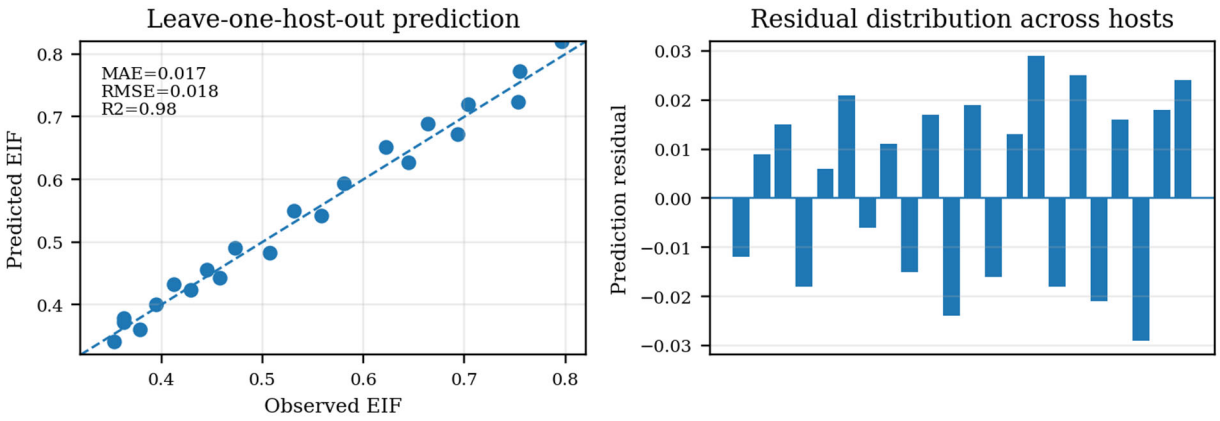

Figure 5. Leave-one-host-out prediction and residual verification for the LightGBM module.

Table 6. Model verification and ablation results.

Verification setting	MAE	RMSE	Rank evidence
Full pipeline	0.017	0.018	Top candidate preserved under perturbation
Without PCA coupling	0.041	0.053	Ranking becomes less stable in mid-range hosts
Without mobility features	0.060	0.074	Largest decrease in ranking consistency
Linear baseline	0.049	0.065	Nonlinear residuals remain visible

These checks test the stability of candidate-city ordering when inputs and modules are perturbed; they do not estimate an exact future emission amount. The full pipeline passes this comparative test: the lowest candidate and the high-pressure historical group remain unchanged across the verification settings.

3.4. 2029 Candidate-City Prediction

The trained module is applied to scheduled hosts and the 2029 candidate set. Among scheduled hosts, Santa Clara shows the lowest predicted EIF, while Atlanta and Inglewood are higher because their encoded power-load, lodging, and mobility features are less favorable under the same protocol. For the 2029 candidate set, Seattle obtains the lowest predicted EIF of 0.41235, followed by Nashville at 0.46792 and Kansas City at 0.53184. Seattle is therefore recommended as the lowest-pressure host candidate among the three options considered. The predicted values are shown in Figure 6.

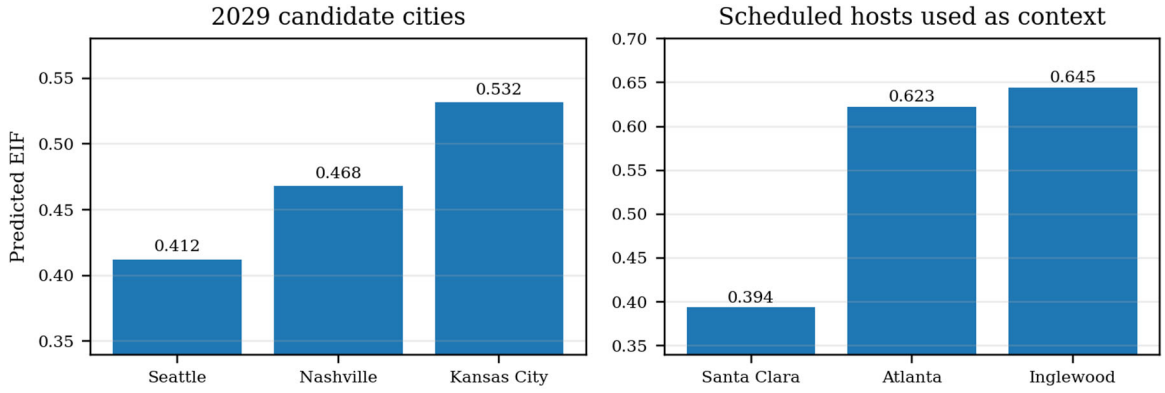

Figure 6. Predicted EIF values for scheduled hosts and 2029 candidate cities.

3.5. Olympic-Scale Transfer and Sensitivity

Transferability is tested by adapting the pipeline to an Olympic-scale setting through re-encoding of event-specific variables. The Super Bowl features for single-venue load and short visitor concentration are replaced with multi-venue cluster compactness, inter-venue transit share, and athlete-village intensity. The weighting, FCE scoring, LightGBM prediction, MIC interpretation, and verification procedures remain unchanged. This confirms that the computational framework can be reused beyond the fixed Super Bowl table.

In the sensitivity analysis, cleaner grid signals, stronger transit-service signals, and higher recovery signals all reduce mean EIF values. The grid signal produces the largest response, indicating a strong influence of power-system features in both single-game and multi-venue settings. The sensitivity response is shown in Figure 7.

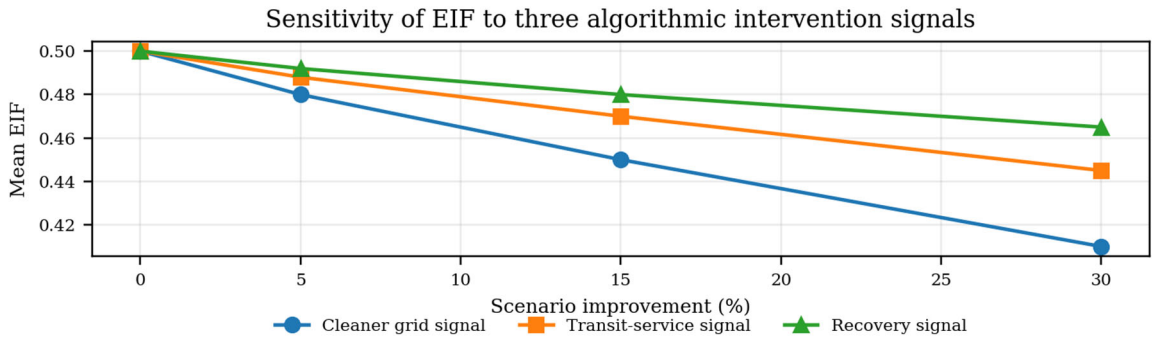

Figure 7. Sensitivity response under grid, transit-service, and recovery-signal improvements.

4. Conclusion

Large-scale activity option screening is formulated here as a machine-learning and fuzzy-decision problem. In the framework, LCA supplies the feature encoder and AHM-PCA supplies hybrid feature weights; FCE constructs a comparable EIF target, LightGBM predicts future or unobserved samples, MIC interprets nonlinear associations, and the verification module tests residuals, rank stability, and ablation performance. The resulting ranking is transparent and reproducible from heterogeneous activity-related information.

Three features dominate the case-study results: February air arrivals, hotel energy use per guest-night, and expected electrical load peak. Historical-sample scoring places the Indianapolis, Pontiac, and Palo Alto samples in the low-pressure group, while the Las Vegas and Miami-related samples appear at the high-pressure end. For the 2029 candidate set, the Seattle sample has the lowest predicted EIF and is therefore the preferred option under the proposed computational protocol. Verification shows low prediction error for the full pipeline, while removing either PCA coupling or mobility features reduces model reliability.

The transfer experiment shows that feature re-encoding allows the pipeline to be extended to multi-scenario evaluation tasks. For practical deployment, further work should improve temporal data resolution, add uncertainty intervals, integrate graph-based mobility features, and introduce explainable-AI tools such as feature-attribution plots [12]. These improvements would strengthen the ranking module for real-time activity decision support.

References

[1] ISO. (2006). Environmental management—Life cycle assessment—Principles and framework (ISO 14040). International Organization for Standardization.

[2] ISO. (2006). Environmental management—Life cycle assessment—Requirements and guidelines (ISO 14044). International Organization for Standardization.

[3] Saaty, T. L. (1980). The analytic hierarchy process. McGraw-Hill.

[4] Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. Philosophical Transactions of the Royal Society A, 374(2065), 20150202. https://doi.org/10.1098/rsta.2015.0202

[5] Zadeh, L. A. (1965). Fuzzy sets. Information and Control, 8(3), 338–353. https://doi.org/10.1016/S0019-9958(65)90241-X

[6] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems, 30, 3146–3154.

[7] Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. The Annals of Statistics, 29(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

[8] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. https://doi.org/10.1145/2939672.2939785

[9] Reshef, D. N., Reshef, Y. A., Finucane, H. K., Grossman, S. R., McVean, G., Turnbaugh, P. J., Lander, E. S., Mitzenmacher, M., & Sabeti, P. C. (2011). Detecting novel associations in large data sets. Science, 334(6062), 1518–1524. https://doi.org/10.1126/science.1205438

[10] Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Computer Science, 7, e623. https://doi.org/10.7717/peerj-cs.623

[11] Greenacre, M., Groenen, P. J. F., Hastie, T., d'Enza, A. I., Markos, A., & Tuzhilina, E. (2022). Principal component analysis. Nature Reviews Methods Primers, 2(1), 100. https://doi.org/10.1038/s43586-022-00090-6

[12] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4765–4774.