Explainable Shuttle Delay Prediction for Smart Campus Mobility

Abin Raj Devarajan*, Haseeb Abdul Mohammed*, Khaled Abdalgader†
*School of Engineering, American University of Ras Al Khaimah
Ras Al Khaimah, UAE
†Department of Computer Science and Engineering, American University of Ras Al Khaimah
Ras Al Khaimah, UAE
Emails: abinraj.devarajan@aurak.ac.ae
2023006075@aurak.ac.ae
khaled.balawafi@aurak.ac.ae

***Abstract*—Reliable campus micro-transit is a key enabler for smart-city ecosystems in the Gulf Cooperation Council (GCC), yet shuttle operations exhibit strong stochasticity due to traffic, weather, and event-driven demand. This paper presents AURAK-Predictor, a proof-of-concept framework for predicting shuttle delay risk and estimating delay magnitude for operational decision support at the American University of Ras Al Khaimah (AURAK). To address small-sample limitations, we construct a 100,005-row *augmented semi-empirical* dataset by scraping published Ras Al Khaimah Transport Authority (RAKTA) timetable entries and replicating schedules across calendar days, while bootstrapping contextual variables (traffic, weather, special events, temperature) and delay-derived labels from an existing campus dataset to preserve realistic marginal distributions. We benchmark an interpretable Logistic Regression model against a high-performance XGBoost ensemble using imbalance-robust metrics. On a stratified 80/20 split, Logistic Regression achieves 0.847 accuracy with 0.807 macro-F1 and 0.864 balanced accuracy, while XGBoost achieves 0.942 accuracy with 0.921 macro-F1 and 0.953 balanced accuracy. For delay regression, an XGBoost regressor reduces mean absolute error from 5.01 minutes (baseline) to 1.31 minutes. Explainability is provided via Logistic Regression coefficient magnitudes and SHAP attributions for XGBoost.**

***Index Terms*—Intelligent Transportation Systems, Machine Learning, Explainable AI, Smart Campus, Smart Cities.**

I. Introduction

The rapid urbanization of the Northern Emirates, specifically Ras Al Khaimah (RAK), has increased the complexity of localized transit networks. Within a university ecosystem, shuttle services ensure timely movement of the academic community. However, variables such as localized sandstorms, traffic bottlenecks, and peak-hour surges during examinations introduce significant stochasticity into shuttle schedules. This phenomenon, termed the “Shuttle Reliability Problem,” creates a disconnect between administrative expectations and student mobility needs [2], [4].

In the GCC region, traffic congestion and extreme weather patterns are primary drivers of transit inefficiency [7]. To address this, we propose AURAK-Predictor, a simulation-based proof-of-concept that transforms transit logs into actionable insights. Unlike purely empirical studies that rely on live telematics, this framework explores the feasibility of predictive modeling using a semi-empirical environment while explicitly stress-testing scalability to 100k+ trips.

II. Related Work

A. Foundations of Transit Prediction

Early efforts in transit modeling focused on historical averages and simple regression. However, Altinkaya and Zontul [2] emphasize that stochastic urban environments require more robust models. Deep learning approaches (e.g., LSTM) can achieve strong performance [3], but may reduce transparency for operational stakeholders. Recent reviews summarize the move toward predictive, connected, and sustainable mobility systems [13].

B. Environmental and Regional Context

Transit in the GCC is uniquely affected by arid climatic conditions. Habib et al. [6] documented the impact of dust storms on transportation safety and speed. UAE smart-city programs increasingly emphasize data-driven mobility as a core pillar of sustainability and service quality [11].

C. Interpretability and Explainable AI (XAI)

As ITS pipelines become more complex, explainability becomes a practical requirement for adoption. Systematic reviews emphasize that operational utility depends on interpretable decision logic in addition to predictive accuracy [12]. We report both global linear weights (LR coefficients) and SHAP attributions for XGBoost [19].

III. Simulation and Methodology

A. Augmented Semi-Empirical Dataset (Scalability Stress-Test)

To avoid ambiguity between empirical observations and scaled data, we explicitly label the 100,000+ dataset as an *Augmented Semi-Empirical Dataset*. Static timetable entries are scraped from the RAKTA public timetable page [10] and replicated across calendar days to reach the target size. Contextual variables (traffic condition, weather, special events, temperature) and delay-derived labels are then *bootstrapped* by sampling with replacement from the original AURAK shuttle dataset to preserve realistic marginal distributions. This augmentation is used to (i) stress-test pipeline runtime and memory behavior and (ii) benchmark models under controlled distributional assumptions; it does **not** replace real telematics ground truth.

Label definition: We define *Late* if the observed delay exceeds 5 minutes, i.e., *delay_minutes* > 5; otherwise the trip is labeled *On-Time*.

B. Operational Capture of "Traffic Condition" (Deployment Realism)

To avoid a purely theoretical "traffic condition" feature, AURAK-Predictor assumes traffic is derived from widely used routing APIs. In a real deployment, the system queries a travel-time endpoint (e.g., Google Distance Matrix API) and computes a congestion ratio:

$$\rho = \frac{T_{\text{traffic}}}{T_{\text{free-flow}}} \tag{1}$$

where T_{traffic} is current travel time and $T_{\text{free-flow}}$ is baseline travel time under free-flow. ρ is discretized into categories (e.g., *Low*, *Medium*, *High*) using configurable thresholds derived from travel-time telemetry [22].

TABLE I
Imbalance-robust classification results on the augmented dataset (*Late* is the positive class; test prevalence ≈ 0.780).

Model	Acc.	Bal. Acc.	Macro-F1	$\text{F1}_{\text{On-Time}}$	F1_{Late}
Majority Baseline (predict Late)	0.780	0.500	0.438	0.000	0.877
Logistic Regression	0.847	0.864	0.807	0.720	0.895
XGBoost	0.942	0.953	0.921	0.880	0.961

Note: Deduplication + group split were applied to reduce leakage in the augmented dataset.

IV. Machine Learning Benchmarking

A. Experimental Setup and Metrics (Imbalance-Robust)

We use a stratified 80/20 train-test split. Because the dataset is class-imbalanced (majority *Late* in our augmented labels), accuracy alone can be misleading. We therefore report **Macro-F1** and **Balanced Accuracy** in addition to accuracy.

B. Models Evaluated: Interpretability vs. High Performance

We benchmark two complementary models:

- **Logistic Regression (LR):** Interpretable baseline for decision support [8].
- **XGBoost (XGB):** High-performance ensemble for capturing non-linear interactions [9].

C. Hyperparameter Tuning

We tune hyperparameters on the training set using stratified 5-fold cross-validation, selecting configurations by **Macro-F1**. For LR, we tune regularization strength C and apply class weighting. For XGBoost, we tune *max_depth*, *learning_rate*, *n_estimators*, *subsample*, and *colsample_bytree*, with early stopping on a validation fold.

D. Classification Results

Table I summarizes benchmarking results on the augmented dataset.

E. Delay Magnitude Regression (MAE in Minutes)

Binary labels (On-Time/Late) are operationally useful for alerting, but administrators also need the expected delay magnitude. We train an XGBoost regressor to predict *delay_minutes* and report mean absolute error (MAE).

TABLE II
DELAY REGRESSION PERFORMANCE (MAE IN MINUTES)

Model	MAE (minutes)
Baseline (mean predictor)	5.01
XGBoost Regressor	1.31

V. EXPLAINABILITY AND OPERATIONAL INSIGHT

A. *Logistic Regression Coefficients*

Figure 1 visualizes coefficient magnitudes of the Logistic Regression model, providing an auditable view of how features contribute to late-arrival risk.

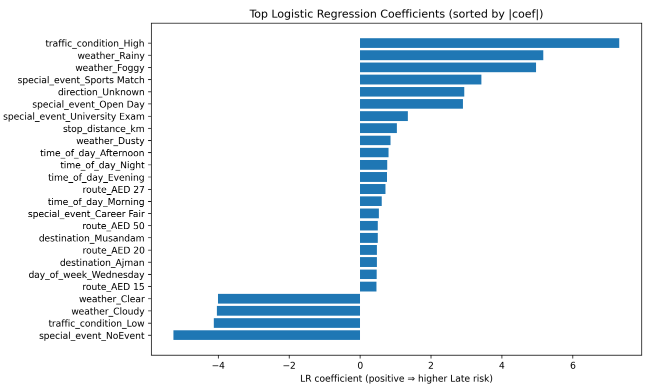

Fig. 1. Logistic Regression coefficient magnitudes for *Late* risk. Positive weights increase late-arrival probability; negative weights decrease it.

B. *SHAP Attributions for XGBoost*

To explain the non-linear XGBoost model, we compute SHAP values [19] and summarize global feature impact using a SHAP summary plot (Fig. 2).

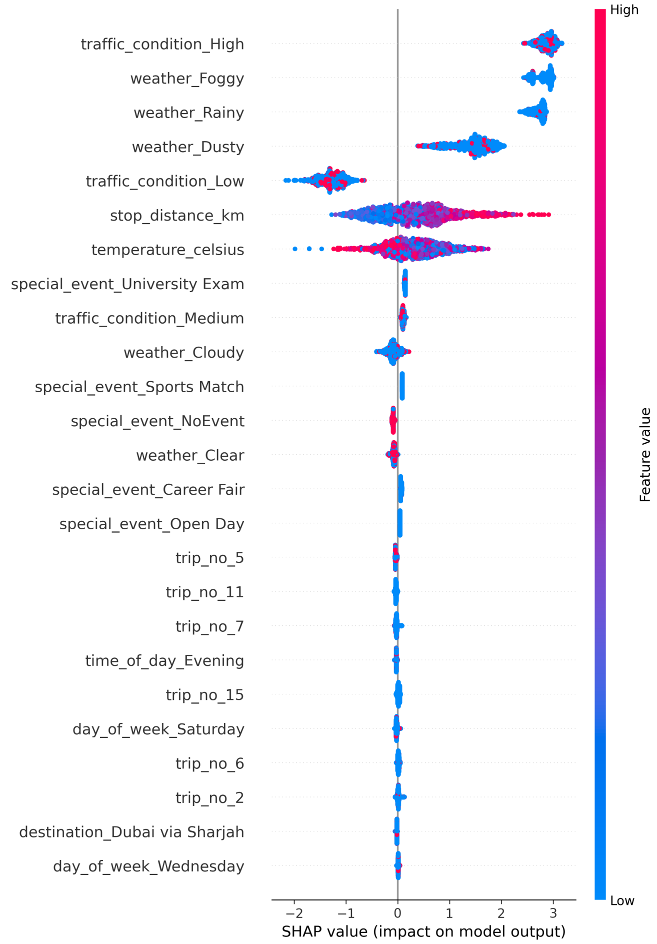

Fig. 2. SHAP summary plot for XGBoost: global feature contributions to late-arrival risk.

VI. SYSTEM ARCHITECTURE AND ADMINISTRATIVE NOVELTY

The novelty of AURAK-Predictor is integrating prediction, evaluation, and explainability into an administrative-facing tool. A lightweight GUI enables operators to (i) query predicted risk and delay magnitude for a scheduled trip, (ii) audit model behavior via coefficient/SHAP visualizations, and (iii) adjust alert thresholds for operational policy.

A. *Imbalance-Robust Evaluation Metrics*

Because accuracy can be misleading under skewed class priors, we visualize Precision–Recall (PR) curves, which are often more informative than ROC curves for imbalanced classification [18].

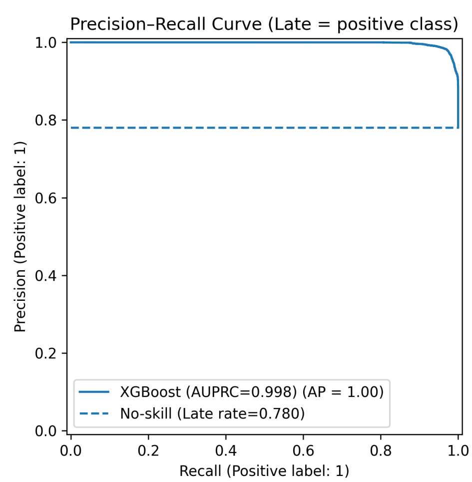

Fig. 3. Precision–Recall curve for the Late class (XGBoost).

B. Runtime and Resource Footprint

All experiments were executed on a commodity laptop platform (12th Gen Intel® Core® i5-12450H CPU, 16 GB RAM). End-to-end inference (including preprocessing) achieved a median single-query latency of 2.59 ms (p90: 3.49 ms) for Logistic Regression and 3.89 ms (p90: 4.97 ms) for XG-Boost. For high-throughput evaluation, batch inference over 1000 queries yielded 0.0054 ms/query (LR) and 0.0087 ms/query (XGB). These results indicate that AURAK-Predictor supports real-time decision support on standard campus hardware.

VII. Implementation Visuals (Selected)

To conserve space, we include only one representative interface view.

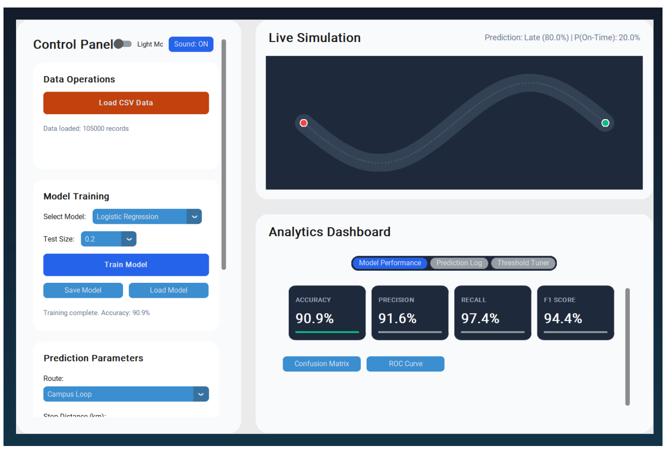

Fig. 4. The AURAK-Predictor Dashboard: integrated view of the ML control panel and evaluation widgets.

VIII. Conclusion and Future Work

This study establishes a proof-of-concept framework for intelligent campus mobility at AURAK. We showed that explainable ML models achieve strong performance under an augmented, class-skewed label distribution, and we report Macro-F1 and Balanced Accuracy to avoid accuracy-only evaluation.

Future work will transition from semi-empirical simulation to deployment by integrating real-time telemetry (e.g., GPS/IoT) and live traffic feeds. We will also explore reinforcement-learning-based control policies for reliability interventions (e.g., holding/dispatch decisions) [21].

Transferability: While our proof-of-concept is trained on AURAK context distributions, the feature schema (time, weather, congestion ratio, route/direction) generalizes to municipal bus systems. A zero-tuning deployment to UAQ/RAK city buses is possible as a baseline but may degrade under domain shift; in practice, we recommend calibration and re-training using a small amount of local telemetry.

Reproducibility: Code is available at `https://github.com/Sedulous-sedu/AI_weather/tree/main`; the augmented dataset is not publicly released due to data-source/usage constraints, but the preprocessing and augmentation scripts are provided to regenerate it.

Acknowledgment

The authors thank the AURAK Engineering Department and Dr. Khaled Abdulgader for their guidance.

References

[1] Z. Fan and S. K. Gurmu, "A Hybrid Machine Learning Approach for Bus Delay Prediction," *IEEE Trans. Intell. Transp. Syst.*, 2022.

[2] M. Altinkaya and M. Zontul, "Urban bus arrival time prediction: A review of computational models," *Int. J. Recent Technol. Eng.*, 2019.

[3] X. Ma et al., "Predicting bus arrival time using a long short-term memory (LSTM) network," *IEEE Trans. Intell. Transp. Syst.*, 2018.

[4] J. Gong et al., "Smart Campus: A Review of the Architecture, Applications, and Challenges," *IEEE Access*, 2021.

[5] A. Al-Fuqaha et al., "Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications," *IEEE Commun. Surv. Tutor.*, 2015.

[6] M. S. Habib et al., "Impact of Dust Storms on Transportation Safety in Arid Regions," *J. Transp. Geogr.*, 2020.

[7] K. Shaaban, "Assessing the Impact of Traffic Congestion in the GCC Region," *Case Stud. Transp. Policy*, 2022.

[8] C. Molnar, *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*, 2020.

[9] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," *Proc. KDD*, 2016.

[10] Ras Al Khaimah Transport Authority (RAKTA), "Public Transport Timetables," Accessed: Dec. 2025. [Online]. Available: https://www.rakta.gov.ae/public-transport/.

[11] S. Alkaabi et al., "Enhancing Urban Mobility with IoT: A Versatile Parking Solution Integrating Real-Time Surveillance and Machine Learning," in *Proc. 2024 IEEE Global Conf. on Artificial Intelligence and Internet of Things (GCAIoT)*, Dubai, UAE, Nov. 2024, doi: 10.1109/GCAIOT63427.2024.10833578.

[12] A. Kuznietsov et al., "Explainable AI for Safe and Trustworthy Autonomous Driving: A Systematic Review," *IEEE Trans. Intell. Transp. Syst.*, vol. 25, no. 12, pp. 19342–19364, 2024, doi: 10.1109/TITS.2024.3474469.

[13] M. Elassy et al., "Intelligent transportation systems for sustainable smart cities," *Transportation Engineering*, vol. 16, p. 100252, Jun. 2024, doi: 10.1016/j.treng.2024.100252.

[14] E. Antypas et al., "A time-series approach for estimated time of arrival prediction in autonomous vehicles," *Transportation Research Procedia*, vol. 78, pp. 166–173, 2024, doi: 10.1016/j.trpro.2024.02.022.

[15] J. dos Santos Sá et al., "Multimodal urban mobility solutions for a smart campus using artificial neural networks for route determination and an algorithm for arrival time prediction," *Engineering Applications of Artificial Intelligence*, vol. 137, 2024, doi: 10.1016/j.engappai.2024.109074.

[16] Q. Zhang et al., "Real-time bus arrival delays analysis using seemingly unrelated regression model," *Transportation*, 2024, doi: 10.1007/s11116-024-10507-3.

[17] Q. Zhang et al., "A causality-based explainable AI method for bus delay propagation analysis," *Communications in Transportation Research*, vol. 5, Art. no. 100178, 2025, doi: 10.1016/j.commtr.2025.100178.

[18] T. Saito and M. Rehmsmeier, "The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets," *PLOS ONE*, vol. 10, no. 3, e0118432, 2015, doi: 10.1371/journal.pone.0118432.

[19] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017. [Online]. Available: https://arxiv.org/abs/1705.07874.

[20] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why Should I Trust You?: Explaining the Predictions of Any Classifier," in *Proc. ACM SIGKDD*, 2016, doi: 10.1145/2939672.2939778.

[21] G. Ai et al., "Deep Reinforcement Learning based dynamic optimization of bus timetable," *Applied Soft Computing*, 2022, doi: 10.1016/j.asoc.2022.109752.

[22] Google, "Distance Matrix API: request and response (duration_in_traffic with departure_time)," Google Maps Platform Documentation. [Online]. Available: https://developers.google.com/maps/documentation/distance-matrix/distance-matrix.