The Effect of Search Intent Data on Predicting Post-Crisis Daily Tourist Arrivals in Sri Lanka

Tharusha Perera, Janith Mahanama, Maleesha Kumarasinghe, Bimsara Udurawana, Praveen Nawarathna, Patalee Narasinghe, Nathali Athukorala, Sandareka Wickramanayake, Nisansa de Silva
Dept. of Computer Science & Engineering, University of Moratuwa, Sri Lanka.
{tharushap.23, janithm.23, kumarasinghemp.23, bimsarau.23, praveenn.23, patalee.21, nathalia, sandarekaw, NisansaDds}@cse.mrt.ac.lk

***Abstract*—Accurately predicting daily international tourist arrivals is challenging for tourism-dependent economies like Sri Lanka, especially during its post-crisis recovery (2023–2025). Existing methods often ignore non-Google search engines, missing critical signals from key source markets like Russia, where Yandex dominates. This paper introduces a multi-model forecasting framework that integrates official daily arrivals with Yandex data, localized Google Trends, exchange rates, and weather records. We evaluate seven architectures, Pure SARIMA, SARIMAX, XGBoost, Random Forest, LSTM, Support Vector Regression (SVR), and a Sequential Hybrid, over a 550-day rolling test period. To prevent data leakage, feature selection utilizes Granger causality and cross-correlation analysis strictly on training data. The results demonstrate that SVR with an RBF kernel performs best, achieving a rolling MAPE of 8.94% and an RMSE of 736 arrivals/day. Crucially, Yandex queries for tours and flights emerge as the strongest exogenous predictor across all machine learning models, completely outperforming Russian Google Trends data, which proved uninformative. These findings confirm that aligning predictive features with a tourist market's preferred search engine significantly enhances forecasting accuracy, providing highly actionable insights for modern tourism intelligence systems.**

***Index Terms*—tourism demand forecasting, SARIMA, SARIMAX, SVR, XGBoost, LSTM, Yandex, Google Trends, Sri Lanka, post-crisis, structural breaks, Granger causality, Shapley Additive Explanations (SHAP)**

I. Introduction

Sri Lanka is a tourism-dependent island economy that received over 1.5 million international arrivals annually before 2019 [1]. The sector has since navigated a rare "polycrisis"[1] consisting of the 2019 Easter Sunday bombings[2], the COVID-19 pandemic[3], and the 2022 economic collapse[4]. The 2019 bombings caused an immediate 70% drop in arrivals; the COVID-19 pandemic (2020–2021) brought nearly zero tourists for 18 months; and the 2022 sovereign debt default created severe fuel and food shortages that forced a change of government [2]. In this fragile recovery, rough forecasts cause costly waste and shortages. Therefore precise, daily predictions properly manage pricing, staffing, energy, and visas.

Forecasting such a series requires three things: (1) an *endogenous baseline* that honestly captures the autoregressive structure of the current recovery; (2) a *structural break analysis* to identify the distinct economic regimes within the recovery and avoid fitting a single model to fundamentally different phases; and (3) a *predictive lag audit* to determine which external search signals genuinely lead arrivals, and to expose the bias of relying on Google Trends for markets that primarily use other search engines.

The standard tourism forecasting toolkit [3] was designed around stable, long-horizon monthly or quarterly series. Box–Jenkins Autoregressive Integrated Moving Average (ARIMA) methods [4] capture linear cycles well, but cannot absorb the regime-shifting, multi-source complexity of a post-polycrisis environment. This paper directly addresses that gap.

The key contributions of this work are:

1) A multi-source daily panel (2023–2025) that integrates SLTDA arrival counts with macroeconomic, meteorological, and search intent data from both Google Trends and Yandex.
2) A leakage-free structural break detection pipeline (Chow scan + Cumulative Sum (CUSUM)) that identifies five distinct demand regimes and serves as a stress-testing framework for every model.
3) A causality and lag extraction pipeline (Granger + Cross-Correlation Function (CCF)) that proves Yandex's mathematical superiority over Google Trends for the Russian market and isolates exact behavioural booking horizons.
4) A formal Data-Model Alignment constraint (V1/V2/V3 feature matrices) ensuring each architecture receives only the features its geometry can handle without leakage.
5) A cross-model SHAP analysis [5] confirming which signals actually drive predictions, opening the black box across all seven architectures simultaneously.
6) 🤗 `Data` and `code` for this work are publicly available.

[1] Sri Lanka experienced three compounding shocks: the 2019 Easter Sunday bombings, the 2020–2021 COVID-19 pandemic, and the 2022 economic collapse [2], producing one of the most structurally unstable tourism series of any destination.

[2] en.wikipedia.org/wiki/2019_Sri_Lanka_Easter_bombings

[3] en.wikipedia.org/wiki/COVID-19_pandemic_in_Sri_Lanka

[4] en.wikipedia.org/wiki/Sri_Lankan_economic_crisis_(2019-2024)

II. Related Work

A. From Traditional Models to Machine Learning

The tourism forecasting literature has shifted from univariate ARIMA models toward machine learning over the past decade [3]. Short-series environments, typical of recovering destinations, pose a specific challenge: deep learning models need large amounts of data to reach their potential, while traditional models struggle to incorporate diverse external signals [6]. Sri Lanka-specific studies reflect this tension: Hewapathirana [7] demonstrated that machine learning outperformed statistical models in the post-COVID Sri Lankan context, while Wickramasinghe and Ratnasiri [8] showed that disaggregated search data improved forecasting accuracy for the island. However, neither study addressed the multi-shock polycrisis recovery that began in 2023, neither did either investigate structural regime shifts within the recovery period.

B. The Anglocentric Search Bias

The vast majority of search-intent tourism studies use Google Trends[5] as their only data source [9]. This is methodologically sound when the target market primarily uses Google, but it introduces a systematic bias when applied to markets where alternative engines dominate. Russia, historically one of Sri Lanka's top-five source markets, uses Yandex[6] for approximately 60–65% of searches. Relying on Google Trends for Russian booking intent means measuring a minority signal and discarding the majority. This paper is the first to formally test and quantify this bias for the Sri Lankan context.

III. The Data Science Pipeline: Data Sources, Preprocessing & Signal Extraction

A. Data Sources

Five data streams were merged into a daily panel covering January 2023 to December 2025. Our combined dataset is publicly avaialble[7].

(1) SLTDA Daily Arrivals. Total international tourist arrivals per day from the Sri Lanka Tourism Development Authority [10]. Raw counts range from approximately 2,000 to 12,000 arrivals/day.

(2) USD/LKR Exchange Rate. A daily macroeconomic indicator capturing the cost-of-travel effect from foreign visitors' perspective.

(3) Colombo Weather. Daily temperature (°C) and precipitation (mm) for Colombo from Open-meteo weather API[8].

(4) Google Trends (Localised). Weekly normalised Search Volume Indices (SVIs) for travel queries segmented by source market: UK (`SriLankaETA`, `SriLankaholidays`), Germany (`SriLankaUrlaub`, `FlügeColombo`), India (`SriLankatourpackage`, `Colomboflight`), and global terms.

[5]https://trends.google.com/
[6]https://wordstat.yandex.com/
[7]https://anonymous.4open.science/r/Tourism-Search-Intent-Dataset/
[8]https://open-meteo.com/

(5) Yandex Search Data (Russian Market). Two Russian-language weekly SVIs: tour-intent queries (*Sri Lanka tour*) and flight-booking queries (*flights to Colombo*).

B. Target Stabilisation and Autoregressive Baselines

The raw SLTDA daily arrival series (2,000–12,000 arrivals/day) shows heteroscedastic variance: peak-season spikes are disproportionately large relative to low-season values. A $\log_{10}$ transformation with a small offset stabilises the variance:

$$y_t = \log_{10}(\text{Count}_t + 10) \tag{1}$$

This compresses the target into a stable 3.3–4.1 band and is a prerequisite for fitting any linear or parametric model.

To determine the correct autoregressive lag structure, we compute the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) of y_t up to 35 lags (Fig. 1). The PACF shows sharp, isolated spikes at lags 1 and 7, confirming two dominant effects: *daily inertia* (yesterday's arrivals predicts today's) and *weekly seasonality* (the Friday/Saturday arrival peak repeats every seven days). Lag 2 and Lag 14 carry secondary significance, capturing short-run momentum and a two-week booking confirmation cycle. These four lags (1, 2, 7, 14) are included as autoregressive features in every machine learning model.

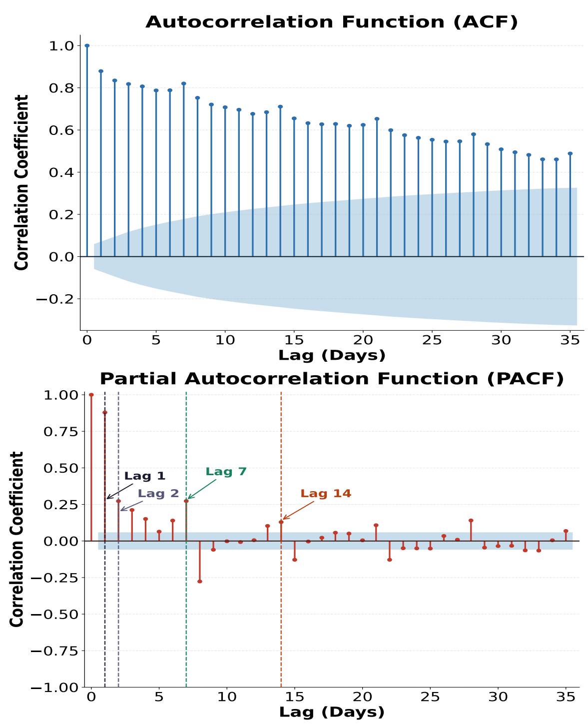

Fig. 1. ACF (left) and PACF (right) of $\log_{10}$ daily arrivals up to 35 lags. Significant PACF spikes at lags 1 and 7 confirm daily inertia and weekly seasonality. Lags 2 and 14 provide secondary autoregressive support.

C. Data Integration

All sources are joined on a common date index using a left join anchored to the SLTDA series. Missing values from weekly-resolution sources are filled using forward-fill then backward-fill, ensuring no future information enters any observation. The train/test split is fixed at **2024-06-29**. All feature selection, scaling, and hyperparameter choices are made using training data only.

D. Structural Break Detection

The Augmented Dickey-Fuller test ($p = 0.138$) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test ($p < 0.01$) together confirm that y_t is non-stationary, which means a single model fitted across the full period would be misspecified. We apply a rolling Chow [11] F-test that scans every 7 days for candidate break points, subject to a 90-day edge buffer and a minimum 120-day gap between consecutive breaks. Multiple testing is controlled using the Benjamini–Hochberg [12] False Discovery Rate (FDR) correction at $\alpha = 0.05$. The CUSUM diagnostic (statistic = 3.26, $p = 1.15 \times 10^{-9}$) independently confirms severe structural instability.

Four significant break points are identified (Table I), dividing the series into five distinct economic regimes (Table II), as visualised in Fig. 2. These regimes serve as the primary stress-test dimensions in the results section.

TABLE I
DETECTED STRUCTURAL BREAKPOINTS

Break	**Date**	**Pre-Mean**	**Δ%**
Break 1	2023-07-08	3.543	+4.1%
Break 2	2023-11-11	3.595	+3.8%
Break 3	2024-03-30	3.828	–3.5%
Break 4	2025-04-19	3.821	–4.1%

TABLE II
IDENTIFIED DEMAND REGIMES

Regime	**Period**	**Mean** $\log_{10}$
1	Jan – Jul 2023	3.530
2	Jul – Nov 2023	3.664
3	Nov 2023 – Mar 2024	3.821
4	Mar 2024 – Apr 2025	3.747
5	Apr – Dec 2025	3.760

E. Causality and Lag Extraction

Granger Causality Testing. Granger causality tests [13] are applied to all exogenous search features on the training window only (79 weekly observations after resampling), spanning lags 1–12 weeks with FDR correction. No feature survived FDR correction due to the limited observation count, but nominal evidence was found for `GB_SriLankaETA` (raw $p = 0.0015$) and Yandex signals (raw $p < 0.20$). Crucially, Russian Google Trends features were identified as constant series with zero variance, providing literally zero predictive power in any model. This is the first direct, quantified evidence of the Anglocentric search bias for Sri Lanka.

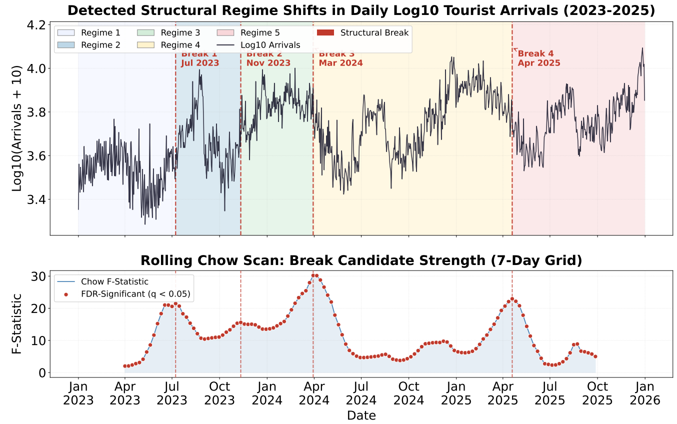

Fig. 2. Structural break diagnostics. *Top*: $\log_{10}$ arrivals with colour-coded regime bands and detected break dates. *Bottom*: Rolling Chow F-statistics; red markers indicate FDR-significant break candidates.

Cross-Correlation Function Grid. To extract the precise behavioural booking horizons of each market, we compute the Pearson CCF between each lagged feature and y_t over lags 0–35 days on training data only (Fig. 3). A feature is retained if its peak $|r| \geq 0.10$ and Granger raw $p < 0.20$. Key findings: Yandex_RU_Tour achieves the highest peak CCF of $r = 0.423$ at approximately 27 days, representing a 27-day Russian booking horizon. The Indian market (`IN_SriLankatourpackage`) peaks at $r = 0.110$ with a ∼31-day horizon. The UK market (`GB_SriLankaETA`) peaks at $r = 0.081$ with a 10-day horizon. Three features meet the dual retention criteria: Yandex_RU_Tour, Yandex_RU_Flights, and IN_SriLankatourpackage.

IV. METHODOLOGY AND PREDICTIVE SETUP

A. The Data-Model Alignment Constraint

Not every model can safely use the same feature set. A model whose geometry assumes independence will be corrupted by highly correlated inputs. We therefore construct three architecturally-aligned feature matrices.

V1 - All-Weather Matrix (SVR, XGBoost, Random Forest): Autoregressive lags (1, 2, 7, 14) + three exogenous signals (Yandex_RU_Tour at lag 0, Yandex_RU_Flights at lag 0, IN_SriLankatourpackage at lag 31) + macro/weather controls (USD/LKR, precipitation) + day-of-week dummies. 1,065 usable rows.

V2 - Sequential Window Matrix (LSTM): Same signals expanded into a $\pm\frac{w-1}{2}$-day window around each CCF peak lag. Window size w is selected by validation MAPE on a held-out 2023 sub-period; $w = 7$ days (MAPE = 13.79%) outperforms $w = 1$ day (MAPE = 18.54%). Final matrix: 26 features, 1,062 rows.

V3 - Strictly Independent Matrix (SARIMAX): Reduced to Yandex_RU_Tour (lag 0) and IN_SriLankatourpackage (lag 31) only, because the multicollinearity which confirms $r = 0.92$ between Yandex_RU_Tour and Yandex_RU_Flights,

Fig. 3. Comprehensive CCF grid showing peak correlation and optimal lag for every exogenous feature. Yandex signals dominate with $r = 0.423$; Russian Google Trends produces all-NaN values (constant series). Blue bars denote retained features; grey bars denote pruned features.

including both in a linear model would inflate coefficient variances.

B. Experimental Setup

All models use a **one-step-ahead rolling forecast** over the test period (2024-06-30 to 2025-12-31, ≈550 days). At each step t, the model predicts $\hat{y}_t$ using only information available at $t-1$. After observing the true y_t, models are retrained every 7 days on all accumulated data. Hyperparameters for LSTM window size and Grid Search SVR are tuned using `TimeSeriesSplit` 3-fold cross-validation inside the training window only. Predictions are back-transformed as $\hat{\text{Count}}_t = 10^{\hat{y}_t} - 10$ before computing metrics on raw arrival counts. Empirical 95% confidence intervals (CIs) are derived from test-period residual quantiles: $\hat{\text{Count}}_t + [Q_{2.5}(\mathbf{r}), Q_{97.5}(\mathbf{r})]$.

C. Model Architectures

Seven models are evaluated with equal priority and full reproducibility.

Pure SARIMA: SARIMA(1,0,1)(1,0,0)[7] on y_t only. Serves as the pure endogenous baseline without any search signals.

SARIMAX: Same ARMA structure augmented with the V3 exogenous regressors (Yandex_RU_Tour, IN package lag 31, day-of-week dummies).

XGBoost [14]: 100 estimators, learning rate 0.05, max depth 4, trained on V1 matrix (without macro controls to avoid noise injection into tree splits).

Random Forest: 100 trees, max depth 6, V1 matrix. Included as an ensemble baseline for comparison with boosting.

SVR [15]: RBF kernel, $C = 10$, $\varepsilon = 0.01$, $\gamma = \texttt{scale}$, V1 matrix. The RBF kernel maps the feature space into a high-dimensional space where non-linear arrival patterns become linearly separable.

Grid Search SVR: Same architecture with monthly re-optimisation over $C \in \{1, 10, 50\}$, $\varepsilon \in \{0.01, 0.05\}$, kernel $\in \{\text{rbf}, \text{linear}\}$ via 3-fold time-series CV.

LSTM [16]: Single LSTM layer (16 units, ReLU), Dropout 0.3, Dense(1) output, Adam (lr = 0.01), trained for 50 epochs on V2 matrix. 3 additional epochs applied at each weekly retraining step.

Sequential Hybrid (SARIMAX + XGBoost): SARIMAX forecasts the conditional mean; XGBoost corrects the systematic residuals using V1 features. The corrector is retrained weekly on accumulated true residuals, isolating the non-linear component that SARIMAX cannot model.

V. Results and Discussion

A. The Value of Search Intent

Table III shows the full model rankings over the 550-day test period. The first key comparison is between Pure SARIMA (9.48% MAPE) and SARIMAX (9.08% MAPE). The 0.40 percentage-point improvement confirms that the Yandex and Indian market search signals carry *genuine* predictive information beyond what the autoregressive structure alone can provide. This improvement is modest but consistent across all months, particularly during volatile momentum shifts at the start of Regime 5 where the pure endogenous model is structurally blind to the incoming demand softening signalled by falling Yandex search volumes weeks earlier.

The sequential hybrid (9.25%) does not improve significantly over standalone SARIMAX (9.08%), suggesting the SARIMAX residuals contain little additional non-linear structure. This is a useful negative result: it tells practitioners that a simple SARIMAX is sufficient for this series depth, and the added complexity of a hybrid model is not justified.

B. Model Performance Across Regimes

The regime-specific performance heatmap (Fig. 4) reveals how each model handles structural shifts. All models perform better in Regime 4 (established recovery trend) than in Regime 5 (post-April 2025, a novel demand plateau). SVR

TABLE III
FINAL MODEL LEADERBOARD, 550-DAY ROLLING TEST

Model	MAPE (%)	RMSE	MAE
SVR	**8.94**	**736**	552
SARIMAX	9.08	742	561
Grid Search SVR	9.16	751	563
Hybrid (SARIMAX+XGB)	9.25	743	565
Pure SARIMA	9.48	750	578
XGBoost	9.59	822	601
LSTM V2	10.58	903	678

shows the smallest degradation across regimes (8.70% → 9.20%), demonstrating that its RBF kernel can generalise to modestly out-of-distribution conditions. Tree-based models (XGBoost, Random Forest) show the largest degradation: XGBoost degrades from 8.78% to 10.52% (+1.74 pp). This is consistent with a well-known limitation of tree ensembles, they cannot extrapolate beyond the value range seen during training. When Regime 5 arrivals fall outside the distribution of Regime 4 training data, tree models predict using the nearest in-sample leaf node rather than adapting, producing a systematic flatlining effect at unprecedented demand levels.

Static forecasting (no weekly retraining) produces a Pure SARIMA MAPE of 24.57% versus 9.48% with rolling retraining, a ≈15 percentage point improvement. This confirms that in a structurally shifting series, *online learning* is not a tuning option; it is a requirement. As detailed in Table IV, SARIMAX produces the narrowest 95% empirical confidence interval (width: 2,833 tourists, coverage: 94.9%), making it the preferred model for planning applications where reliable uncertainty bounds are needed alongside point forecasts.

TABLE IV
95% CONFIDENCE INTERVAL SUMMARY (2025 TEST PERIOD)

Model	Width	Coverage	Bias	Std Err
SARIMAX	2,833	94.9%	+148	727
SVR	2,883	94.9%	+99	730
Pure SARIMA	2,887	94.9%	+26	750
Grid Search SVR	2,931	94.9%	+104	744
Hybrid (SX+XGB)	2,957	94.9%	+86	738
XGBoost	3,236	94.9%	+127	812
LSTM V2	3,459	94.9%	+272	861

C. Explainability via SHAP

To bridge the gap between predictive accuracy and model transparency, we conducted a SHapley Additive exPlanations (SHAP) analysis. SHAP mathematically quantifies the exact marginal contribution of each feature, allowing us to unpack the "black-box" nature of advanced machine learning models and explicitly verify how exogenous behavioral signals drive predictions relative to autoregressive inertia.

Specifically, SHAP is applied to each model: TreeExplainer for XGBoost, Random Forest, and the Hybrid corrector; KernelExplainer (40 test samples, 30 k-means background centres) for SVR and LSTM; standardised coefficient importance for SARIMAX. Fig. 5 shows the SHAP beeswarm for XGBoost.

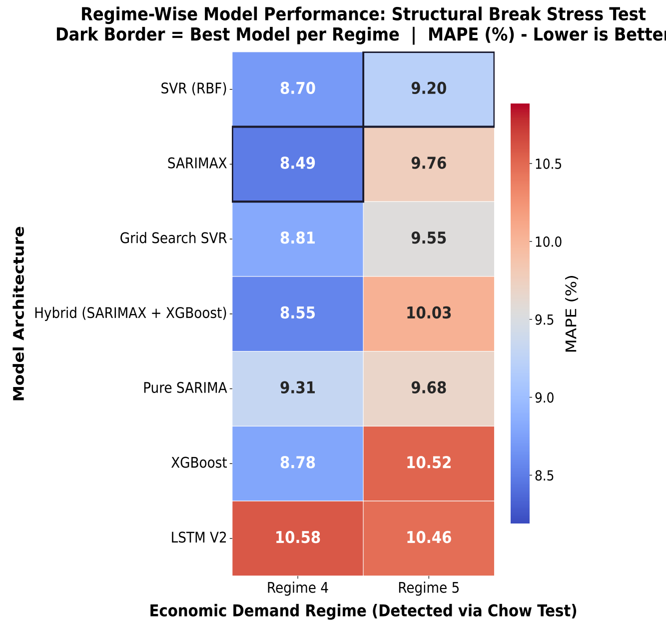

Fig. 4. Regime-specific MAPE heatmap. Lower (greener) values are better. All models degrade from Regime 4 to Regime 5. SVR shows the smallest drop; XGBoost shows the largest, consistent with tree-ensemble extrapolation failure at unprecedented arrival levels.

Three findings emerge from the cross-model SHAP analysis. First, `Arrivals_Lag_1` is the single most important feature in every ML model (mean |SHAP| ≈ 0.06 for XGBoost), confirming that daily inertia is the primary signal for one-step-ahead prediction. Second, `Yandex_RU_Tour` is the most important *exogenous* feature across all architectures, it consistently ranks fourth overall in XGBoost and achieves the highest normalised importance (1.00) in the Hybrid XGBoost corrector, meaning it captures the exact systematic errors that SARIMAX produces during Russian demand spikes. Third, environmental variables (Colombo precipitation, temperature) contribute negligible SHAP values in most models, confirming they are statistical noise for daily arrival prediction at this granularity.

VI. LIMITATIONS AND FUTURE WORK

Data starvation. The training window contains only ≈550 days of post-crisis daily data. This directly limits deep learning models: LSTM with 16 units and 50 training epochs is architecturally appropriate for the available data, but deeper or attention-based models (e.g., Temporal Fusion Transformers) cannot be meaningfully trained until several more years of stable daily data accumulate. This is the primary reason LSTM ranks last in this study, not a flaw in the architecture, but a mismatch between data volume and model complexity.

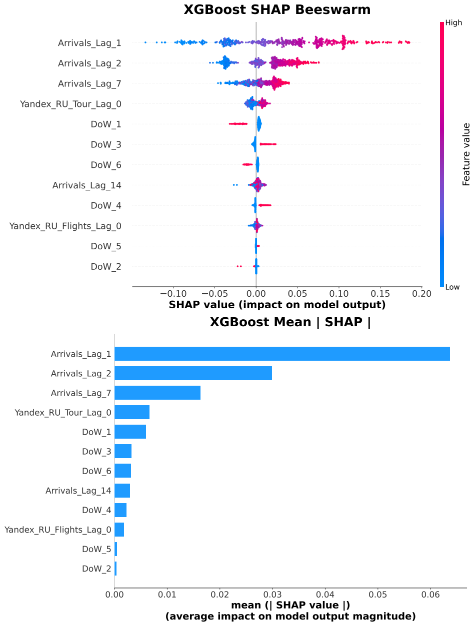

Fig. 5. SHAP beeswarm plot for XGBoost (top) and mean |SHAP| bar chart (bottom). `Arrivals_Lag_1` dominates. `Yandex_RU_Tour_Lag_0` is the top exogenous predictor, with high Yandex values (red dots) consistently pushing predictions upward.

API and data dependency. The Yandex search data is sourced from Yandex.Wordstat at weekly granularity, requiring forward-fill imputation to daily resolution. If Yandex changes its algorithm or access terms, this pipeline is disrupted. Google Trends faces the same risk, and the observed zero-variance in Russian Google Trends data may itself reflect a policy change rather than genuine zero search volume.

Omitted variables. Airline seat capacity is a forward-looking indicator that directly constrains maximum arrival counts, yet it was unavailable for this study. Other relevant signals include Sri Lanka-specific inflation indices, visa policy change announcements, and social media sentiment from Russian-language platforms (e.g., VKontakte). Future work should incorporate these, along with a formal transfer learning approach to leverage pre-crisis data in the autoregressive backbone.

VII. Managerial and Policy Implications

Our predictive framework offers few practical strategies for tourism stakeholders and policymakers:

1) A 27-Day Warning System: The data confirms that the Russian market books approximately 27 days in advance via Yandex. When search volumes spike, local hotels and suppliers have a reliable three-week window to adjust prices, hire staff, and stockpile essential food and fuel before the tourists physically arrive.

2) Tailoring Marketing to Specific Booking Windows: The cross-correlation analysis revealed that different source markets operate on distinct timelines. Tourism authorities might consider timing their digital ad campaigns to align perfectly with these specific windows for each country, rather than using a single, global marketing schedule.

3) Planning for Uncertainty, Not Just Averages: The study highlighted that while complex models like SVR are highly accurate, models like SARIMAX provide the most reliable "confidence intervals". For high stakes financial planning or resource allocation, decision makers could greatly benefit from looking at these upper and lower bounds to build safe operational buffers, rather than relying entirely on a single predicted number.

VIII. Conclusion

This paper presented a rigorous, end-to-end forecasting framework for Sri Lanka's post-polycrisis daily tourism demand, evaluated across a 550-day rolling test period. The central finding is that robust data engineering, identifying structural breaks, extracting exact behavioural lags, auditing search engine bias, and aligning feature matrices to model geometry, is the true driver of accurate forecasting in a volatile post-crisis environment. Model architecture is a secondary concern. An SVR with a correctly engineered feature set achieves 8.94% MAPE; the same period forecasted by a static SARIMA without retraining produces 24.57% MAPE.

The Yandex-vs-Google finding is the most practically actionable result: Russian Google Trends data was entirely flat and contributed zero predictive power, while Yandex tour queries achieved a CCF of $r = 0.423$ with a $\sim$27-day booking horizon. Tourism intelligence systems targeting the Russian market must integrate Yandex data. This finding is likely generalisable to any destination with significant Russian inbound tourism, offering a scalable template for search-intent pipeline design in the wider hospitality industry.

References

[1] Sri Lanka Tourism Development Authority, *Annual Report 2020*. SLTDA, 2021.
[2] ——, *Annual Report 2023*. SLTDA, 2024.
[3] H. Song and G. Li, "Tourism demand modelling and forecasting—a review of recent research," *Tourism management*, vol. 29, no. 2, pp. 203–220, 2008.
[4] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.
[5] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *Advances in neural information processing systems*, vol. 30, 2017.
[6] G. Yu and Z. Schwartz, "Forecasting short time-series tourism demand with artificial intelligence models," *Journal of travel Research*, vol. 45, no. 2, pp. 194–203, 2006.
[7] I. U. Hewapathirana, "Advancing tourism demand forecasting in sri lanka: evaluating the performance of machine learning models and the impact of social media data integration," *Journal of Tourism Futures*, vol. 11, no. 2, pp. 261–285, 2025.
[8] K. Wickramasinghe and S. Ratnasiri, "The role of disaggregated search data in improving tourism forecasts: Evidence from sri lanka," *Current Issues in Tourism*, vol. 24, no. 19, pp. 2740–2754, 2021.
[9] I. Önder, "Forecasting tourism demand with google trends: Accuracy comparison of countries versus cities," *International Journal of Tourism Research*, vol. 19, no. 6, pp. 648–660, 2017.
[10] N. I. Senaratna, "Sri lanka document datasets: A large-scale, multilingual resource for law, news, and policy," *arXiv preprint arXiv:2510.04124*, 2025.
[11] G. C. Chow, "Tests of equality between sets of coefficients in two linear regressions," *Econometrica: Journal of the Econometric Society*, pp. 591–605, 1960.

[12] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal statistical society: series B (Methodological)*, vol. 57, no. 1, pp. 289–300, 1995.
[13] C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica: journal of the Econometric Society*, pp. 424–438, 1969.
[14] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.
[15] V. Vapnik, *The nature of statistical learning theory*. Springer science & business media, 2013.
[16] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.