

MERCon 2026 - Engineering Research Unit

Forecasting Weather-Driven Price Dynamics Across Sri Lankan Tea Market Catalogues

Hesandi Mallawarachchi Senilka Madurapperumage Nadil Kulathunge Thilokya Angeesa Nethsith Gunaweera Sandeepa Weerasekara Patalee Narasinghe Nisansa de Silva Sandareka Wickramanayake

Department of Computer Science & Engineering, University of Moratuwa.

Background on Sri Lankan Tea

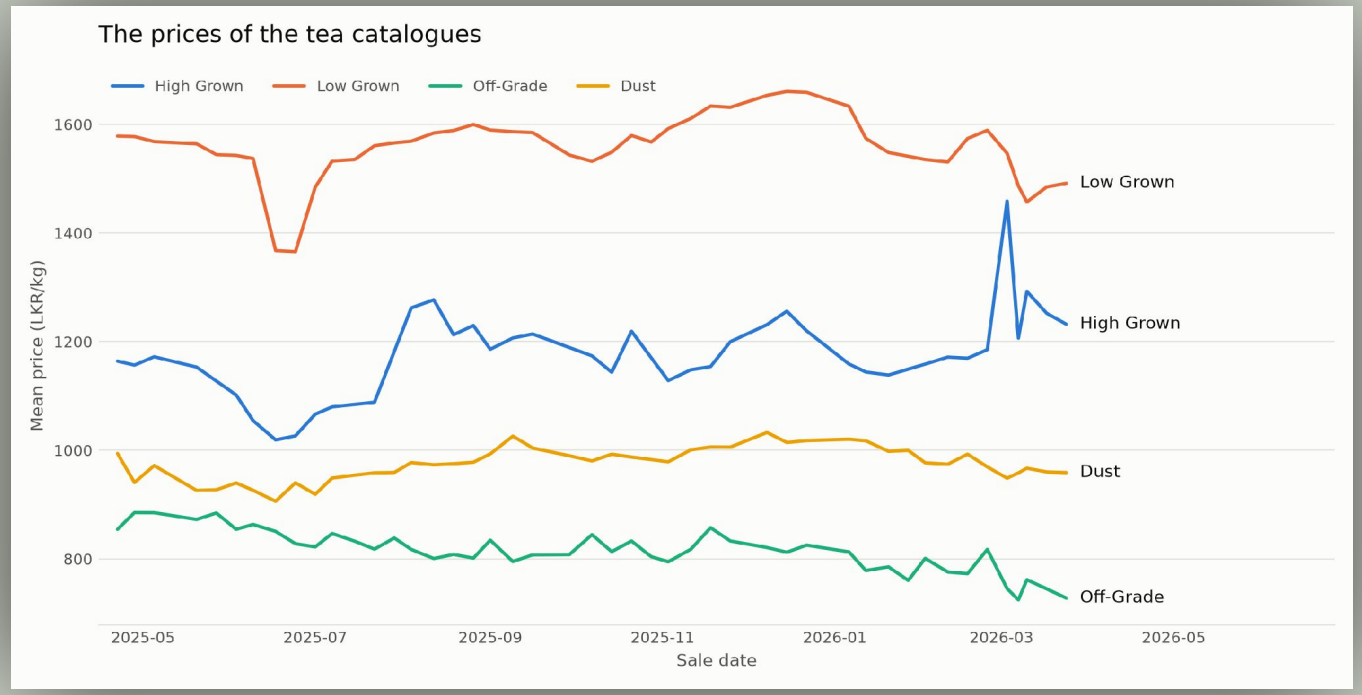
Tea growing areas - Classified by elevation

High Grown	> 1,200 m
Medium Grown	600 - 1,200 m
Low Grown	< 600 m

Classified by catalogue - how it is sorted and sold at auction

High Grown	Mainly tea grown from High and medium areas
Low Grown	Low Grown production, subdivided by leaf size and style, prized for visual appeal and brisk taste.
Off Grade	Teas outside standard leaf grades, often with more fiber, stalk, or broken leaves.
Dust	The smallest particle fraction, the finest tea dust.

Background on Sri Lankan Tea



- Most previous work has been done considering a single mean price for tea in Sri Lanka.
- But the Sri Lankan tea is clearly four different markets and therefore cannot be considered as one single price.

Research Objectives

- Measure the statistical relationship between weather variables and tea auction price movements
- Evaluate whether weather-based features improve the predictive accuracy of tea price forecasting models compared to baseline models

What We Asked – The Problem Statement

Does weather information provide predictive value for tea auction prices and how it affects each tea catalogues separately ?

From Raw PDFs to a Modelling-Ready Table

1. Collection → 2. Preprocessing & Feature Engineering → 3. Time-Series Diagnostics → 4. Modelling & Evaluation → 5. Pattern Analysis

Price stream

- Forbes & Walker weekly market reports (105, one of 8 licensed CTA brokers)
- Custom multi-table parser built on pdfplumber

Weather stream

- Open-Meteo historical archive API
- 4 region centroids: Maskeliya, Nuwara Eliya, Uva/Uda Pussellawa, Matara-Galle
- Weekly precipitation, mean temperature, sunshine duration → 7-day aggregation windows

Preprocessing, Features & Lag Design

1. Collection → **2. Preprocessing & Feature Engineering** → 3. Time-Series Diagnostics → 4. Modelling & Evaluation → 5. Pattern Analysis

$$\text{Lag reference date} = \text{auction date} - 7 \times \text{lag}$$

Why a short primary window?

- Targets near-term harvest and supply
- Explicitly not a model of the full biological crop-to-auction pathway

Feature groups

Group	Features
Lagged weather	precipitation, mean temp, sunshine
Macro	fx_usd (LKR/USD rate)

Time-Series Diagnostics

1. Collection → 2. Preprocessing & Feature Engineering → **3. Time-Series Diagnostics** → 4. Modelling & Evaluation → 5. Pattern Analysis

Stationarity

(ADF - Augmented Dickey-Fuller)

- Low Grown non-stationary in levels → first-differenced.
- High Grown, Off-Grade, Dust stationary ($p < 0.05$). Weather variables generally stationary.

Catalogue	p-value	Verdict
High Grown	0.0007	Stationary → levels
Low Grown	0.1087	Unit root → differenced
Off-Grade	0.0325	Stationary → levels
Dust	0.0024	Stationary → levels

Time-Series Diagnostics

Granger Causality Test

Catalogue	Strongest driver	Best lag	p-value
Dust	Temperature	4 weeks	0.0003
Off-Grade	Temperature	1 week	0.015
Low Grown	Precipitation	2 weeks	0.015
Low Grown	Sunshine	1 week	0.020

- **96 Granger tests were attempted** (4 catalogues × 3 weather variables × 8 lags)
- There was a risk that some of these significant results happened just by chance.
- Therefore, we applied the ***Benjamini-Hochberg correction*** to control the false discovery rate (FDR).

Modeling Protocol & Evaluation Strategy

1. Collection → 2. Preprocessing & Feature Engineering → 3. Time-Series Diagnostics → **4. Modeling & Evaluation** → 5. Pattern Analysis

Price-only baselines

Naive Persistence : $\hat{Y}_{t+1} = Y_t$

AutoRegression : $\hat{Y}_{t+1} = \propto + \beta Y_t$

Tree ensembles

- Random Forest
- Gradient Boosting
- XGBoost
- LightGBM

No Single Model Wins Everywhere

Catalogue	Best method	RMSE (LKR)	R^2	vs. Naive Persistence
High Grown	LightGBM	167.02	0.448	**−33.31%**
Off-Grade	LightGBM	102.15	0.511	**−45.78%**
Dust	XGBoost	114.82	0.633	**−52.42%**
Low Grown	**Naive Persistence**	**79.76**	**0.990**	**ML is worse**

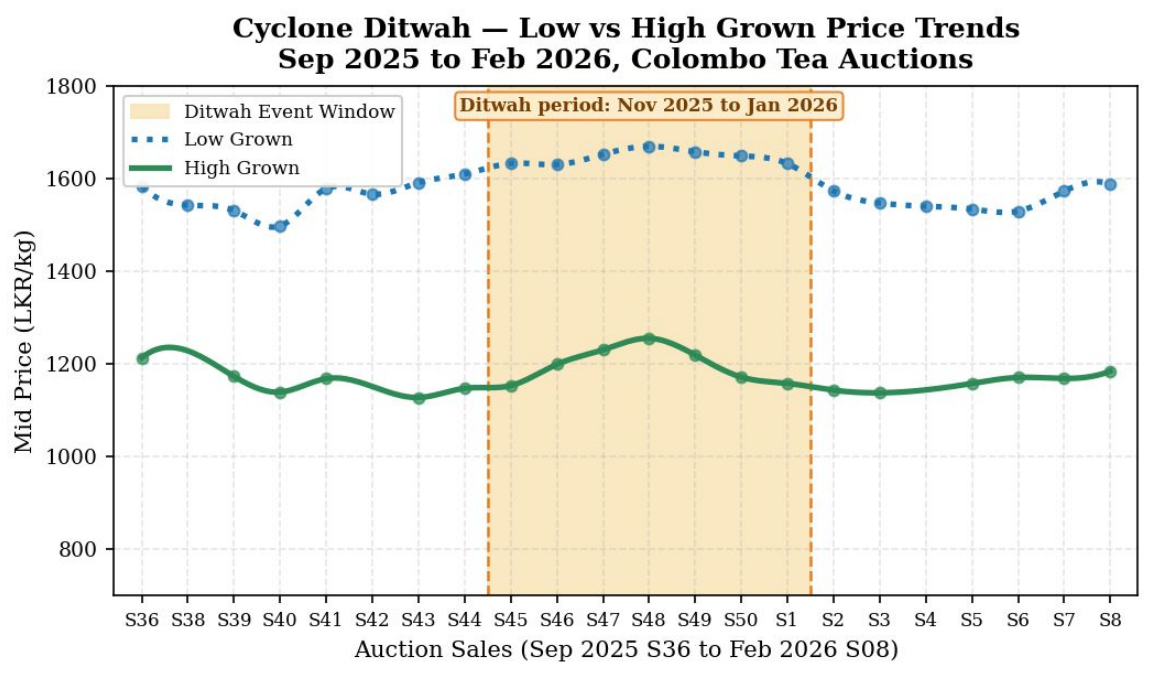
- **High Grown, Off-Grade and Dust catalogues reward machine learning while Low Grown rewards the Naive Baseline.**

- **In Low Grown, Week-to-week prices are dominated by short-term autocorrelation - extra features add variance, not signal.**

A Natural Experiment: Cyclone Ditwah

Event-window analysis, 28 November 2025

1. Collection → 2. Preprocessing & Feature Engineering → 3. Time-Series Diagnostics → 4. Modeling & Evaluation → **5. Pattern Analysis**



What the window shows

- Heavy flooding and landslides across major growing areas - a supply shock
- Both catalogues rise into a mid-window peak; High Grown spikes most sharply

Main Contributions

Public Dataset

105 broker PDFs parsed into 12,233 records, 26 features, 105 sales, Nov 2023 - Mar 2026. Publicly released.

Catalogue-specific ML models

Four tree ensembles benchmarked against price-only baselines, per catalogue.

See the dataset - Hugging Face https://huggingface.co/datasets/colombo-tea-auction-prices/cta-price-prediction

See our work - Github https://github.com/hesandism/data-analysis-for-tea-industry

Practical implications, limitations & future work

For practitioners

- Match the model to the catalogue
- Simple baselines are legitimate deployments - Low Grown persistence costs nothing to run

Limitations

- Auction-day signals unobserved: buyer participation, bidding behaviour, blend-maker decisions
- 105-week window, grid search limited to 8 configs per algorithm

Future work

- Incorporate real-time bidding and buyer-activity data
- Extend the study window for more sales

Market dynamics are the primary driver. Weather is a secondary, catalogue-specific refinement.

Conclusion

1. **The CTA is not a monolithic series - catalogue-specific evaluation is necessary, not optional**

2. **Weather is a limited, catalogue-specific signal**

3. **Tree models beat persistence in 3 catelogues - but persistence wins Low Grown outright**

4. **A public, machine-readable CTA dataset now exists with 12,233 records including 26 features spanned through 105 sales**



References

1. Sri Lanka Tea Board, Annual Report 2024. Colombo, Sri Lanka: Sri Lanka Tea Board, 2024, accessed: 2026-06-27.

2. K. B. A. Pasandul, R. V. Pathirana, S. S. Chandrasinghe, H. A. B. S. Mahanama, and G. Pereraa, ***"Tea price trend analysis and forecasting in Sri Lanka: A time series approach,"*** in Transformative Applied Research.

3. S. Aslan and H. Ombao, ***"A practical approach for exploring Granger connectivity in high-dimensional networks of time series"***

4. M. W. A. De Silva and N. S. Cooray, ***"The export performance of the Sri Lankan tea: An econometric analysis,"*** International Journal of Research and Innovation in Social Science.

5. B. A. T. K. Rupasinghe and S. H. P. Malkanthi, ***"An investigation of the relationship between exchange rate volatility and volume of tea export in Sri Lanka"*** Faculty of Social Sciences and Languages, Sabaragamuwa University of Sri Lanka.

References

6. M. A. Wijeratne, ***"Vulnerability of Sri Lanka tea production to global climate change,"*** Water, Air, and Soil Pollution

7. S. Ahmed, T. Griffin, S. B. Cash, W.-Y. Han, C. Matyas, C. Long, C. M. Orians, J. R. Stepp, A. Robbat, and D. Xue, ***Global climate change, ecological stress, and tea production,"*** in Stress Physiology of Tea in the Face of Climate Change

8. P. Zippenfenig, ***"Open-Meteo.com weather API,"*** Zenodo. https://doi.org/10.5281/zenodo.14582479

9. V. Verma, "***A review of the Benjamini–Hochberg method for multiple hypothesis testing,"*** International Journal of Research.

Thank You !!!

Any Questions ?