Department of Statistics,
Faculty of Science, University of Colombo
B.Sc. (Honors) in Industrial Statistics
Final Year Research Project Proposal-2025

Name	K.G.A. Yohan
Title of the Research Project	*A Hybrid Machine Learning Framework for Forecasting Onion, Coconut, Beans, and Carrot Prices in Sri Lanka's Dambulla Market: Integrating Climatic, Economic, and Sentiment Drivers*

For Coordinator's use only			
Decision	**Accepted/Not Accepted**	**Date**	
Remark			
Supervisor			
Potential Examiners			

Introduction / background	Price volatility in key agricultural commodities—onion, coconut, beans, and carrot—severely disrupts Sri Lanka's food supply chains, impacting farmers, traders, and consumers. The Dambulla Dedicated Economic Centre, a major agricultural trading hub, experiences extreme price fluctuations due to climate shocks, market speculation, and policy instability. Traditional forecasting models (e.g., ARIMA, VAR) fail to capture non-linear interactions between these factors, leading to unreliable predictions. This study proposes a machine learning (ML)-driven price forecasting model that integrates climatic, economic, and socio-political data to improve prediction accuracy. By providing real-time insights, this research aims to enhance market stability, support policymaking, and safeguard livelihoods in Sri Lanka's agricultural sector
Research question or Objectives	➢ To develop a hybrid machine learning model (LSTM + Attention Mechanisms) for forecasting daily wholesale prices of onion, coconut, beans, and carrot in Dambulla market ➢ To Identify key price drivers (e.g., monsoon impact on onions, fuel costs on coconuts) using explainable AI (SHAP/LIME) ➢ To Compare model accuracy (RMSE, MAPE) against traditional methods (ARIMA, Prophet) and baseline ML models (Random Forest) ➢ To Design a prototype dashboard for real-time price forecasts accessible to farmers, traders, and policymakers

Data

This study utilizes historical datasets spanning from January 2015 to present,

Data Type	Source	Frequency
Carrot, Beans, Onion, Coconut Price	Dambulla Market Wholesale Section https://www.cbsl.lk/eresearch/	Daily
Weather Data (Temperature, Precipitation(mm), Precipitation hours(h), Wind Speed (Km/h), Sunshine)	https://open-meteo.com	Daily
News Headlines	https://www.infolanka.com/news/	Daily
Fuel Prices	https://ceypetco.gov.lk/historical-prices/	
Exchange Rate	https://www.cbsl.lk/eresearch/	Monthly
Festival Dates	Internet	

Exchange rate and fuel price are not reported daily. Therefore, I apply the forward fill method to extend the last known value across the missing days, aligning all features to a daily frequency.

Significance of the study

This research addresses a pressing gap in agricultural economics by developing a multi-commodity price forecasting system using machine learning. By integrating real-time market sentiment and cross-commodity dependencies, the study aims to:

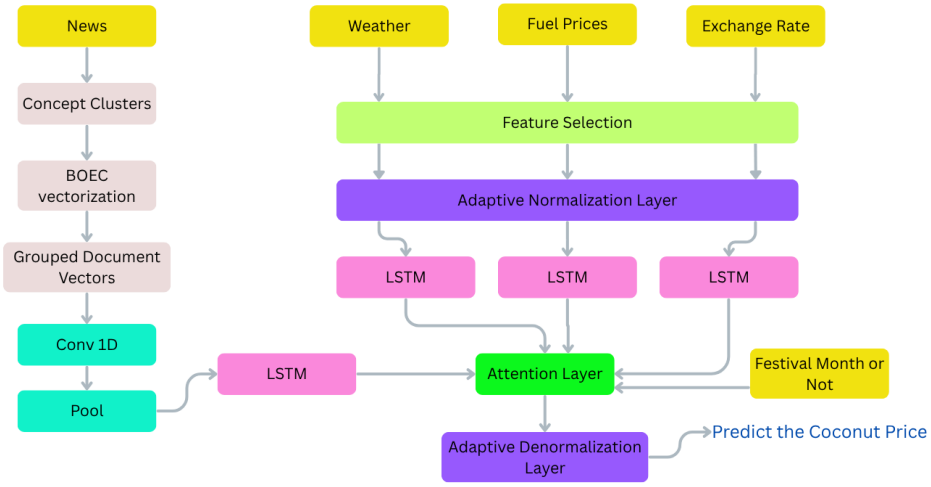
1. Enhance predictive accuracy beyond traditional econometric models.
2. Empower stakeholders with actionable insights to mitigate financial risks.
3. Strengthen food security through data-driven supply chain planning. The outcomes will directly benefit Sri Lanka's agricultural sector and serve as a scalable framework for other regions.

Suggested Methodology

This study implements a rigorous two-stage feature selection process to identify the most influential price drivers. First, Granger causality tests statistically evaluate potential factors (e.g., weather patterns, fuel prices, exchange rates) to determine their causal relationship with commodity price movements. Second, XGBoost machine learning further refines these candidates by quantifying each feature's predictive importance (e.g., identifying rainfall as a key driver for onion prices while discounting less relevant factors).

The selected features feed into an advanced hybrid neural network architecture featuring:

- **LSTM recurrent layers** that capture complex temporal patterns in price and weather data across defined historical windows
- **Attention mechanisms** that automatically identify and emphasize high-impact events by analyzing synchronized news sentiment trends alongside market indicators
- **Adaptive normalization layers** that maintain consistent feature scaling despite differing units and volatility levels across input types

	The model undergoes comprehensive training using 2015-2021 historical data, with rigorous validation against the turbulent 2022-2023 post-pandemic period. Performance metrics including MAE and MAPE provide quantitative comparisons against traditional benchmarks (ARIMA, Prophet models). Finally, SHAP explainability analysis transparently reveals each factor's contribution (e.g., demonstrating how diesel price fluctuations disproportionately affect coconut transportation costs), delivering actionable insights for agricultural stakeholders 
References (optional)	https://www.mdpi.com/2077-0472/13/9/1707 https://iopscience.iop.org/article/10.1088/1742-6596/2425/1/012068/pdf https://www.mdpi.com/2227-7390/11/5/1132