Forecasting Weekly Wholesale Tomato Prices in Major Indian Markets: Project Summary

May 12, 2025

# Abstract

This report summarizes the development, validation, and practical implications of an ensemble‑based forecasting framework designed to predict weekly wholesale tomato prices across twenty‑five major Indian markets for the period 2015–2024. Leveraging 10 years of Agmarknet price data, locational weather variables, and policy/event indicators, we benchmarked classical statistical, machine‑learning, and deep‑learning models. A stacking ensemble reduced forecast error by 10–15 percent relative to the best standalone model, offering stakeholders an early‑warning tool for price volatility management.

# 1. Project Overview

Volatile tomato prices create income uncertainty for farmers, procurement risk for traders, and inflationary pressure for consumers. The objective of this project is to deliver a data‑driven early‑warning system that forecasts weekly wholesale prices one to four weeks ahead, enabling evidence‑based decisions on harvest timing, storage, logistics, and policy intervention.

# 2. Data Sources

• \*\*Agmarknet Weekly Prices (2015–2024):\*\* 25 principal markets spanning four metropolitan hubs and key regional centres.  
• \*\*Indian Meteorological Department (IMD):\*\* Average weekly maximum and minimum temperature, cumulative rainfall, and relative humidity.  
• \*\*Policy & Event Flags:\*\* COVID‑19 lockdown (Q2 2020), export bans, festival‑driven demand shocks.

# 3. Methodology

## 3.1 Model Development

Five model families were calibrated for each market:  
1. Autoregressive Integrated Moving‑Average with eXogenous regressors (ARIMAX)  
2. Seasonal ARIMAX (SARIMAX)  
3. Facebook Prophet (additive seasonality with changepoints)  
4. Extreme Gradient Boosting (XGBoost) with lagged features & weather covariates  
5. Long Short‑Term Memory (LSTM) neural network using sliding windows

## 3.2 Ensemble Strategy

A two‑layer stacking ensemble was trained using rolling origin cross‑validation. Base‑learner predictions served as inputs to a ridge‑regularized meta‑learner that optimally weighted each model for every forecast horizon.

# 4. Validation & Performance

Rolling time‑series cross‑validation with a 52‑week sliding window generated out‑of‑sample forecasts. Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) were averaged across all markets.  
  
\* \*\*Stacking Ensemble:\*\* MAE ≈ ₹1.45 kg⁻¹; RMSE ≈ ₹2.10 kg⁻¹  
\* \*\*Best Single Model (LSTM):\*\* MAE ≈ ₹1.63 kg⁻¹; RMSE ≈ ₹2.35 kg⁻¹  
\* Weather and policy covariates lowered MAE by an additional 5–10 percent relative to price‑only baselines.

# 5. Key Findings

• Ensemble forecasts consistently outperformed individual models across all 25 markets.  
• Deep‑learning (LSTM) captured non‑linear dynamics, while tree‑based XGBoost handled sudden shocks; their complementary strengths improved the meta‑learner.  
• Weather variables were most influential during monsoon months; policy flags dominated variance during COVID‑19 lockdown weeks.

# 6. Practical Implications

Accurate price foresight empowers farmers to optimise harvest schedules and negotiate contracts, allows traders to allocate storage and logistics assets efficiently, and provides policy‑makers with an early warning signal for consumer price inflation. A weekly dashboard integrating these forecasts could reduce post‑harvest losses and improve farm‑gate realisations.

# 7. Recommendations & Future Work

1. \*\*Operational Deployment:\*\* Build a cloud‑hosted dashboard that automatically refreshes forecasts every week.  
2. \*\*Crop Extension:\*\* Replicate the modelling pipeline for onions, potatoes, and other perishables.  
3. \*\*Higher‑Frequency Data:\*\* Incorporate daily arrivals, satellite vegetation indices, and retail prices to improve short‑horizon accuracy.  
4. \*\*Economic Impact Study:\*\* Quantify potential income gains and waste reduction under different adoption scenarios.

# 8. References

Agmarknet. (2024). Wholesale price dataset. Ministry of Agriculture & Farmers Welfare, Government of India.  
Indian Meteorological Department. (2024). Climatic Data Services Division Reports.  
Taylor, S. J., & Letham, B. (2018). Forecasting at Scale. \*The American Statistician\*, 72(1), 37‑45.