Forecast Fusion

A Project-Based Learning Report Submitted in partial fulfilment of the requirements for the award of the degree

of

Bachelor of Technology

in The Department of Computer Science and Engineering

Deep Learning - 23AVI3101A

Submitted by

2310030020: V. Shiva Shanmukh Teja

2310030002: K. Sri Aashritha

Under the guidance of

Dr. Sumit Hazra

Department of Computer Science and Engineering

Koneru Lakshmaiah Education Foundation, Aziz Nagar

Aziz Nagar – 500075

AUG - 2025.

Introduction

Context and Importance

Spices play a crucial role in agriculture, cuisine, and commerce, particularly in countries like India, the world's largest producer, consumer, and exporter of spices. The spice industry not only supports the livelihoods of millions of farmers and traders but also contributes significantly to the national economy through domestic consumption and exports.

Despite their importance, the pricing of spices is highly unpredictable. Multiple factors contribute to this volatility, including:

- Seasonality: Many spices are harvested during specific periods, leading to supply fluctuations throughout the year.
- Arrivals: The volume of spices arriving at markets can suddenly increase or decrease, directly affecting prices.
- Quality Variations: Differences in variety or grading of spice crops influence price points.
- Climatic Conditions: Weather patterns such as monsoons, droughts, or unseasonal rains impact both production and prices.
- Market Dynamics and Policy Changes: Changing export policies, international demand, and competition also contribute to price swings.

Given these influences, spice prices can rise or fall unpredictably, creating significant challenges for those involved in the supply chain.

Problem Statement

One of the main challenges in the spice sector is the lack of reliable and accurate price forecasting. Stakeholders, especially small and marginal farmers, often make selling or stocking decisions based on guesswork or outdated information. This exposes them to financial risks such as:

- Selling goods at inopportune times when prices are low.
- Holding stock too long and missing opportunities for better returns.
- Making procurement decisions without understanding future trends.

Consequently, poor forecasting can lead to substantial financial losses, inefficient resource allocation, and increased vulnerability for those dependent on the spice market.

Objective

The primary objective of this project is to develop a data-driven, machine learning-based forecasting system capable of predicting the prices of selected spices for the next 30 days. This system aims to harness historical and recent price data, along with other relevant factors, to provide accurate and timely forecasts. With such a tool, stakeholders in the spice industry can:

- Anticipate future price trends with greater confidence.

- Make well-informed decisions about buying, selling, or holding inventory.
- Strategically plan procurement, sales, and exports to maximise gains and minimize losses.

Benefits

Implementing an effective price forecasting model brings tangible advantages:

- Reduced Financial Losses: Stakeholders can mitigate losses by avoiding sales at unfavourable prices and capitalizing on profitable market windows.
- Improved Decision-Making: Timely information enables data-driven choices instead of reliance on speculation or rumours.
- Optimized Procurement and Storage: Traders and manufacturers can plan inventories more efficiently by purchasing raw materials when prices are expected to be favourable.
- Supply Chain Stability: Predictable pricing benefits everyone from farmers to retailers, fostering stability in the spice trade ecosystem.
- Empowerment of Farmers and Small Traders: Access to accurate market forecasts bridges the information gap and helps smaller players compete fairly.

Overall, the ability to accurately forecast spice prices not only supports economic growth but also enhances the livelihoods and resilience of those who rely on this vibrant sector.

Literature Review/ Application Survey

Spices such as **large cardamom, tamarind, turmeric, and mace** are considered critical export-oriented crops in countries like **India, Nepal, and Indonesia**, which collectively dominate the global spice trade. These crops are highly valued not only for their culinary and medicinal properties but also for their contribution to rural livelihoods and foreign exchange earnings. However, their prices exhibit **significant volatility**, driven by multiple interacting factors such as:

- **Seasonality of production** (linked to specific harvest windows).
- **Weather shocks**, including variations in rainfall, temperature, and humidity that directly affect yield and quality.
- **Pest and disease outbreaks**, which can drastically reduce supply in certain years.
- **Shifts in international demand** due to changing consumer preferences, health awareness, and global spice trade dynamics.
- **Supply chain disruptions**, including transportation bottlenecks, storage challenges, and export policy changes.

Given these complexities, **accurate forecasting of spice prices** is essential. Forecasts enable farmers and cooperatives to plan cropping and storage strategies, exporters and traders to hedge against market risk, and policymakers to implement price stabilization mechanisms and design appropriate market interventions.

Data Sources for Spice Price Forecasting

The forecasting of agricultural commodity prices depends heavily on the **availability and reliability of data**. For spices, the following sources are commonly referenced:

1. **Government portals**: The Indian Agmarknet and the Spices Board of India publish daily mandi (market) prices, providing granular, region-specific information.
2. **Commodity boards and exporters**: Reports from the Spices Board and trade statistics from FAO (Food and Agriculture Organization) serve as valuable secondary datasets.
3. **International trade statistics**: Datasets such as **UN Comtrade** and **ITC Trademap** reflect global demand and supply shifts, particularly relevant for export-sensitive commodities like cardamom and turmeric.
4. **Auxiliary data features**: Weather datasets (from IMD, NOAA), seasonality indicators (festivals, export cycles), and financial variables such as currency exchange rates (USD–INR) are increasingly used as explanatory variables in advanced models.

Forecasting Methods in Agri-Commodities

1. Statistical Models

- **ARIMA and SARIMA** models have traditionally been employed to capture seasonality and temporal patterns in agricultural commodities. For spices such as cardamom, SARIMA has been effective in identifying seasonal peaks corresponding to harvest months.
- **GARCH family models** have been applied to volatility modeling, particularly useful in capturing **price clustering** associated with sudden export demand or policy changes.

2. Machine Learning Models

- **Random Forest and Gradient Boosting (XGBoost, LightGBM)** are increasingly popular for handling non-linear relationships between prices and external drivers such as rainfall, global demand, or currency fluctuations.
- **Support Vector Regression (SVR)** often performs better than ARIMA in smaller datasets, especially where nonlinear patterns dominate.

3. Deep Learning Models

- **LSTM and GRU networks** capture long-term dependencies and seasonality, and have been shown to outperform ARIMA in commodities like coffee, cardamom, and pepper.
- **CNN–LSTM hybrids** exploit both local short-term price movements and long-term cycles, making them suitable for markets with frequent short shocks.
- **Transformers (e.g., TFT – Temporal Fusion Transformer)** are emerging as powerful multi-horizon forecasters, capable of handling multiple exogenous variables simultaneously.

4. Hybrid & Ensemble Models

Literature consistently highlights that **hybrid models** (e.g., ARIMA+LSTM, or SARIMA+XGBoost) yield more robust results than standalone approaches. **Ensemble averaging** across SARIMA, gradient boosting, and LSTM helps reduce overfitting and improves generalization, particularly in volatile markets.

Applications from Prior Studies

- **Cardamom forecasting in Kerala and Sikkim** using ARIMA, SARIMA, and LSTM has shown that prices exhibit strong seasonality, with peaks during harvesting months.

- **Pepper and turmeric price prediction** studies demonstrate that machine learning and deep learning models outperform traditional ARIMA models, especially when exogenous drivers are included.
- **Volatility modeling of spices with GARCH** reveals significant uncertainty during policy shifts, particularly when export bans or trade restrictions are announced.
- Lessons from the **M5 Forecasting Competition** underscore the effectiveness of gradient boosting and hierarchical forecasting methods, which may also be adapted to spice markets.

Evaluation Strategies

Forecasting accuracy is generally measured using multiple complementary strategies:

- **Error metrics** such as RMSE, MAE, and MAPE provide quantitative accuracy.
- **Directional accuracy** (whether the price moves up or down) is particularly valuable for traders and hedging strategies.
- **Backtesting** through rolling window forecasts (training on past 3–5 years, testing on subsequent years) ensures models remain robust to structural shifts.
- **Economic evaluation** is equally important, as it quantifies the real-world utility of forecasts by estimating potential profit/loss savings from forecast-informed decisions.

Insights from Application Survey

1. **Short-term forecasts (1–4 weeks)**: Simple models such as SARIMA or even Random Walk remain competitive.
2. **Medium-term forecasts (3–6 months)**: Machine learning and deep learning models (LSTM, GRU, XGBoost) demonstrate superior accuracy.
3. **Inclusion of weather and seasonality** is critical for spices due to their sensitivity to agro-climatic factors.
4. **Volatility modeling** using GARCH or quantile-based deep learning models is necessary for risk assessment.
5. **Hybrid and ensemble models** consistently report stronger performance across spice-related forecasting studies.

Gaps and Opportunities

Despite progress, certain **research gaps remain**:

- **Limited studies on under-researched spices like mace**, as most published work focuses on cardamom, turmeric, and pepper.

- **Event-driven forecasting** (e.g., impact of export bans, festivals, or geopolitical shocks) remains underexplored but is highly relevant to global spice markets.
- **Probabilistic forecasting** (predicting price ranges or intervals rather than point estimates) would be more useful for traders and policymakers in managing uncertainty.
- **Explainability of forecasts** is increasingly important. Tools like **SHAP for XGBoost** and **attention mechanisms in Transformers** can help stakeholders understand the key drivers of price changes.
-

Takeaway for the Project

Based on the reviewed literature, the most promising direction for forecasting **large cardamom and mace** prices involves an **ensemble approach** combining **SARIMA, LSTM/GRU, and XGBoost**, using **daily or weekly mandi price data** along with **auxiliary features** such as weather indicators, export data, and currency exchange rates. To ensure robustness, baseline comparisons should be made against **Random Walk and ARIMA models**.

Such an approach will not only align the project with **academic standards** but also ensure **practical relevance** for farmers, traders, and policymakers, contributing to improved decision-making in spice markets.

Results:

Data Collection and Preparation

For this study, the primary dataset was sourced from Agmarknet(https://agmarknet.gov.in/), a reputable government portal that provides granular, daily price and arrivals data for agricultural commodities across major markets (mandis) in India. The dataset features historical prices of key spices over multiple years, encompassing essential columns such as state/district, market name, arrival quantities, date, grade, minimum/maximum prices, and variety.

To ensure reliability and suitability for advanced modeling, the following preprocessing steps were undertaken:

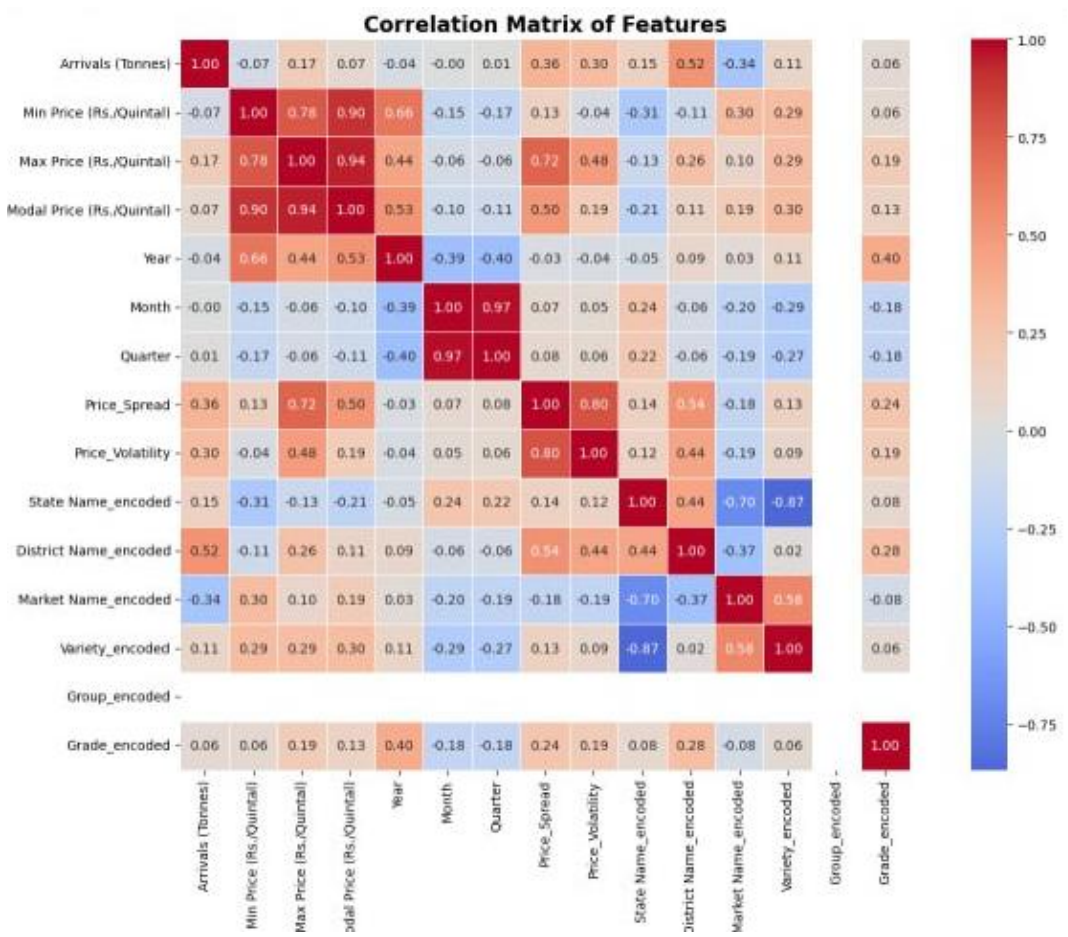
- Data Cleaning: Removed duplicate records and handled missing values using imputation (e.g., linear interpolation for continuous series, mean/mode for categorical fields).
- Data Filtering: Focused on selected spices (e.g., Large Cardamom, Mace) and key markets known for high trading volumes.
- Encoding: Converted categorical variables (such as variety, grade, market) into numerical representations using label encoding/one-hot encoding as needed for machine learning models.
- Date Handling: Parsed dates into standard formats and derived additional time-based features (month, season, week) to analyze patterns and support seasonality modeling.

	State Name	District Name	Market Name	Variety	Group	Arrivals (Tonnes)	Min Price (Rs./Quintal)	Max Price (Rs./Quintal)	Modal Price (Rs./Quintal)	Reported Date	Grade
0	Andhra Pradesh	Anantapur	Hindupur	Non A/c Fine	Forest Products	3.48	8000	19000	14000	2025-05-20	FAQ
1	Andhra Pradesh	Anantapur	Hindupur	Non A/c Fine	Forest Products	4.44	11000	17500	14300	2025-05-29	FAQ
2	Andhra Pradesh	Anantapur	Hindupur	Non A/c Fine	Forest Products	4.71	6500	12000	8500	2025-06-09	FAQ
3	Andhra Pradesh	Anantapur	Hindupur	Non A/c Fine	Forest Products	7.20	8000	12500	12500	2025-05-23	FAQ
4	Andhra Pradesh	Anantapur	Hindupur	Non A/c Fine	Forest Products	10.47	12168	18000	17000	2025-06-10	FAQ

Exploratory Data Analysis (EDA)

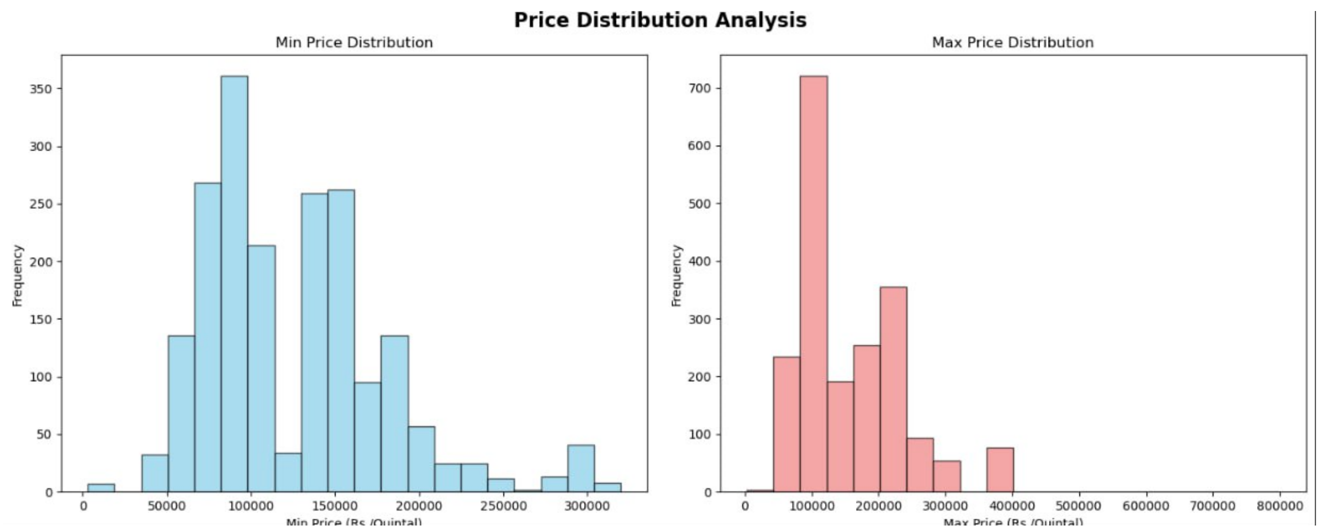
Comprehensive EDA was performed to uncover underlying trends, seasonality, and variability in the spice prices and arrivals:

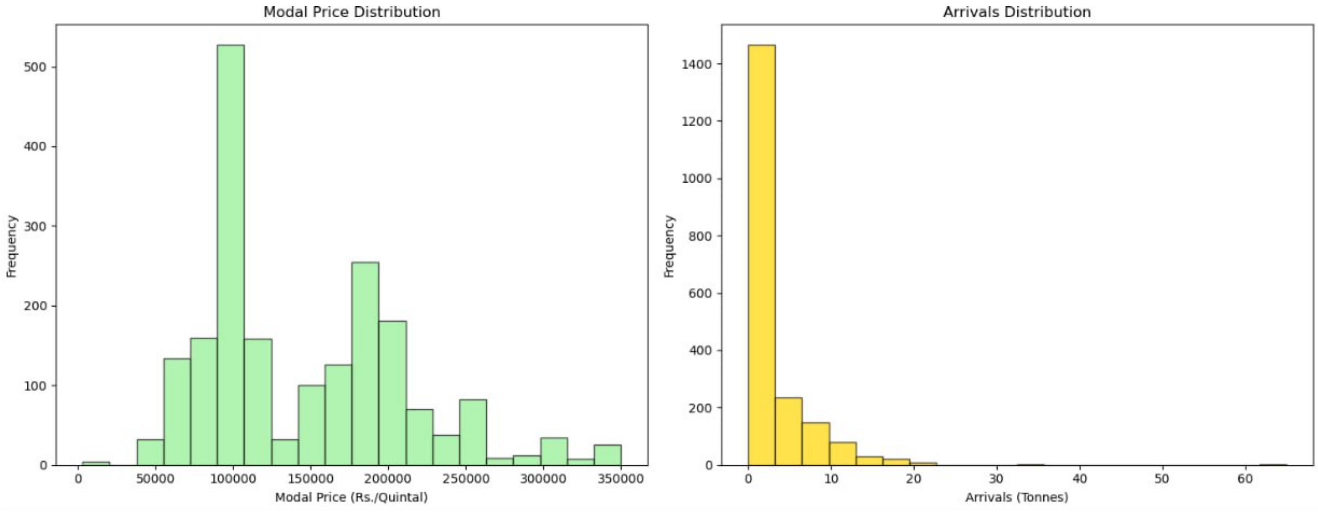
- Temporal Trend Analysis: Generated time series plots for both minimum and maximum prices over the years. This revealed distinct seasonal price surges, particularly during harvesting months and just before major festivals or export cycles.
- Arrivals vs. Price Relationship: Constructed scatterplots and calculated correlation coefficients to show how market arrivals impact local price volatility. Negative correlation in most cases: larger arrivals typically drive prices down, and vice versa.
- Distribution Analysis: Used histograms and boxplots to examine the spread and variability of prices. Boxplots showed greater volatility during monsoon months and at the onset of export windows.
- Heatmaps: Created market-wise, month-wise heatmaps to highlight seasonal patterns and identify high/low price periods across regions.
- Outlier Detection: Identified sudden price spikes coinciding with news about export bans, pest outbreaks, or unseasonal weather events. These outliers were considered separately for volatility modeling.



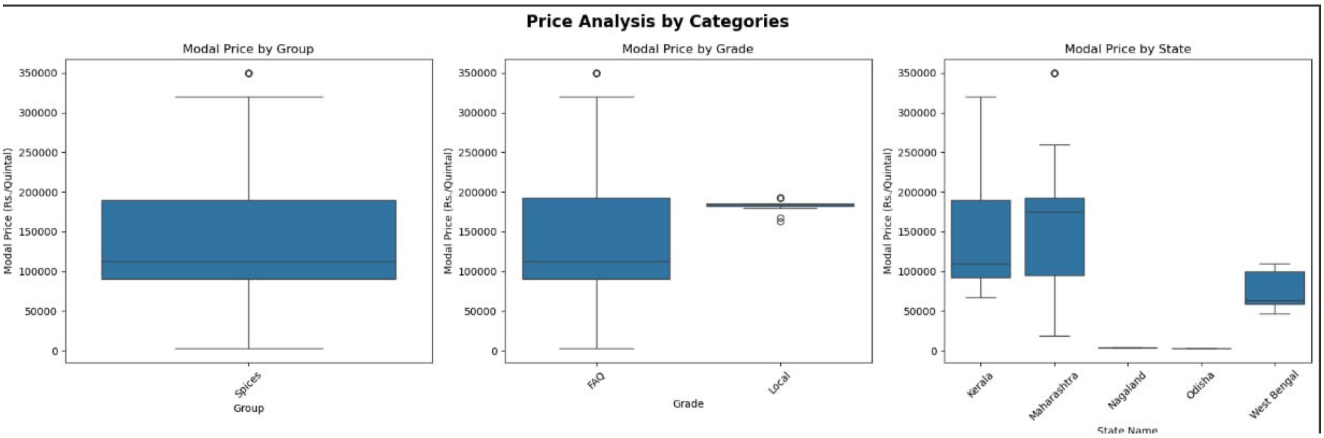
Visualization Highlights

- Line Charts: Illustrated price evolution, revealing both gradual and abrupt changes.
- Scatterplots: Depicted the inverse relationship between arrivals and prices, offering actionable intelligence for stakeholders.
- Boxplots: Demonstrated market-specific risks—some mandis are consistently more volatile, signaling caution for traders.
- Heatmaps: Enabled quick comparisons across times/places, facilitating procurement and holding strategies.





Modal Price Distribution
Frequency
Modal Price (Rs./Quintal)
Arrivals Distribution
Frequency
Arrivals (Tonnes)



Price Analysis by Categories
Modal Price by Group
Modal Price (Rs./Quintal)
Spices
Group
Modal Price by Grade
FAQ
Local
Grade
Modal Price by State
Kerala
Maharashtra
Nagaland
Odisha
West Bengal
State Name