Major Research Project

Comprehensive Analysis and Machine Learning Frameworks for Agricultural Price Stability in India

Submitted By:

Tarun Sarvang

24/DMBA/250

Under the Guidance of

Prof. P.K. Suri

DELHI SCHOOL OF MANAGEMENT

Delhi Technological University

Bawana Road Delhi-11004

DECLARATION

I, Tarun Sarvang, MBA (2024–2026) Student, Delhi School of Management, Delhi Technological University, do hereby certify that the Major Research Project titled:

"Comprehensive Analysis and Machine Learning Frameworks for Agricultural Price Stability in India"

is an original work submitted by me as partial fulfilment of the Master of Business Administration degree requirements.This project work has been accomplished by me, and the research findings included herein are based upon my own effort and analysis. To the best of my belief and knowledge, this work is not submitted for any other university or institution to any degree, diploma, or certificate.

Signature

Tarun Sarvang

Roll No: 24/DMBA/250

Date:

Place:New Delhi

CERTIFICATE

Mr Tarun Sarvang, Roll No. 24/DMBA/250 has submitted the Major research project ***"Comprehensive Analysis and Machine Learning Frameworks for Agricultural Price Stability in India"***

in partial fulfilment of the requirements for the award of the degree of Master of Business Administration (MBA) from the Delhi School of Management, Delhi Technological University, New Delhi during the academic year 2024-25.

(Prof. Pradeep Kumar Suri)

(Dr. Saurabh Agrawal)
Head (DSM)

ACKNOWLEDGEMENT

This research project would not have been completed without the guidance and support of many persons. To them I express my deepest gratitude.

I would like to especially thank my supervising professor, Dr. P.K. Suri from Delhi School of Management, Delhi Technological University, for the unlimited yet invaluable guidance, constructive feedback, and ongoing inspiration he provided during the entire process of this project. His knowledge and insightful advice have greatly contributed to developing this project in both its direction and quality.

Additionally, I am most grateful to the professors of the Delhi School of Management for the academic support provided to me and the opportunity of acquiring a theoretical and analytical basis from which to conduct this research.

Tarun Sarvang

24/DMBA/250

MBA (2024–2026)

Delhi School of Management, DTU

EXECUTIVE SUMMARY

Agriculture significantly contributes to the Indian economy as it provides millions of farmers with job security and helps ensure food security. However, agricultural commodity prices in India can be very unpredictable and unstable due to various factors, including crop seasonality, weather patterns, transportation issues, government policies, and the fluctuation of demand. Therefore, this makes it challenging for farmers, traders, and policymakers to make educated decisions about planting crops, how to store them, and when to sell them.

The project will create a predictive analytics framework to forecast the prices of agricultural commodities based on machine learning methods and time series analysis. Data from multiple states and mandis (markets) across India for various crops (such as wheat, rice, onions, potatoes and tomatoes) were collected and compiled into a dataset so that they could all be assessed and analysed together. This study has used an extensive data preprocessing pipeline to enhance the quality of the data by cleaning it, standardising it, correcting errors and inconsistencies, detecting outliers using Interquartile Ranges (IQR) and creating additional features.

To capture how prices behaved over time, several lagged (i.e., historical) and rolling statistical features were created, including lagged prices, rolling averages and rolling volatility measures. Once the final dataset had been created, several different predicting models (e.g., ARIMA, linear regression, random forest, gradient boosting and XGBoost) were developed to forecast future agricultural commodity prices. The predicting models were evaluated using mean absolute error (MAE), R2 and Pearson correlation coefficients (R) to determine how accurately they were able to predict and how reliable they were.

Overall, the experimental data indicated that machine learning techniques generally outperformed traditional statistical methods when applied to the data set chosen for this project. As such, of all methods tested, the Linear Regression technique produced the most accurate estimate of the future price of Wheat (Deshi) with the lowest average error as well as the greatest correlation between predicted and actual prices. Conversely, ARIMA was not able to effectively model the non-linear fluctuations in market price. In addition to the forecasting objectives of this project, prescriptive analytics were applied using the developed forecasting techniques through the use of market-wise price comparisons and sell-or-hold recommendations to assist agricultural producers/market participants in making more informed and practical decisions regarding the timing of their sales. Lastly, clustering analysis will allow for the

(All Multidisciplinary Journal,2024)(All Multidisciplinary Journal, n.d.)identification of "like" markets characterized by similar pricing patterns and volatility profiles.

This research illustrates how utilizing a data-driven forecasting system can lead to better decision making in agriculture by providing farmers, traders, and policy makers with valuable information about the future of agricultural markets in India. Intelligent agricultural advisory systems developed on this framework will enable better price stability, lower levels of uncertainty, and a more efficient functioning of agricultural markets within India.

Agmarket.net

Under Market Research and Information Network (MRIN) sub scheme of Integrated Scheme for Agricultural Marketing (ISAM) an ICT based Agricultural Marketing Information Network (AGMARKNET) was launched in March 2000 to link important agricultural produce markets spread all over the country and the State Agriculture Marketing Boards and Directorates. More than 4000 APMCs are on boarded on the AGMARKNET portal and more than 2700 APMCs are regularly reporting data on the portal. More than 450 commodities and 2000 varieties are covered under the scheme.

The Directorate of Marketing and Inspection, Ministry of Agriculture & Farmers Welfare, Govt of India is implementing the scheme in association with the State Agricultural Marketing Boards/Directorates and APMCs.

Objectives of the Scheme

To establish a nation-wide information network for speedy collection and dissemination of market information and data for its efficient and timely utilization.

To facilitate collection and dissemination of information related to better price realization and market access by the farmers. This would cover:

- Market related information such as market fee, market charges, costs, method of sale, payment, weighment, handling, market functionaries, development programmes, market laws, dispute settlement mechanism, composition of market committees, income and expenditure, etc.
- Price-related information such as minimum, maximum and modal prices of varieties and qualities transacted, total arrivals and dispatches with destination, marketing costs and margins, etc.
- Infrastructure related information comprising facilities and services available to the farmers with regard to storage and warehousing, cold storage, direct markets, grading, re-handling and

repacking etc.

- Market requirement related information covering accepted standards and grades, labeling, sanitary and phyto-sanitary requirements, pledge finance, marketing credit and new opportunities available in respect of better marketing.

To improve efficiency in agricultural marketing through regular training and education for traders/major specific farmers in their local language.

To provide assistance for marketing research to generate market information for its dissemination to farmers and other market functionaries at grass root level to create an ambience of good marketing practices in the country.

For information relating to the Schemes in respect of agricultural marketing implemented by the Govt. (Central and States) and their financial assistance. Once the farm produce is standardized and labeled, backed by quality certification, it can be directly offered for sale in spot exchange in national and international markets.((Directorate of Marketing and Inspection, 2026)

Table of Contents

1. Literature Review

There has been considerable change over the last decade in the field of agricultural price forecasting research in India. Most of the previous works focused on using simple time series forecasting techniques. With more availability of large data sets and available computational tools, researchers have been gradually moving away from traditional forecasting methods and utilizing advanced machine learning and deep learning methodologies. Today's research focuses on combining environmental, economic, seasonal and logistical factors to provide more insight into the determination and prediction of agricultural price levels. Additionally, the current body of literature pertaining to agricultural price forecasting can generally be classified into three broad categories: price volatility in agricultural commodities, the creation of forecasting models and the effects of external economic and policy-related factors on agricultural market prices.

1.1 Volatility and Seasonality in Perishable and Non-Perishable Commodities

(Choices Magazine, n.d.; Invade Agro Global)

The study majorly concentrates on research concerning instability of Indian agricultural commodity prices, especially in the case of perishable crops namely Tomato, Onion and Potato popularly known as ‘’TOP’’ commodities. These crops are significant in determination of inflation indicators such as Consumer Price Index (CPI) and Wholesale Price Index (WPI), their price changes have a direct impact on household expenditure. Compared with potato, many tomatoes during the month keep a very high price increase, which can be observed from studies conducted between 2014–2024, while the long-term fluctuations of fresh potatoes tend to be more intensive because they are not stored well and have higher post-harvest losses. Onion prices also exhibit strong seasonality with much of the price movement during supply-tight weeks - see figure 3 below. Monsoons, restrictions on transport and interruptions to supply chains all played a considerable role in how the season impacted fluctuations in agricultural production in India. Conversely, the price of wheat and rice—the two main non-perishable samplings—has been stable for many states because of government intervention to purchase large quantities of these goods from producers through Minimum Support Price (MSP).

While this is true, research indicates that some products are much more vulnerable to shifts beyond their control. The war in Ukraine has had a major impact on the price of wheat and rice due in part to rising fertilizer costs resulting from this conflict, as well as adverse weather patterns resulting from El Niño. Additionally, research has demonstrated that there are significant differences between how regional agricultural commodity markets behave depending on local supply and demand, availability of storage capacity, transportation capabilities, and seasonal fluctuations in local production across various states in India.

Overall, these findings indicate that there are no standard prices for most agricultural commodities in India, due to the different price behaviours exhibited by each type of commodity across each geographic region within India.

1.2 Technological Progression: From Statistical to Hybrid Models

(ResearchGate,2022;arXic. 2025)(Researchgate,n.f.; MDPI.n.d.)Arid Agriculture Research Trends in Prediction Strategies have gone through significant change over recent years. Statistical forecasting methods such as ARIMA were usually the preferred method for the earlier forecasting methodologies, due to their relative ease of interpretation. The following describes the function of ARIMA Models. ARIMA Models usually work well for time series data that display both long-term trend and seasonality as their data set is usually derived from a stable set of measurements taken over time. Seasonal ARIMA ((SARMA) models work better at forecasting than ARIMA models, based upon the periodic nature of the time series data that comprise most Agricultural Data Sets, as harvest seasons are a primary influence affecting price (they will always occur in a predictable sequence). Both ARIMA and SARMA perform poorly in periods where external factors are impacting prices (ex. a conflict might occur in remote nations that could delay product imports to the USA) will significantly shift the market for a period of time. As a result, researchers have moved to using machine learning techniques for example Random Forest, Gradient Boost, Support Vector Machine (SVM) and XGBoost to improve upon the predictive ability of ARIMA and SARMA, to be able to model non-linear associations between variables and also give researchers a scalable option to work from in the event that they are working with larger data sets. In particular, XGBoost has generated interest among all of these models because of its historical accuracy as a prediction tool, high level of performance in modeling and utilization for missing/sparse data. XGBoost would therefore be the preferred predictive model given all of the reasons stated above. Predictive performance of XGBoost was even much better than traditional regression models, with high R^2 metrics and significantly lower forecasting errors in several studies. From that perspective, we will also find many feature engineering techniques like lag variables, rolling averages and time indicators (to help the ML to understand price behavior in short-term / seasonal frequencies).(IJIRT, n.d.; IJSDR, 2025)

Once again deep learning methods have opened-up untold possibilities of agricultural forecasting. LSTM networks, a class of RNN is commonly used for sequential and time-dependent data due to their ability to retain long-term patterns and historical dependencies. Agricultural prices are driven by past market data and seasonal climatological data, among other things, which makes LSTM applicable to this forecasting problem. Multiple studies have shown LSTM-based models outperforming traditional statistical methods, especially on commodities with either extremely volatile or long seasonal cycles. The hybrid system like ARIMA-LSTM and ARIMA-XGBoost are designed by researchers with the essence of both statistical and ML approach. The significant improvement in forecasts accuracy comes from the hybridization of ARIMA with machine learning algorithms, where ARIMA captures linear trends as a pre-processing step and machine learning algorithms handles all non-linear variations.

1.3 External Determinants and Policy Implications

(IDEAS/RePEc, 2026 n.d)Among the major issues discussed in the literature is the impact of government policies and external economic shocks on agricultural price stability. India's agricultural markets are very sensitive to policy changes on exports, imports, subsidies and procurement. An example of this would be the ban on rice exports (2023) that was put in place in order to control inflation at home and ensure national food security. Although it provided some stabilisation to local markets, there were also global supply shortages and rising international rice prices, showing how much of an influence India has on global agricultural markets.

In particular, many have identified the increasing financialisation of agricultural markets as being a factor. Financial instruments (e.g., commodity derivatives, futures trading and ETFs) have introduced greater speculative activity within food markets. Therefore, investor behaviour and general macroeconomic trends affect the prices of commodities, in addition to the basic supply and demand dynamics. This increased volatility has resulted in increased frequency of sudden price fluctuations that are detrimental to both farmers and consumers.

Finally, additional studies create a case for including environmental and climatic variables within forecasting systems. Weather-related factors affecting agriculture and the global supply of agricultural products (e.g., rainfall, temperature and humidity, drought and/or extreme weather events) are inherent to agricultural production and the global supply of the commodity market. Current research shows meteorological data is essential to any high-quality forecasting system. Climate disruptors, once they occur, produce local shocks which quickly affect yields and the prices of agricultural commodities.In a general overview, the literature definitely shows a clear progress from elementary statistical forecasting approaches to more sophisticated data driven predictive systems. Historically, previous models were mainly based on historical price trends, however, more recent forecasting frameworks are starting to combine market behavioural insight with climatic conditions.

2. Problem Statement

(AGMARKNET, n.d.; Kaggle, n.d.)(Agmark,net, n.d.; Kaggle, n.d.; Kaggle, 2026)The most serious challenge that the Indian agricultural sector faces is uncertainty and instability in crop pricing. Prices of agricultural commodities often change quickly in response to seasonal production cycles, weather patterns, transport disruptions, market demand changes and government actions. This unpredictability will directly impact farmers income, market efficiency and consumer affordability. Since agriculture remains the mainstay of a significant section of the Indian population, fluctuation in agricultural prices is one of the most important economic and social issues that needs to be dealt with in rural India. Evidence from previous work indicates that such uncertainty is associated mainly with three structural problems in the agrarian marketing system.

Information Asymmetry and Accessibility:

One of the biggest hindrances of Indian agriculture is that small & marginal farmers do not have access to real time market intelligence. Even today, selling decisions by the majority of farmers are guided more by traditional experience, advice from a local trader or prevailing mandi prices and not by scientific price forecasts. The non-availability of information results in farmers often being in a rush to sell their crops directly after harvest, due to the high supply of goods in the market at that time, which means lower prices overall as well. Many of these farmers are limited in their opportunity to take advantage of price spikes due to two factors: primary storage infrastructures are typically not developed and/ or financial needs for cash require producers to sell immediately upon harvest. Consequently, the combination of these two factors means that they have lower profitability and contribute, over time, to long-term financially instability in their rural communities. Additionally, many consumers are affected because when there are sudden and large disruptions in the amount of food available on the market, severe food inflation prices can occur.

Limitation of the Use of Conventional Forecasting is a Shortcoming in Non-Linear Markets:

Another disadvantage to the use of traditional forecasting techniques is that many standard forecasting frameworks are still employed in agricultural business sectors. Instead of utilizing the current developments in machine learning based algorithms that characterize more sophisticated and complicated forecasting, many forecasting practitioners are still utilizing basic statistical time series models or manual regression analysis by examining trends with the assumption that the behaviour of the market is generally stable and linear. However, agricultural market behaviour is inherently variable and is dependent upon various external events, many of which (i.e. pest infestations, abnormal weather conditions, trucking union strikes, export restrictions, price fluctuations for fuel, and geo-political tensions) are not linear in nature and are highly unpredictable. The events in this case are beyond what conventional pricing models can accommodate; and the existing systems are unable to collectively handle the massive amounts of data (such as environmental variables, customer arrivals, seasonal patterns, and economic data) related to the sudden increase in price due to extreme market behaviour, resulting in very high uncertainty that greatly reduces the accuracy of many forecasts and thus obstructs the ability of farmers and policy-makers to make effective decisions.

The delay in institutional price setting also causes a lack of efficiency, making the problem worse and more complex.

The MSP system, which is designed to provide farmers with a safety net against the devastating effects of the marketplace and is based on past cultivation costs, past prices, and past trends in existing markets, serves some purpose in ensuring some measure of income security; however, it does not encourage farmers to think ahead in relation to anticipated consumer demand and climatic threats, or to variability in supply. This has resulted in a misalignment between the realities of the marketplace and the manner in which the government buys. In some situations, the procurement of too many goods creates a surplus and a place for the goods to be stored, whereas in other situations, the number of goods that are procured is not sufficient, thus causing market shortages and/or price spikes. These types of delays impact the ability of the agricultural sector to respond quickly; and consequently, limit its capacity to adapt to changes in all areas of the economy, and all changes in the environment. To sustain Indian agriculture, it needs a scalable, intelligent and data scientific forecasting framework that works on real-time data – so that the sector is not subjected to repeated market shocks and price instability. Therefore, there is a considerable demand for predictive models combining ML with time-series forecasting and real-time market intelligence for improved accuracy in predictions and ultimately better agricultural decision-making. This project aims to solve these problems by building an integrated framework for agricultural price forecasting which analyses past data, generates market patterns and gives recommendations based on actionable statistics that help farmers, traders and policy makers. To ensure that agriculture in India continues into perpetuity, it is important that a scalable data-science based forecasting framework exists that analyses real-time data, to reduce repeated shocks to the agricultural market as well as reduce price instability. This creates a very large requirement for predictive models that combine Machine Learning (ML) with time-based forecasting models, as well as real-time market Intelligence, to allow for much greater accuracy in predictions and therefore improving the decision-making process within the agricultural sector. This research project will develop an integrated price forecasting framework that will use historical data for analysis, develop market patterns, and provide actionable statistical information for the use of farmers, traders and policy makers.

3. Objectives

This project's primary objective is to create a precise, data-driven agriculture pricing prediction system using machine learning/forecasting algorithms. By researching historical data from Indian mandis (agriculture markets), we can predict the future prices of various agro-based commodities. In addition to helping to create accurate price forecasts, implementing machine learning will enable us to provide actionable insights to farmers, traders, and policymakers. To reach our goal, we will work on accomplishing the following:

Price Trend Analysis

Conduct exploratory data analysis on the Agriculture_price dataset to identify regional price differences, cycles associated with seasons/varieties of wheat, rice, and TOP, and determine whether or not these characteristics vary by geographical region while using historical price data.

Multi-Model Forecasting Framework Development and Validation:

Build and test 4 differing methods of developing a prediction model that will utilize machine-learning techniques to create a commodity class-specific model estimating future prices of wheat, rice, and TOP, as well as providing performance comparison based only on standard errors for all four models.(IJIRT, n.d.; IJSDR, 2025)

Create Lag-Based and Rolling Features for Feature Engineering:

Create lag-based and rolling window features (e.g., 7-day rolling mean, multi-day lag) to increase machine learning model response time to current vs. historical long-run equilibrium pricing.

To Evaluate and Compare Accuracy of Models:

The goal of the project is to assess model performance based on various statistical evaluation metrics (e.g., Mean Absolute Error, R^2, and the Pearson Correlation Coefficient). We use time-series cross-validation methods for an unbiased and robust evaluation of models. Comparison of models using the above methods can determine which modelling approach best suits the requirement for performing an agricultural price forecast.

Market clustering and volatility study:

As another goal of the project, we identify groups of similar agricultural markets based on pricing data using clustering methods (e.g., K-Means). We conduct both 12 and volatility analyses to create rolling means of the market to identify a stable (non-volatile) and unstable (volatile) market segment. By understanding how a market is structured, researchers can then focus their energies on identifying areas of the market with similar pricing behaviour.

4. Methodology and Data Preprocessing Pipeline

A data-centric approach is used to create "high quality" datasets by performing rigorous data cleaning and feature engineering to derive heterogeneous datasets from vehicle, trauma registration (TR), and literature sources for time series forecasting/regression throughout this project.(IJIRT, n.d.; IJSDR, 2025)

4.1 Data Acquisition and Structural Overview

Data on price data by mandi level for the following types of agricultural products; Wheat, Rice, Onion, Potato and Tomato. Various varieties of these crops have been sourced. Information on the auction of these particular items are also included including their prices within each district. Additionally, all other relevant product information such as State, District and other characteristics have been incorporated into the data set.

The authors of the previous study used primary data sets of historical prices and real time prices obtained from the various mandis within India. In order to understand how they could predict the maximum and minimum prices, they incorporated their model into the research analysis after mapping their existing data. They currently only predict modal prices; however expect that they will be able to possibly make some predictions about the maximum and minimum prices. Dataset also has information for each record including state name, district name, market name, commodity type + variety + grade along with minimum price for that transaction, maximum price and modal price along with date of transaction. The modal price, which is the most frequently traded market price, has been chosen as the primary target variable for forecasting analysis.

The process applied these tools to include NumPy, Scikit-learn, Statsmodels, XGBoost and Matplotlib. The Pandas library was also used for importing the data set into Python.

Table 4.1- Variables and Datatypes

Column Name	**Description**	**Data Type**
STATE	Indian state where the mandi is located.	String/Categorical
Market_Name	The specific marketplace reporting the price.	String/Categorical
Commodity	The agricultural product (e.g., Wheat, Onion).	String/Categorical
Modal_Price	The most frequent trading price for the day.	Numerical (Float)
Price_Date	The date of the transaction.	Datetime
Price_Spread	Derived feature (Max_Price-Min_Price)	Numerical (Float)

Source(Own analysis, Agriculture_Price_Dataset)

4.2 Data Cleaning and Preprocessing

Due to inconsistent records and missing information in agricultural datasets a preprocessing pipeline was applied before the model training. This included removing spaces and inconsistencies in formatting in the column names to avoid confusion; this is called cleaning. Standardizing all headers of text based fields (state names, district name, market names, commodity name and variety), by converting to title-case across the dataset.

Some known state name inconsistencies were fixed manually through mapping dictionaries. Uniformity was maintained across records; for example: old names like "ted;Orissa", were erradicated in favour of the new name "Odisha". Variable date fields were converted to datetime objects for further timeseries work and to get temporal features. The original dataset was then filtered to remove invalid dates, records with no price, problematic duplicate records or otherwise illogical or inconsistent pricing structures. This preprocessing step also guaranteed that the following price condition was still true: ***Min_Price≤Modal_Price≤Max_Price***

To have a consistent data set, records with zero or negative modal prices were removed.

4.3 Outlier Detection and Removal

The agricultural price datasets with extreme price peaks are mainly due to entries made in error or anomalies in the market that last for a very short period. To minimize the influence of these extreme values, outlier detection was done using the IQR method and the outlying entries eliminated. The first quartile (Q1), third quartile (Q3), and interquartile range were calculated using modal prices. Since all rows have to be accounted for, the outliers were considered as having data points beyond the acceptable range and were removed from the dataset. The following IQR formula was employed in the project:

IQR=Q3−Q1IQR = Q_3 - Q_1IQR=Q3−Q1

The range of valid observations was computed as follows:

Q1−1.5(IQR)≤X≤Q3+1.5(IQR)Q_1 - 1.5(IQR) \leq X \leq Q_3 + 1.5(IQR)Q1−1.5(IQR)≤X≤Q3+1.5(IQR)

This process established model stability and minimized the impact of extreme anomalies on prediction accuracy.

4.4 Feature Engineering and Time-Series Preparation

Feature engineering, as described in this section, transforms a univariate time series into a supervised learning problem for machine learning models.(IJIRT, n.d.; IJSDR, 2025)

The following examples illustrate how the techniques are applied at each step in creating a model from the original time series date.

1. **Lag Features**

Lagged values of the daily mean price were created to capture the autocorrelation of commodity prices for 1-day, 2-day, 3-day, 7-day, 14-day, and 30-day lags, which gives the model a series of independent features to consider.

2. **Rolling Statistics**

A rolling mean of the previous 7 days (e.g., Rolling_Mean_7) serves as an indicator of trend, which effectively filters out the daily noise that occurs through the filtering of the market's direction.

3. **Missing Value Imputation**

To prevent breaks in the continuous data stream necessary for training the model, the time series was forward-filled for all of the time series missing values where there were reporting gaps.

4.5 Evaluation of the Model

The models were tested with standard statistical measures such as Mean Absolute Error (MAE), R-square (R2), and Pearson Correlation Coefficient (R) to determine how well the models predicted the future market price, how much variance was explained, and how much associated relationship exists between the actual price and the predicted price.

The formulas used to evaluate the models are described:

Mean Absolute Error:

MAE=1n∑i=1n|yi−y^i|MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|MAE=n1∑i=1n|yi−y^i|

R-squared:

R2=1−∑(yi−y^i)2∑(yi−y¯)2R^2 = 1 - \frac{\sum (y_i-\hat{y}_i)^2}{\sum (y_i-\bar{y})^2}R2=1−∑(yi−y¯)2∑(yi−y^i)2

The models were compared based on these evaluation metrics to determine the most suitable forecasting approach for agricultural price prediction.

4.6 Prescriptive and Clustering Analysis

(All Multidisciplinary Journal,2024)(All Multidisciplinary Journal, n.d.)The project, apart from predictive modelling, also involved prescriptive analytics that suggested the best course of action. A methodology was developed to provide a simple decision-support mechanism: Sell or Hold, where the sell recommendation was generated accounting for the prices for future periods. Average modal price analysis was used to derive recommendations for market wise and state wise.

Finally, we applied K-Means clustering to group markets based on average pricing metrics and price dispersion behaviour. Market clusters delineated a space of not only similar pricing paths but of volatility. Clustering analysis was also used for the identification of parts and territorial cultivator behaviour.(All Multidisciplinary Journal,2024)(All Multidisciplinary Journal, n.d.)

4.7 Predictive Modelling and Evaluation Strategy

The project implements a diverse suite of models to handle the varying degrees of volatility in agricultural data [5]:

- **ARIMA (pmdarima):** The auto_arima function is utilized to optimize the p, d, q parameters. The model is configured to account for weekly seasonality ($m = 7$) and is particularly effective for stable, trend-based commodities like wheat.[6]
- **Tree-Based Ensemble Models:** XGBoost and Random Forest Regressors are deployed to capture non-linear relationships. These models are evaluated on their ability to minimize Mean Absolute Error (MAE) and maximize the R^2 score, which indicates the percentage of variance explained by the model.[3]
- **Evaluation Metrics:** Model performance is assessed using MAE, RMSE, and Pearson correlation (R). These metrics provide a multidimensional view of accuracy, from the average deviation in units to the strength of the linear relationship between predicted and actual values.[3]

5. Descriptive Analysis of Indian Agricultural Markets

A detailed examination of the Cleaned_Agri_Data.csv file presents critical location and variety specific dynamics to aid in our predictive modeling efforts.(ResearchGate, n.d.; MDPI. n.d.)(ResearchGate, n.d.; IRE Journals. n.d.)

5.1 Regional price variations

Regional differences in prices for various commodities throughout the country do not often reflect uniformity but rather reflect shocks to the supply of the product locally, or problems with transportation. An example for wheat is that there is a 1,200 to 1,100 ₹ difference in modal prices between Hoshangabad, Madhya Pradesh at ₹2,120 and Bangalore and Bidar, Karnataka at ₹3,300 respectively.6 The importance of distance from production hubs i.e. how far you have to transport the product from the producer to the final market are reflected in this price disparity.

Table 5.1 - State Wise Commodity and their Price Characteristics

State	Primary Commodities Reported	Price Characteristic
Maharashtra	Wheat, Onion, Tomato, Potato	High volume of trade in Nashik (Onion) and Pune.[6]
Uttar Pradesh	Potato, Onion, Rice, Wheat	Dominant reporter of local potato varieties like "Desi".[6]
Karnataka	Tomato, Onion, Potato	Exhibits extreme price spreads in tomato markets like Mulabagilu.[6]
West Bengal	Rice, Potato, Wheat	High historical prices for "Jyoti" and "Other" potato varieties.[12]
Kerala	Tomato, Onion, Potato	Frequent reporting of high price ceilings for perishables (e.g., Pothencode).[6]

Source(Own analysis, Agriculture_Price_Dataset)

5.2 Variety-Specific Analysis

This comprehensive dataset portrays many distinct types of agricultural commodities (i.e. wheat, onions, tomatoes) along with their specific characteristics regarding the market as well as their associated shelf lives that affect the price volatility indicators of each type.

Wheat Types - The most commonly used are Average, Lokwan, Sonalika and Maharashtra 2189. Although these types are classified as Fair Average Quality (FAQ), they exhibit unique pricing clusters within all regions of India; the pricing for Lokwan usually exceeds other types when sold in centralized locations (i.e. Central India).

Onion Types - The two most common types are Nasik; and Red; the latter are the dominant type for most of North and Central India. Additionally, prices in the international market influence onions exported from India, with Nasik onions being particularly susceptible on account of export policy developments.

Tomato Varieties - Early season volumes of tomatoes were typically made of Deshi varieties; but as hybrid types became popular (due to higher market prices), there were substantial differences in the pricing between the two varietal types (specifically in Karnataka).

6. Exploratory Data Analysis (EDA)

The analysis began with the execution of an Exploratory Data analysis (EDA), which allows the analyst to see what the distribution is, structure of the price market, and the characteristics of the respective commoditized processes so they can begin to develop predictive marketing models for future periods. In analysing the data, we found several key indicators to help determine seasonality, patterns, and their variation; as well as the price differences among different regions; and also determined what behaviours exist within those three different behaviours. Since predicting farm prices using only previous periods is a common occurrence, EDA is also essential as it will provide insights regarding the data being analysed as well as assist in choosing credible models with reasonable confidence.

6.1 Dataset Overview

After cleaning and pre-processing, we generated a dataset containing record sets from 2.9 million wholesale marketplaces throughout India, covering the period of time from 2023 through 2025 with submissions for each day listed from 2023 through 2025 (containing June 2023 through December 2025 submissions). The five crops we used for input factors included; Wheat, Rice, Onion, Potatoes, and Tomatoes, on overall representation of 27 of India's state by recording the input data from January to March 2023 through to December 2025.

The dataset has many attributes that are considered the key variables in predicting farm prices.

- State Name
- District Name
- Market Name
- Commodity
- Variety
- Minimum Price
- Maximum Price
- Modal Price
- Price Date

Features that were engineered such as Price Spread and Temporal Variables. The main target variable chosen was the modal price, which captures the most common market price from trading and depicts a better approach to many real usages compared to using minimum or maximum prices.

6.2 Commodity-Wise Analysis

The exploratory analysis indicates that there are significant differences in the pricing behavior of different commodities, which is due to the perishability of the commodity, seasonality, and the demand/supply trend of the commodity within the market place. For instance, onions and tomatoes exhibited higher volatility than both wheat and rice when looking at the other commodities and comparing their volatility. In addition, perishable commodities, such as onions and tomatoes, exhibited volatile price fluctuations as a result of storage limitations, transportation delays and/or supply and demand mismatches.

By contrast, staple goods such as wheat and rice exhibited a relatively stable and normal trend due to the presence of government procurement systems in place and the use of USDA-approved prices as a cost in the procurement process. However, there was still some price fluctuation in staple commodities (such as wheat and rice) when there were climate shocks and/or changes in agricultural policy.

6.3 Time-Series Price Trend Analysis

The use of time-series can also be used to show how agricultural prices change over time. A time-series analysis is conducted by visually analyzing average daily modal prices for each selected commodity over the period(s) for which they were produced. This is done by representing average daily modal price for each commodity on the Y-axis and daily transaction date on the X-axis. The results of the time-series analysis for the various commodities show that as time progresses, there are many cycles and times where the price of a commodity will increase rapidly and then drop back down.

For example, wheat (Deshi) would typically have a relatively consistent price throughout most months of the year but would fluctuate moderately within the seasonal transition months. The trend lines for the prices of these commodities also appear to be cyclical, these cyclical trends provide evidence that agricultural commodities in this market place exhibit a seasonal pattern of pricing.

The time series helped a lot with identifying the:

- Long-term price trend
- Seasonal cycles

- Sudden market shocks
- Stable and unstable periods
- Market recovery phases post volatility

6.4 Comparison by States of Prices

Regional variations in prices exist among the different states. To assess these regional differences, a state level comparison was performed. There was a large difference in price across various agricultural commodities for the states due to the following: transportation costs, level of production in local areas, strength of mandi infrastructure, availability of storage facilities, and patterns of demand in different regions.

Through the data collected from the price of wheat in the markets in Delhi and Gujarat, wheat prices (deshi) were compared with other regions to determine that deshi wheat was being purchased at the lowest average price possible. In addition to price variations, the price variations displayed by provinces with more considerable production capacity and greater market infrastructure also appeared more stable.

The results of the analysis provided evidence that prices for agricultural commodity products in India exist primarily on a local basis within a local economy.

6.5 Seasonality Examination

To study the seasonality of agricultural commodities, monthly average modal prices were calculated for each month (Barroca et al.). The results of the study confirmed that seasonal trends exist for agricultural commodity prices.

In particular, distinct peaks were noted for crops considered perishable (i.e., tomatoes and onions) due to scarcity and off-season supplies. Conversely, when examining the seasonality of crops with a longer shelf life such as wheat, the seasonal patterns were noted to be smoother because of the availability of government purchases.

The seasonality study showed that:

- Poor supply periods reflect in rising prices
- Prices usually drop during harvest seasons due to higher arrivals in the market
- Seasonal patterns are driven by climate
- Specific commodities display annual cycles

Knowing seasonality was important because it could impact forecasting accuracy and informed feature engineering.(IJIRT, n.d.; IJSDR, 2025)

6.6 Price Distribution Analysis

Histograms were used to conduct a histogram distribution analysis to assess the frequency and dispersion of the modal prices. The analysis showed that a large proportion of commodities were positively skewed; they had a large number of moderate-priced commodities compared to a small number of extremely priced commodities leading to long tails.

The datasets used for predictive modeling were skewed and included a few extremely high observations, so preprocessing was required to remove outliers using the IQR processes. Once this occurred, the distributions were more stable and appropriate for predictive modeling.(ResearchGate, n.d.; MDPI. 2024)(ResearchGate, n.d.; IRE Journals. n.d.)

6.7 Market-wise analysis

The project also identified which markets had the highest average modal price per kg. In terms of average modal prices for any individual year, the maximum average modal prices for Deshi wheat were among the highest that was reported in the dataset for Najafgarh, Bavla, and Kasrawad.

The analysis provided prescriptive recommendations to support selling into the best-priced markets, along with insights for other prescriptive recommendations in the later phases of the project.

6.8 Volatility Analysis

Using percentage price changes, we determined how volatile commodities were within a set timeframe. When comparing the volatility values of different types of wheat (local) to those of other products that tend to have smaller degrees of variation such as tomatoes and onions, wheat (local) exhibited significant volatility.

To further investigate long-term commodity trends and remove short-term price fluctuations that occurred as a result of noise in the data, we conducted a rolling mean analysis. A rolling 7-day mean was used in the analysis to observe larger trends in the overall market without being affected by the daily fluctuations associated with time series data.

Overall, through exploratory data analysis, we were able to identify both the structure of the market and the regional differences in the seasonally related behaviour of volatility in the markets, which were important when performing feature engineering, selecting a model, and interpreting results from predictive modelling during the course of this project.(IJIRT, n.d.; IJSDR, 2025)

7. Predictive Modeling and Model Performance Analysis

(ResearchGate, n.d.; IRE Journals. n.d.)(IRE Journals, n,d,; MDPI, 2024; ResearchGate, 2022)Using percentage price changes, we determined how volatile commodities were within a set timeframe. When comparing the volatility values of different types of wheat (local) to those of other products that tend to have smaller degrees of variation such as tomatoes and onions, wheat (local) exhibited significant volatility.

To further investigate long-term commodity trends and remove short-term price fluctuations that occurred as a result of noise in the data, we conducted a rolling mean analysis. A rolling 7-day mean was used in the analysis to observe larger trends in the overall market without being affected by the daily fluctuations associated with time series data.

Overall, through exploratory data analysis, we were able to identify both the structure of the market and the regional differences in the seasonally related behaviour of volatility in the markets, which were important when performing feature engineering, selecting a model, and interpreting results from predictive modelling during the course of this project.(IJIRT, n.d.; IJSDR, 2025)

7.1 ARIMA Forecasting Model

For forecasting, the project's primary statistical model was the ARIMA (AutoRegressive Integrated Moving Average) method. Due to its ability of identifying seasonal, cyclical, and trend patterns in data collected over distinct time intervals, ARIMA is currently one of the most frequently used statistical models for forecasting time series. The project utilized the functionality of auto_arima to perform the task of selecting the best possible parameter combinations to utilize in predicting agricultural prices. Additionally, the configuration of the ARIMA model included the weekly seasonality component so as to identify the short-term cycles that tend to occur in agricultural price fluctuations as a result of seasonal variation.

In addition to the use of the ARIMA model, the forecasting was conducted by the historical price of the selected commodity and variety over a specific period. In order to determine how well the ARIMA model was able to forecast prices, statistical metrics such as Mean Absolute Error (MAE), R2, and Pearson Correlation (r) were calculated using these data. The results of the study indicated that although the ARIMA model was able to find some levels of both average and trend price movements over the period; there was a notable lack of predictive accuracy for agricultural prices when faced with random spikes in the respective markets. This demonstrates one of the main weaknesses of most statistical methods that are used for forecasting agriculture, i.e. they generally function as linear and fixed systems despite the dynamic nature of the agricultural commodity markets and the additional uncertainty from unknown external influences to agricultural prices.

7.2 Machine Learning-Based Forecasting Models

This project implemented a number of machine learning algorithms to overcome the shortcomings of traditional methods of statistic forecasting. These included:

- Linear Regression
- Random Forest Regressor
- Gradient Boosting Regressor
- XGBoost Regressor

These models, in contrast to ARIMA, can take many engineered features at a time and learn non-linear relationships on the data. Feature engineering, where lag-based features, rolling statistics, and temporal variables were extracted from the data in advance of using them to train machine learning models.(IJIRT, n.d.; IJSDR, 2025)

In order to make the models learn about historical dependencies of agricultural prices, we implemented lag features that took into account the previous price from 1 day, 2 days, 3 days, 7 days, 14 days and even 30 days. Finally, it added rolling mean, rolling standard deviation, rolling min/max and median features to record the market behaviour and short time volatility of crypto. Seasonality was also introduced to help the models detect market patterns over time, blending specific entities based on month, day and day-of-the-week in the datasets.

7.3 Hyperparameter Tuning and Cross-Validation

Grid Search CV was used for hyperparameter tuning to improve the prediction accuracy and model reliability. Since the task was time-series forecasting, rather than using random train-test splitting Time Series Split cross-validation was used. This way, past data was always used to predict representations on future observations and that resulted in a model evaluation more representative of how it would be used in production or when the nature of the problem involves sequential data structures.

Each of the machine learning models was tested with different combinations of parameters. For example:

Random Forest was tuned with tree depth and number of estimators.

A few hyper-parameters were tuned for Gradient Boosting and XGBoost with respect to learning rate, maximum depth and estimator count.

Different configuration of the intercept was tested for Linear Regression.

The most successful combinations of parameters were selected automatically according to evaluation scores from cross-validation.

7.4 Model Evaluation Metrics

The forecasting models were assessed through various statistical performance metrics to provide an overall assessment of prediction accuracy.

Mean Absolute Error (MAE)

MAE=n1∑i=1n|yi−y^i|

R-squared (R^2)

R^2 measures how much variance in the actual prices is explained by the forecasting model. Higher values indicate better explanatory capability.

R2=1−∑(yi−y^i)2∑(yi−y¯)2R^2 = 1 - \frac{\sum (y_i-\hat{y}_i)^2}{\sum (y_i-\bar{y})^2}R2=1−∑(yi−y¯)2∑(yi−y^i)2

Pearson Correlation Coefficient (R)

The correlation coefficient calculates how strong the linear association is between actual and predicted values of price. Higher positive correlation and values near 1.

These metrics are providing a balanced check on the performance of the model and also measures prediction error and relationship strength between actuals versus forecast.

7.5 Model Performance Analysis

This project may generate the outcomes of prediction, solely for the specific commodity and variety used in execution, its variations (Wheat (Deshi)). So, it should not be considered as universal conclusions for all agricultural commodities. Agricultural products vary widely in perishability, shelf life, seasonality, market demand and vulnerability to external shocks. Hence, the correlation of forecasting models may be quite different by commodities and market regimes.

Of all the models we implemented, Linear Regression performed the best overall in this particular Wheat (Deshi) dataset. This model had the lowest MAE with the highest R 2 and correlation scores which indicates that it captured well the pricing trend of the selected dataset. The relatively good performance of Linear Regression shows that the Wheat (Deshi).

Their analysis also finds that there is no one-size-fits-all best model for predicting prices for all agricultural commodities. Products with finer granularity, such as tomatoes and onions, that are subject to sharp seasonality and rapid price shocks, may benefit more from advanced ensemble or deep learning models than from simpler regression-based approaches. Ideally, forecasting systems in agriculture would be commodity specific and flexible across market conditions, rather than a single highly generalised model.

Overall, the predictive analyses demonstrate that there are benefits and drawbacks to both approaches of machine learning and statistical forecasting. The results emphasise the importance of feature engineering, suitable preprocessing [$preprocessing$], time-series cross-validation [$time_series_cv$] and commodity-wise evaluation for building robust agricultural price forecasting systems.(IJIRT, n.d.; IJSDR, 2025)

8. Prescriptive Analysis and Decision Support System

In addition to providing forecasts for agricultural commodities, the second part of this project involved the use of prescriptive analytics to support decision-making by agricultural stakeholders. While predictive forecasting models provide predictions of future market prices, prescriptive analytics provides answers to an important practical question: What action can the agricultural stakeholders take today based on the forecasted price? The purpose of the prescriptive analytics portion of this project was to take the forecasted prices and recommend actions that agricultural producers and traders could take in order to make better decisions about selling their commodities.

8.1 Sell vs Hold Decision System

We designed a simple decision support system that will automatically generate a recommendation of either sell or hold based on the forecasted price. The decision support system uses the current observed price of a commodity in the market (or an index) and compares it with the forecasted price generated by the forecasting methodology that was used (e.g., ARIMA) to create the forecast.

The decision rule developed for the project is the following:

1. If the forecasted price for the future is greater than or less than the current price, then issue a hold recommendation.
2. If the forecasted price for the future is less than the current price, then issue a sell recommendation.

These decision rules were implemented into software code based on the forecast output from the ARIMA forecasting model. In providing an example of the application of this decision rule in the case of the Wheat (Deshi) market at the current time, the following applies:

- Current Market Price = 2,673.33
- Expected Forecasted Price in 7 hours = 2546.96
- Recommendation = Sell

The expected price was said to be lower than the observations made in the current market at this point in time, therefore, expecting future prices in the same type of market to remain stable if not decline. Selling the commodity during that period could limit the risk of further loss.

The recommendation system we developed clearly indicates how predictive analytics can be turned into a helpful tool for agricultural producers. By helping farmers determine when to sell their crops (whether now, or wait for a price increase), this tool could ultimately help improve the livelihood of farmers.

8.2 Market Recommendation Analysis

The second component of this project included research on the mode of the sales price by each region of South America. The objective of this part of the project was to identify potential selling opportunities for farmers and traders by grouping by the market name, finding the mode for each market and discussing the differences between them.

For example, wheat produced in the Deshi variety of wheat would be sold in the following open markets with the highest mode ($1.00 per bushel): Najafgarh, Bavla and Kasrawad.

The markets identified above have had the best success in selling wheat within the Deshi variety:

Market Name	**Average Modal Price**
Najafgarh	2824.61
Bavla	2755.33
Kasrawad	2540.00
Paatan	2375.82
Pratapgarh	2373.04

This analysis demonstrates that agricultural prices vary considerably across markets due to differences in transportation costs, local demand, mandi infrastructure, and regional production patterns. Such information can help farmers select better markets for selling their produce whenever transportation and logistics permit.

8.3 State-Wise Recommendation Analysis

State-level analysis was also conducted to identify regions where the selected commodity fetched higher average prices. The analysis grouped records according to state names and calculated average modal prices for each state.

For Wheat (Deshi), the following states showed relatively higher average prices:

State	**Average Modal Price**
Delhi	2824.61
Gujarat	2755.33
Uttar Pradesh	2373.04
Rajasthan	2320.31
Chhattisgarh	2307.98

The results indicate that agricultural pricing behavior differs significantly across states because of local production levels, storage facilities, transportation networks, and regional market demand. Such regional analysis can help policymakers and agricultural traders better understand market dynamics across different geographical areas.

8.4 Volatility and Trend-Based Decision Support

The project included an analysis of volatility and rolling means to study the market's stability and to characterize the long-term behaviour of the prices analysed. The measurement of volatility was determined based on the changes in the fluctuation of the data set over time in relation to price changes in the market. The change in price (% change) was used to represent the impact of inflation; thus, the change in price also shows the actual increase/decrease in the price of the same commodity. Volatility is always defined as the standard deviation of price movements over time. The higher the value, the greater the uncertainty associated with price movements or price regimes.

Wheat (Indian) was moderately volatile compared to more volatile commodities like onions and tomatoes when compared to levels of volatility between the two. Therefore, wheat was generally more stable compared to other commodities in the added years.

The 7-day rolling mean was calculated to eliminate short-term fluctuations from prices and to be able to identify long-term price trends more easily. The primary purpose of rolling averages is to reduce the daily noise in price movements in order to show a clearer picture of market trends. Trend analysis can be a helpful tool for farmers or traders, enabling them to buy and sell commodities in the short term.

8.5 Practical Significance of Prescriptive Analytics

The third stage of analysis: shows how agricultural forecasting systems may in the future be able to go beyond predictions only and offer decision-support capabilities. The project is not only aimed at price forecasting but also other actionable insights which can help stakeholders in practical agricultural decisions.

The proposed framework may have several practical applications.

- Helping farmers to select the better time to sell
- Spotting profitable states and sectors
- Reduce the loss of gains from selling early
- Market planning assistance for traders
- Help policymakers to monitor regional price movements
- Improve decision making in the agricultural supply chain

Our recommendation system is rudimentary, but it provides a basis for future intelligent advisory platforms that might couple weather forecasts, transport costs, storage availability and real market arrivals to produce higher level agricultural recommendations .

9. Clustering Analysis and Market Segmentation

(All Multidisciplinary Journal,2024)(All Multidisciplinary Journal, n.d.)(Machine learning prediction of agriculture produces for Indian Farmers using LSTM, 2026)In addition to forecasting and prescriptive analytics, the project included some clustering analysis to find commonalities and anomalies of agricultural markets based on their pricing behaviour. Of course, agricultural markets in India are neither uniform nor national; commodity prices depend on local supplies, transport networks and storage facilities, and consumer demand across regions of India economic conditions. Therefore, clustering techniques were applied to cluster markets with similar price features and to identify different market clusters in the whole dataset.

9.1 Purpose of Clustering Analysis

(All Multidisciplinary Journal,2024)(All Multidisciplinary Journal, n.d.)Cluster Analysis The Cluster analysis was focused on classifying Agricultural markets into different groups, according to their price patterns and volatility behaviour. Clustering is an unsupervised learning method (as opposed to supervised machine learning models which rely on predefined labels — i.e.

In this project we managed to use clustering analysis to:

- Identifying high-price and low-price market groups
- Studying market volatility patterns
- Understanding regional pricing similarities
- Detecting unusual or isolated market behavior
- Supporting market segmentation for decision-

While this analysis is not a forecasting tool, it serves to give stakeholders greater context and understanding around the relative behavior of different markets.

9.2 Feature Selection for Clustering

For market segmentation, four price-based features were summarized at the market level. Clustering Done using the Following Variables:

- Average Minimum Price
- Average Maximum Price
- Average Modal Price
- Average Price Spread

The price spread represented the maximum minus the minimum prices and was an indicator of variability and trading range in the market.

The features were organized in market-wise groups and transformed into a cohesive structured dataset market-level for the clustering analysis.(All Multidisciplinary Journal,2024)(All Multidisciplinary Journal, n.d.)

9.3 Data Standardization

Since clustering algorithms are sensitive to differences in feature scales, standardization technique (StandardScaler) was used for all clustering variables before applying the clustering model. The process of standardization made the variables comparable with other standard scores because it creates a common scale where mean equals zero and standard deviation equals to one. This prevented any variable with larger numerical magnitude from dominating the clustering process.

9.4 K-Means Clustering Algorithm

K-means clustering is a simple and efficient algorithm that has been used frequently in segmentation analysis because it is an uncomplicated, easy-to-implement technique for numerical datasets. K-Means is an unsupervised clustering algorithm that attempts to partition n observations into k clusters, in which each observation belongs to the cluster with the nearest mean.

The algorithm iteratively:

- Initializes cluster centroids
- Assignment of observations to the nearest centroid
- Updates centroid positions
- Repeats the process until convergence

The Elbow Method is used to estimate the most appropriate number of clusters for this project. WCSS was computed for different cluster counts and the “elbow point” in the curve to identify a good number of clusters. The three clusters were chosen from summary of the analysis for market segmentation.

9.5 Clustering Results

The clustering analysis was performed on the selected Wheat (Deshi) dataset. The markets were divided into three clusters based on average price characteristics and price spread behavior.(All Multidisciplinary Journal,2024)(All Multidisciplinary Journal, n.d.)

Cluster 0 – Moderate Price and Stable Markets

Cluster 0 contained markets with relatively moderate average prices and controlled price spreads. Markets in this cluster displayed comparatively stable pricing behavior and lower volatility.

Some markets included in this cluster were:

- Ajaygarh
- Durg
- Kareli
- Kasrawad
- Rajnandgaon
- Seoni
- Tonk

The average characteristics of this cluster were:

- Average Modal Price: 2298.33
- Average Price Spread: 131.42

These markets may represent relatively balanced agricultural trading environments with moderate fluctuations and stable market activity.

Cluster 1 – Isolated Low-Price Market

Cluster 1 contained only a single market, Naila, which showed unusually low pricing behavior compared to all other markets in the dataset.

The average characteristics of this cluster were:

- Average Modal Price: 851.00
- Average Price Spread: 0.00

This isolated clustering behavior may indicate:

- Abnormal reporting patterns
- Limited trading activity
- Data irregularities
- Extremely localized market conditions

Such isolated clusters are important because they help identify outliers or markets requiring further investigation.

Cluster 2 – High-Price and High-Spread Markets

Cluster 2 included markets with comparatively high modal prices and larger price spreads, indicating more profitable but potentially more volatile trading conditions.

Markets in this cluster included:

- Bavla
- Najafgarh
- Pratapgarh

The average characteristics of this cluster were:

- Average Modal Price: 2650.99
- Average Price Spread: 363.81

These markets demonstrated stronger price potential but also higher variability in pricing behavior. Such markets may offer better selling opportunities during favorable market conditions but could also involve greater uncertainty.

9.6 Interpretation of Clustering Results

The study of clustering analysis has established that different types of agricultural markets exist in India based on their respective behaviours, prices and volatility of price action. Some agricultural markets are very stable and therefore the price technicals indicate how much potential exists within those markets. On the other hand some less stable agricultural markets will be classified as "Boom/Bust" markets where at least twice a month there has been dramatic fluctuations of sales and/or volume of the traded crops.(All Multidisciplinary Journal,2024)(All Multidisciplinary Journal, n.d.)

The following conclusions are made from our cluster analysis:

- They illustrate the habitual characteristics of buying across the area.
- Most of the ESA (type) prices tend to have higher levels of volatility than they do in a high value (price) area.
- Some less stable (non-tested), clusters require additional analysis

9.7 Importance of Clustering in Agricultural Analytics

These clusters can be used for helping buyers develop strategies to select and develop appropriate purchasing decisions for whatever agricultural products are traded.

Using the generated predictive models from the analysis, clustering analysis allows farmers to study the geographical location of similar type land by the estimation of the variariables relevant to that area. The benefits of using the clustering analysis for forecasting agricultural products insufficiently delineates specific potential market structures that are hidden closets from the statistical forecast methods that are conducted in a one-on-one basis. Forecasting techniques would allow us to predict what future pricing would be, whereas clustering analysis may further clarify the organizational structure and the behaviour so that appropriate actions can be taken into the marketing.(All Multidisciplinary Journal,2024)(All Multidisciplinary Journal, n.d.)

10. Socio-Economic Impact and Volatility Drivers

(Choices Magazine, n.d. Invade Agro, n.d., 2026)Forecasting/Prescription analytics and clustering will create a more shipment and market research plan for avocado stake holders in the future. Adding to in future implementations of additional enhancements could be developed around the presentational column models:

- Weather conditions
- The quantities at which goods arrive at the market
- Transportation costs
- New event types discovering
- Storage infrastructure data
- Farmer production capacity
- Regional consumption patterns

In summary, the research accomplished its goal of examining possible clustering trends within agricultural markets regarding price characteristics; therefore, improving our understanding and providing a more knowledgeable decision-making process in regards to agriculture.

10.1 Impact of Price Volatility on Farmers

Agricultural price instability causes problems for both producers and purchasers in India. Since agriculture provides the main source of income/livelihood for a large percentage of the citizens in India, fluctuations in crop value could trigger economic distress. These fluctuations affect the profitability and sustainability of Indian farmers, the consumer's ability to purchase food, and contribute to the stability in the overall marketplace and inflation rates nation-wide. The study conducted as part of this project identifies that agricultural price movement is not only influenced by prior market changes, but also by various social-economic and environmental variables.

When agricultural prices are erratic, it is farmers who suffer the most. Often, farmers put a lot of money into growing an agricultural product — buy seeds and fertilizers irrigating the crops at a transport stir it up without knowing how much they will get for product at harvest season. In case prices suddenly fall in a glut, or poor market mechanism compel the farmer to sell whatever he has below production cost as it becomes difficult to store with trees and bushes yielding lower quantities of produce.

There is shortage of this problem manifold for the small and marginal farmers whose are normally lacking in proper storage facilities for cold storage. The last of the three major success factors:

- Market intelligence access
- Financial backup
- Transportation flexibility

Consequently, a number of farmers sell their harvest as soon as it is harvested, during times with the most arrivals in mandi, usually associated with cheaper prices. This squeezes profits and makes rural communities financially unstable over the long run.

10.2 Consumer Impact and Food Inflation

Price volatility can adversely affect consumers as well in the case of key food items such as onion, tomato, potato, wheat and rice. Spikes in prices is likely to have a very direct impact of household outlays and overall food inflation. Because TOP commodities constitute a large share of daily consumption in India, even small short-term price spikes can exert significant overall economic pressure on consumers.

For example:

- Tomatoes prices frequently spike fiercely through interventions with deliveries due to the storms
- Retail food inflation can be quickly amplified by onion shortages
- Seasonal price shocks in potatoes arising from shortages of storage

High income households face these fluctuations, but low-income households are affected to a greater degree as they spend a larger share of their income on food consumption. Stable agricultural pricing thus serves not just the farmers but helps maintain larger macroeconomic equilibrium.

10.3 Climatic and Environmental Factors

Climate is an important determinant of agricultural production and market supply. Weather parameters like rainfall, temperature, drought, floods, humidity and heatwaves have high impact on agricultural commodities. Even small climatic disturbances have a strong impact on crop yields and transport systems.

Major climate related factors affecting agricultural prices:

- Relative weakness in the monsoon or they start later than expected
- Getting rainfall out of the harvesting season
- Drought conditions
- Flooding and waterlogging ·
- Increased temperature and heat stress

These disruptions tend to diminish supply in markets, resulting in unexpected price hikes. Tomato and onion being perishable items are more vulnerable than any other commodity specifically because they have shorter shelf lives and a greater risk to hold from the hours of delay in transportation, storage.

10.4 Transportation and Supply Chain Disruptions

Prices of agricultural products are also strongly affected by logistic systems and transport. Transportation delays, rising fuel costs, road blockages, shortage of labour or storage will have an immediate impact on carrying the crops from farms to mandis.

This leads us to the fact that many agricultural markets in India still have:

- Inefficient storage infrastructure
- Insufficient cold chain infrastructure
- Transportation wastes resources silly
- Significant post-harvest losses

Such problems are more serious for perishable commodities, as these take time to transport and otherwise lead to artificial urban shortages with consequential quality losses. This will result in a significant price increase even if global production is above the level.

10.5 Government Policies and Market Interventions

Indian agricultural market behavior flows from a lot of government influence. Available measures for government intervention to stabilize prices and safeguard farmers and consumers are Minimum Support Price (MSP), export restrictions, procurement policies, subsidies and stock limits. Nonetheless, policies can also have unintended consequences on markets. For example:

- Potential domestic inflation relief and local commodity oversupply from export bans
- Procurement on a large-scale increases storage and warehousing pressure
- Further delays in policy decision may keep the uncertainty in markets

India is a significant player in global agricultural trade, and rice export bans highlighted how domestic agricultural policies can extend well beyond geographic borders.

10.6 Financialization of Agricultural Markets

One of the key components discussed in modern agronomic research is financialization of commodity markets. Agricultural commodities are no longer shaped by supply and demand in the physical world only, but also financial maneuvers:

- Commodity Future Trading
- ETFs (Exchange-Traded Funds)
- Speculative investing
- Trends in the Global Financial Market

The increasing role of financial investors can lead to instability in markets, since this can amplify price movements over the short term which do not necessarily correspond to market fundamentals for agriculture. This also adds another layer of ambiguity for farmers and domestic merchants who relied on in-person trades.

10.7 Predictive analytics to resolve Market Uncertainty

Overall, this project demonstrated the potential for predictive analytics and machine learning to mitigate some of the uncertainties around agricultural markets. Forecasting systems include early indicators datapoints of:

- Expected price trends
- Market seasonality
- High-risk jumping shocks
- Greater opportunity to sell

This kind of system can help farmers to take better decisions on:

- Crop planning
- Harvest timing
- Storage management
- Market selection
- Selling strategy

Further, these predictive systems can help policymakers in formulating procurement planning improving the monitoring of inflationary risk and managing the operation supply chain better.

10.8 Overall Impact

The socio-economic analysis points out that forecasting the price of agricultural products is a technical problem, but it is also an important economic and social issue. Key factors for stable and predictably agricultural markets:

- Income protection of farmers
- Keeping food prices low
- Oxygen, increasing the pressure of inflation
- There is an improvement in the supply chain process.
- Providing a contribution to national food security

Thus, the integration of machine learning, forecasting systems and decision-support analytics in agricultural markets holds the promise of providing sustained benefits to producers and consumers alike while enhancing the resiliency of the sector as a whole.

11. Strategic Framework and Future Scope

he findings of this study show that there is a lot of potential for better decision making in the Indian agricultural market by using such price forecasting systems. But for these models to make a real-world difference, they also need to be linked to scalable digital infrastructure, real-time data networks and advisory systems that farmers can trust. We then suggest a strategic model based on the findings of this study to enhance agricultural market intelligence and advanced forecasting ecosystems in the future.

11.1 Development of Real Time Agricultural Advisory System

The main future application of this project is real time android application for Farmer & Trader advisory. This system proved that machine learning models can be used to predict pricing trend and give recommendation on market using historical data of mandi. These forecasting capabilities can be embedded in mobile-assist applications or web-based dashboards that provide farmers:

- Daily price forecasts
- Sell-or-hold recommendations
- Price compared to other markets
- Volatility alerts
- Head over to seasonal trends analysis

These systems can close the information asymmetry between farmers and mega market participants by providing predictive insights at the bottom of the pyramid.

11.2 Combining Weather and Environmental Data

The current project is mainly based on forecasting by taking the price history. Agricultural markets are very sensitive to environmental factors such as rain, temperature, humidity, drought and extreme weather events. The combination will make forecasting systems of the future more accurate and robust:

- Real-time weather data
- Rainfall patterns

- Soil Moisture Data
- The change of temperature
- Climate Risk Indicators

Now incorporating environmental variables would help forecasting models capture climate-driven supply shocks and sudden production disruptions in a far better way. Such an integration would prove beneficial, especially for temperature sensitive perishables such as tomato and onion.

11.3 Integration of Market Arrival & Supply Chain Data

Future versions of the forecasting framework can also incorporate mandi arrival quantities and supply chain indicators. Agricultural prices are no exception. Not only does yield history influence agricultural prices, but also:

- Daily arrivals in the market• Transportation costs •
- Fuel costs
- Availability of storage
- Cold chain facility
- Regional demand drivers

Together, these variables will increase predictive accuracy and provide valuable context to explain what moves the market. Supply chain analytics can also be used to identify bottlenecks in the supply chain that lead to artificially induced shortages and sudden price increases.

11.4 Expansion Toward Deep Learning and Hybrid Models

This project employed ARIMA as well as various machine learning models. Further research should investigate more advanced deep learning architectures such as:

- Long Short Term Memory (LSTM)
- Gated Recurrent Unit (GRU)
- Transformer based forecast models
- ARIMA-LSTM hybrid models

The following step is hybrid forecasting models that combine the two statistical and machine learning approaches in a manner that their advantages cancel out. Statistical models are capable of identifying broad linear trends over extended periods, while deep learning algorithms can recognise non-linear variations and complex temporal correlations.

11.5 Commodity-Specific Forecasting Frameworks

The findings from this project show that commodities display heterogeneity in their pricing impact and volatility. Hence, the further step for future forecasting systems is to be directed towards modeling approach according to specific types of commodities rather than universal forecasting model.

For example:

- Wheat and rice being more stable in nature are likely to be better predicted using linear / statistical models.
- Advance machine learning or deep learning approaches may have to be used for tomato and onion due to their perishability nature, volatility

Designing commodity-specific forecasting pipelines will enhance the accuracy and applicability of your predictions.

11.6 Integration with Government Policy Systems

Predictive analytics can also play a crucial role in helping policy managers enhance their agricultural planning as well as the market intervention strategies. Predictive systems can help agencies analyze and predict:

- Observing the risk of inflation
- Scheduling supply chain processes
- Administration of stock buffers
- Spotting gaps in the supply chain
- Targeted subsidy designing

By applying predictive analytics to decisions related to MSP, a more proactive and data-driven regime may be developed for agricultural pricing policies rather than one based solely on past production cost experience.

11.7 Smart Decision-Support Systems for Farmers

This project implements a simple sell or hold recommendation system. This will be the foundation to build on an intelligent advise platform that can offer;

- Crop Planning Recommendations
- Optimal harvest timing
- Recommendations for how long to keep them
- Increased efficiency in transportation routes
- Selling in multiple markets

Chatbot systems based on AI likewise would help integrate these systems also with farmers without extensive technical knowledge or digital literacy.

11.8 Scalability and Deployment Opportunities

This project presents a scalable forecasting framework. If integrated with a cloud and automated data pipelines, the system can be extended to:

- All India mandis – Real-time forecasting
- National commodity dashboards
- Monitoring systems for agriculture at the state level
- Farmer advice apps for mobile devices

With the recent growth in the availability of agricultural data and digital infrastructure driven by pioneering initiatives in India, we find enormous opportunities for extending existing AI-driven agricultural intelligence systems across geographies.

11.9 Future Research Opportunities

Notwithstanding the fact that the project has led to non-trivial forecasting and analytical results, some interesting prospects are left for future research:

- Streaming data integration in real-time• Integration of remote sensing and satellite data
- Deep learning based volatility prediction
- Agriculture news and social media sentiment analysis
- Multi-step forecasting will be useful for long-term planning of agriculture
- Reinforcement learning-based dynamic pricing

Future work could also test forecasting systems under very different market conditions, for instance, droughts, pandemic or international trade interruptions.

In summary, the strategic framework for development of this study suggests that predictive analytics, machine learning and intelligent decision support systems are becoming much more important in today's agriculture. In the long-term, forecasting models and real-time market intelligence, combined with scalable digital infrastructure can transform agricultural systems into resilient, economically efficient and resource-sustainable systems.

12. Conclusion

The focus of this project is to explore predictive analytics and machine learning techniques to predict agricultural prices at the Mandi level in India using Mandi-level data. The objectives of this stream of work include understanding the operation of agricultural markets, developing a process for cleaning large datasets of major commodities, forecasting based on historical dynamics and developing actionable decision support information systems for farmers, traders and policymakers to use.

According to our most recent research, some of the main factors influencing agricultural commodity prices include: seasonality, weather-related disruptions to transportation, regional climate conditions, and government level intervention through price controls or trade restrictions. Additionally, the research indicates that pricing behavior of perishable items (i.e. tomatoes, onions and potatoes) tends to be much more volatile compared to wheat and rice, which have a lot of land area specified in agricultural policy for procurement and storage support systems.

An extensive automated pipeline for pre-processing and feature engineering was implemented to enhance data quality and improve the forecasting. This pipeline used procedures such as text standardization; removal of duplicate records; outlier detection based on Interquartile Range (IQR); lagged feature generation; rolling statistical analysis; and temporal feature extraction, to make agronomic data suitable for predictive modeling. The procedures used in preprocessing and feature engineering increase the robustness of the resulting models, while also accounting for data irregularities (i.e. short- term fluctuations) at the Mandi level.(IJIRT, n.d.; IJSDR, 2025)(ResearchGate, n.d.; IRE Journals. n.d.)(ResearchGate, n.d.; MDPI. 2024)

Table 1 outlines a number of different forecasting models tested and evaluated: ARIMA, Linear Regression, Random Forests, Gradient Boosting, XGBoost.....In general, the evidence shows it matters MORE how you have chosen to evaluate forecast accuracy, with respect to the type of commodity you are forecasting, and your chosen methodology. For example; with the wheat (various) analysis performed, it was concluded that on a general basis, the use of linear regression to analyse this dataset was much more successful than using ARIMA. This suggests that prices were stable over the period being analysed, and that high frequency noise within the price series was a limiting factor when using the ARIMA method. Therefore this provides evidence that no model does the best job in forecasting all agricultural commodity prices, but rather than change the model to fit the historical data, it is generally best to adapt your forecast methodologies to fit your commodity fundamentals based upon changing market conditions or fundamentals.

Another component of this research was to use prescriptive analytics for identifying actionable insight, as well as clustering what would become market insight (expected price into buy/hold recommendation). In addition, each state's relative comparison for the potential to improve pricing over time was also captured within the report, which would provide some insight as to which states would have the greatest opportunity for price increases. The use of cluster analysis also provided an opportunity to more accurately identify regional agricultural market segments based upon price behaviour and volatility behaviour; therefore providing a more accurate depiction of agricultural market segment characteristics.

This report highlights trends that show how critical machine learning and data-driven decision-making systems are becoming for agriculture in an evolving climate. Forecasting systems can guide farmers' futures by helping them select proper timing for selling their crops and predicting future prices with more precision, thus reducing their risk of loss due to price volatility. Forecasting systems can also support policy makers in procurement and improve market efficiency while also facilitating better price discovery. While somewhat limited due to only using historical prices and not including any external variables, forecasting systems can still serve as a strong foundation for future, more advanced, agricultural advising systems.

The development of this document illustrates that there are many areas within the agricultural and market intelligence sectors where predictive analytics, machine learning, and decision-support systems have, and will continue to have, numerous uses in the context of agriculture in India. In conjunction with existing crop forecasting systems, near real-time weather data, modern supply chain analysis, mature deep learning models, and comprehensive large-scale digital platforms can be utilized to bolster confidence in the agricultural marketplace and improve food security and the quality of life for farming communities.

13. Limitations of the Study

A project to apply machine learning or predictive analytics to agricultural price forecasting successfully showed the predictions made above, however limitations were noted with the approach taken in both development and evaluation process. Such constraints can limit the generalizability and applicability of the proposed forecasting framework in real-world settings, thus opening avenues for improvement in future work.

Dependence on Historical Price Data

As far as the forecasting models that were derived in this work . They are mostly based on past mandi price data . Seasonality provides some clue to the commodity markets from historical trends but agricultural markets are also heavily dependent on unpredictable factors in a constantly changing real-world scenario such as weather shocks, transport disruptions, geopolitical events and government interventions. And because these external factors were not fully mapped in the current setup, extreme market-side shocks or volatility may cause traditional forecasting models to perform poorly.

Limited External Variables

The existing methodology thus primarily dealt with price and time-based aspects like lag values, temporal indicators, rolling averages. Important external variables such as:

- Rainfall and Weather
- Quantities of arrivals on the market
- Fuel prices
- Transportation costs
- Fertilizer prices
- Global commodity prices
- Changes in the policies with respects to Export and import

the ones that were not directly part of the meta-modeling process Since the best predictor of future agricultural prices is its current price and in very dynamic market conditions, external conditions tend to dominate sustainable long run forecasting accuracy, these variables may help predictive power but remove the need for higher frequency data.

Commodity-Specific Model Performance

They were determined by commodity variety on which the project forecasting results were returned. Agricultural commodities are less volatile, seasonal and perishability. This project has illustrated that a model finding success with one commodity does not produce the same results on another.

For example:

- Linear Regression worked well on the selected Wheat (Deshi) dataset as prices did not fluctuate that much.
- Non-linear, or deep learning approaches may be required for more volatile commodities like tomato or onion.

Thus, the effect cannot simply be transferred across all agricultural commodities without commodity-wise appraisal.

Limited Deep Learning Implementation

The discourse in the literature review mentioned more advanced techniques using deep learning such as LSTM and hybrid forecasting methods, however, this project only practicalised basic statistical and traditional machine learning algorithms. Deep learning models generally require:

- More computational resources
- Longer training time
- Larger parameter tuning
- Large sequential datasets

These models were not implemented in the current scope of work due to time and computational limits.

Data Quality and Reporting Issues

Most of the datasets on agricultural mandi have errors as this data is manually collected and reported. Even after preprocessing and cleaning, there is still a possibility of some potential limitations with our dataset:

- Missing records
- Delayed reporting
- Human error of entering the data
- The irregularity of regional naming
- Inconsistent Q1-Q4 report coverage of markets

While techniques such as means for removing outliers and standardization enhanced data quality, the forecasting system still relies heavily on the credibility of original data source.

Short-Term Forecasting Focus

The forecasting framework implemented was mainly designed for long-term agricultural market predictions. Long-term forecasting is much more difficult in agriculture as the markets change over time due to climate change, policy reforms, technological progress and macroeconomic variables.

As a result:

- The models can become less accurate at longer forecasting horizons.
- Unanticipated structural shifts in market behavior may not be reflected accurately.
- Model retraining may be necessitated periodically for ensuring reliability.

Simplified Decision-Support Logic

In the project we implemented a simple rule-based sell-or-hold recommendation system based on predicted future prices. This is an example of how predictive analytics works in practice, but agricultural decision-making involves many more factors in the real world such as:

- Storage costs
- Travel cost
- Crop perishability
- Liquidity needs of farmers

- Risk tolerance
- Domestic demand circumstance

So, the recommendation system at this moment is a good wholesale prototype but not yet a full scale production advisory platform.

Computational and Resource Constraints

There were some high-level modeling techniques and large-scale experiments that we could not perform from the limitation of computation resources and execution time. For example, XGBoost, Gradient Boosting and cross-validation procedures need considerable amount of time to run to get optimal model in every 5 folds during hyperparameter tuning with multiple hyperparameters on large datasets. More sophisticated experimentation may help to further improve performance, given the resources.

Limited Geographic and Temporal Scope

While this dataset had a large breadth with multiple states and commodities it was ultimately limited to the available mandi records for that state over time. Depending on local economic and climatic conditions, agricultural market behaviour may differ considerably across regions and years. Consequently, larger datasets covering longer periods in history may give greater forecasting powers.

Overall Limitation Summary

However, despite these limitations, the project has provided us with some insight into the potential of machine learning and predictive analytics for agricultural price forecasting and decision systems. The highlighted limitations also point to critical directions for future research and improvement of the system, namely on the inclusion of external variables, advanced deep learning architectures (e.g. CNNs or GANs) as state-of-the-art in classification problems, real-time market intelligence through automated parsing of news and social media traffic, and more sophisticated decision-support mechanisms.

14. Implementation Outputs

Hyperlink to code https://github.com/TarunSarvang/Machine-Learning-Frameworks-for-Agricultural-Price-Stability-in-India

'Agriculture_price_dataset.csv' found locally. Loading data...

Cleaned data saved to 'Cleaned_Agri_Data.csv'

Final Shape: (282701, 15)

States: 27

Commodities: 5

Date Range: 2023-01-07 00:00:00 to 2025-12-05 00:00:00

--- Starting Data Analysis ---

Available Commodities:

1. Wheat

2. Onion

3. Potato

4. Tomato

5. Rice

Please enter the number corresponding to the commodity you want to analyze (1-5): 1

Selected commodity: Wheat

Available Varieties for Wheat:

1. 147 Average

2. Other

3. Lokwan

4. Maharashtra 2189

5. Sonalika

6. Red

7. Mill Quality

8. Local

9. Lok-1

10. Dara

11. 2189 No. 2

12. Hybrid

13. Rajasthan Tukdi

14. Medium

15. Sharbati

16. Kalyan

17. Deshi

18. 1482

19. Dara Mill Quality

20. Lok -1 (Nilami Rate)

21. Mp(Desi)

22. Mexican

23. 147 Best

24. Mp Sharbati

25. Super Fine

26. Sechor No. 1

27. Pissi

28. White

29. Jawari

30. Lokwan Mp No. 1

31. Medium Fine

32. Lokwan Gujrat

33. Milbar

34. Farmi

35. Bansi

36. 2189 No. 1

37. Wh-542

38. 343

39. Sona

40. Chandausi

41. Pbw-373

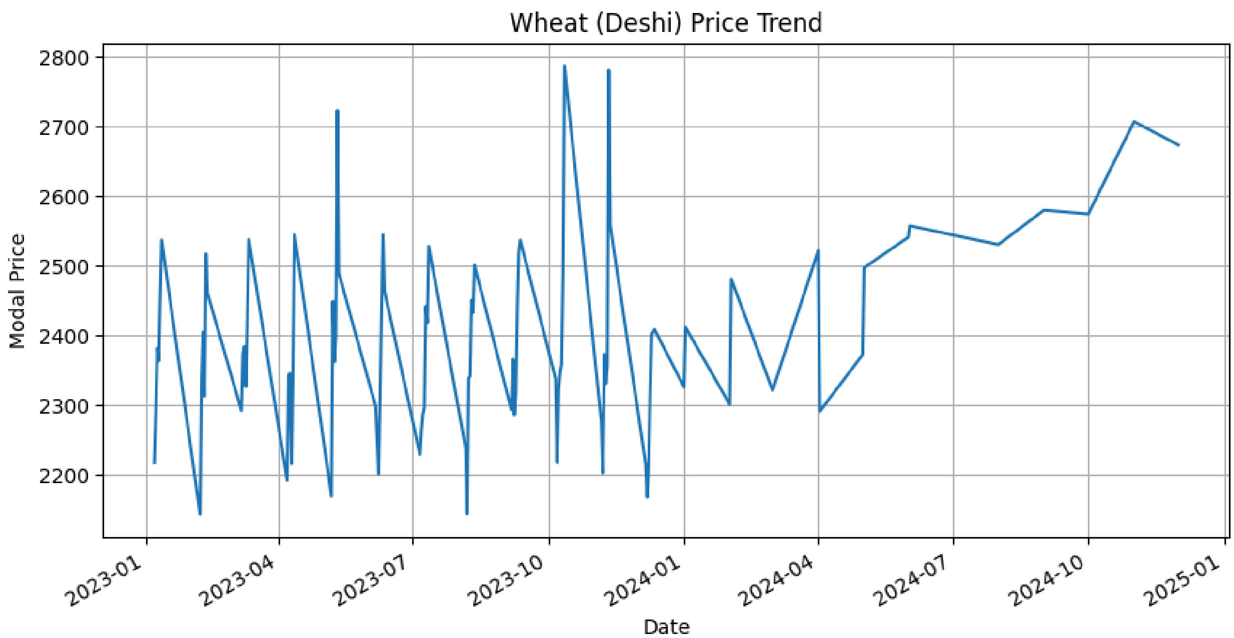
42. H.D.

43. Analyze all varieties for Wheat

Please enter the number corresponding to the variety you want to analyze (1-43), or choose to analyze all varieties: 17

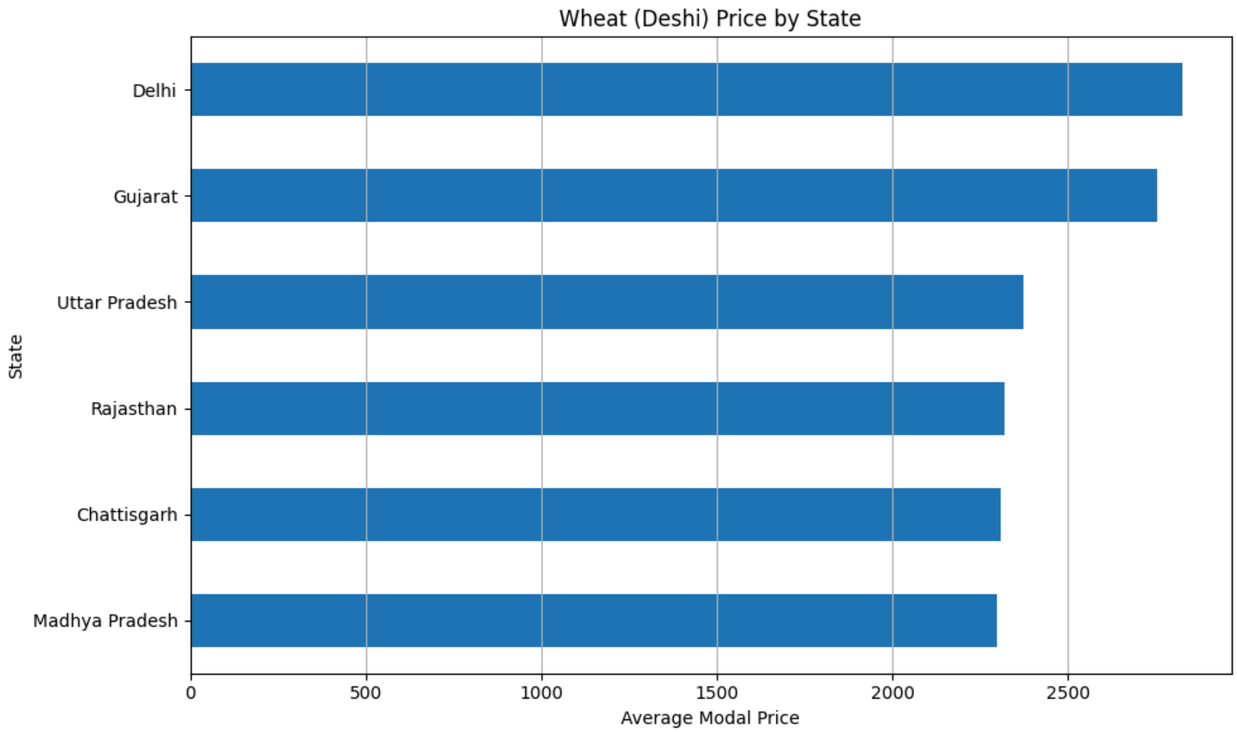
Selected variety: Deshi

Figure 14.1 - Overall Price Trend for Wheat (Deshi)



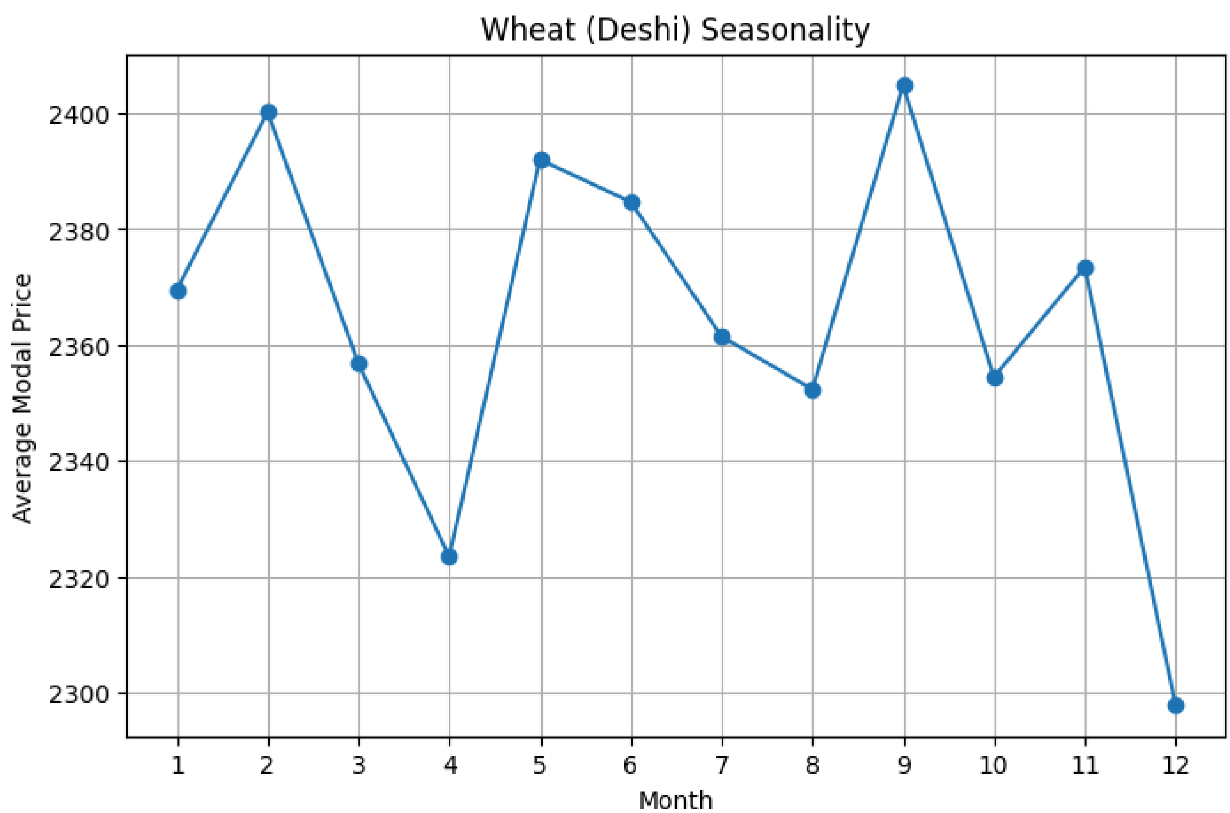
Source(Own analysis,Python Code)

Figure 14.2 - State-wise Price Comparison for Wheat (Deshi)



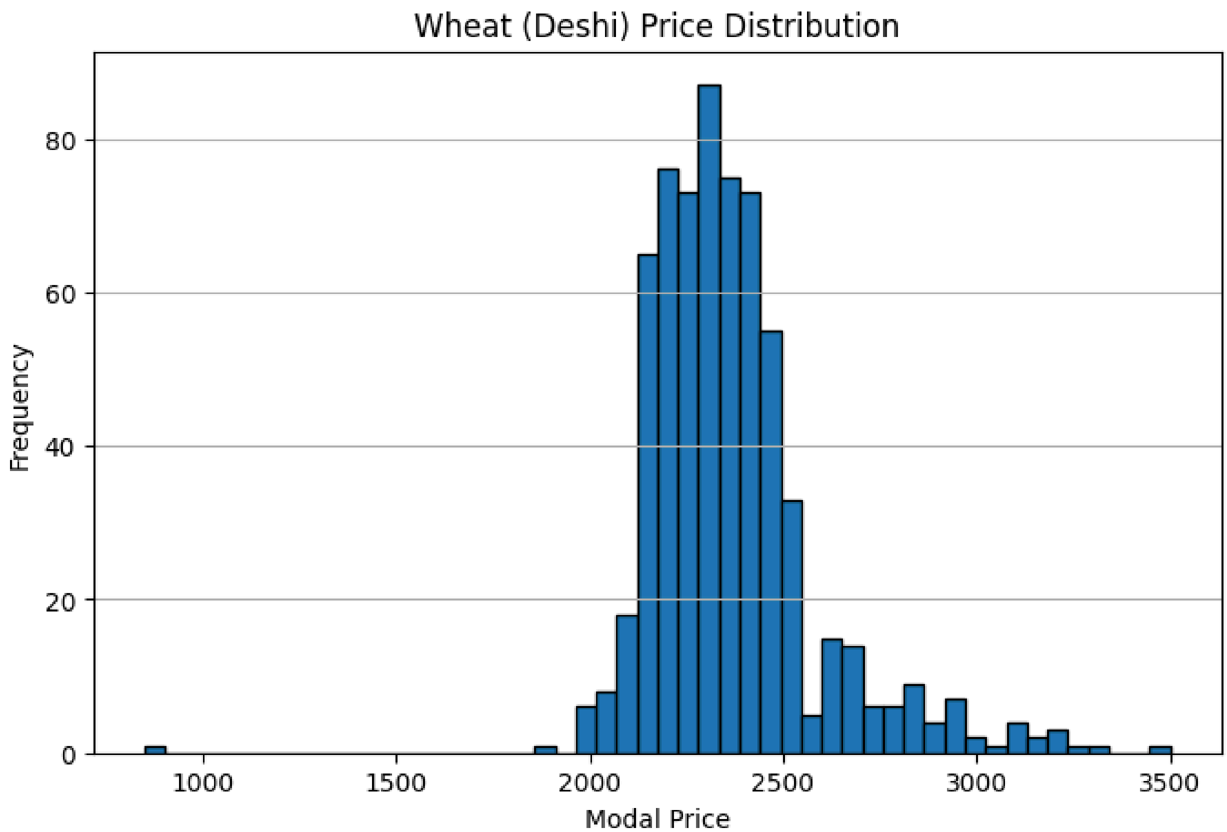
Source(Own analysis, Python Code)

Figure 14.3 - Seasonality Analysis for Wheat (Deshi)



Source(Own analysis, Python Code)

Figure 14.4 - Price Distribution for Wheat (Deshi)



Source(Own analysis, Python Code)

Table 14.1 - Top Markets for Wheat (Deshi)

Market_Name	
Najafgarh	2824.608696
Bavla	2755.333333
Kasrawad	2540.000000
Paatan	2375.818182
Pratapgarh	2373.041667
Durg	2365.000000
Kareli	2349.769231
Rajnandgaon	2330.564516
Sabalgarh	2323.062500
Tonk	2305.428571

Name: Modal_Price, dtype: float64

Source(Own analysis, Python Code)

--- Starting Predictive Analysis ---

Time Series Preparation for Wheat (Deshi)

ARIMA Forecast for Wheat (Deshi)

ARIMA training set length: 556

ARIMA test set length: 139

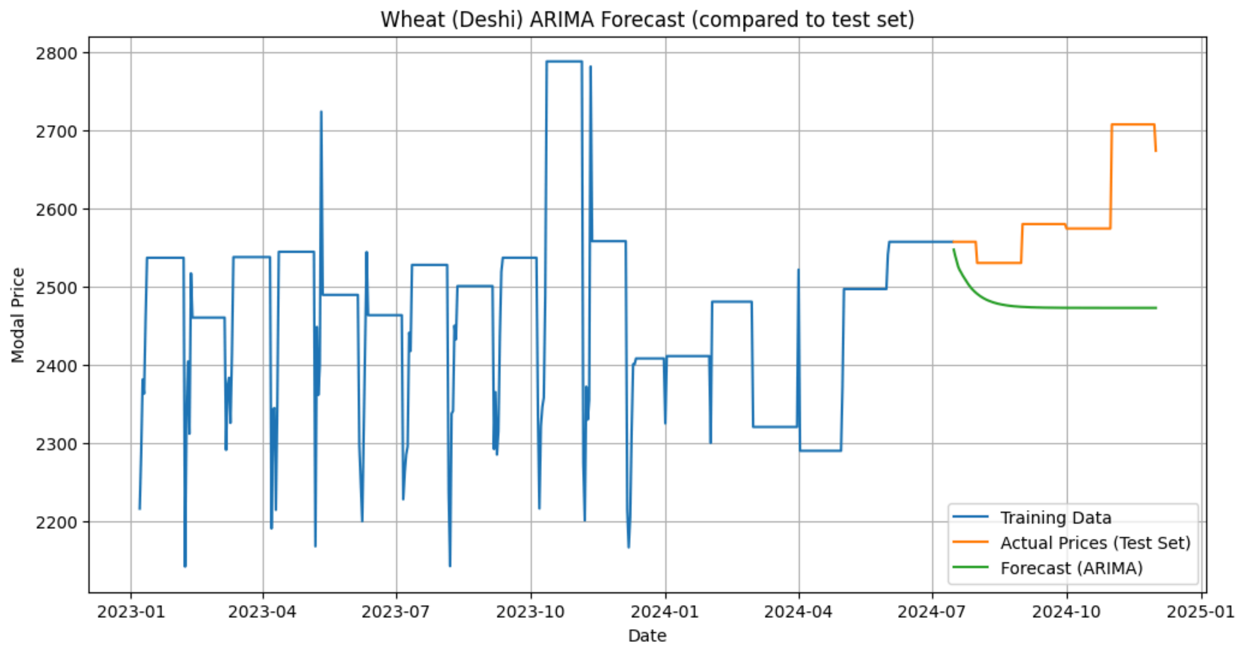
MAE for Wheat (Deshi) (ARIMA): 114.10

R-squared for Wheat (Deshi) (ARIMA): -3.43

Correlation (R) for Wheat (Deshi) (ARIMA): -0.30

Interpretation: On average, the ARIMA model's price predictions for Wheat (Deshi) were off by approximately 114.10 units. R-squared indicates -342.89% of the variance in actual prices is explained by the model. R (-0.30) shows the linear relationship between actual and predicted prices.

Figure 14.5 - ARIMA Forecast Plot for Wheat (Deshi)



Source(Own analysis, Python Code)

Machine Learning Models (Lag-Based) for Wheat (Deshi) and Comparison (Predicting Daily Average)

Tuning Random Forest...

Best Random Forest parameters: {'max_depth': 10, 'min_samples_split': 2, 'n_estimators': 200}

Tuning Gradient Boosting...

Best Gradient Boosting parameters: {'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}

Tuning Linear Regression...

Best Linear Regression parameters: {'fit_intercept': False}

Tuning XGBoost...

Best XGBoost parameters: {'learning_rate': 0.1, 'max_depth': 5, 'n_estimators': 100}

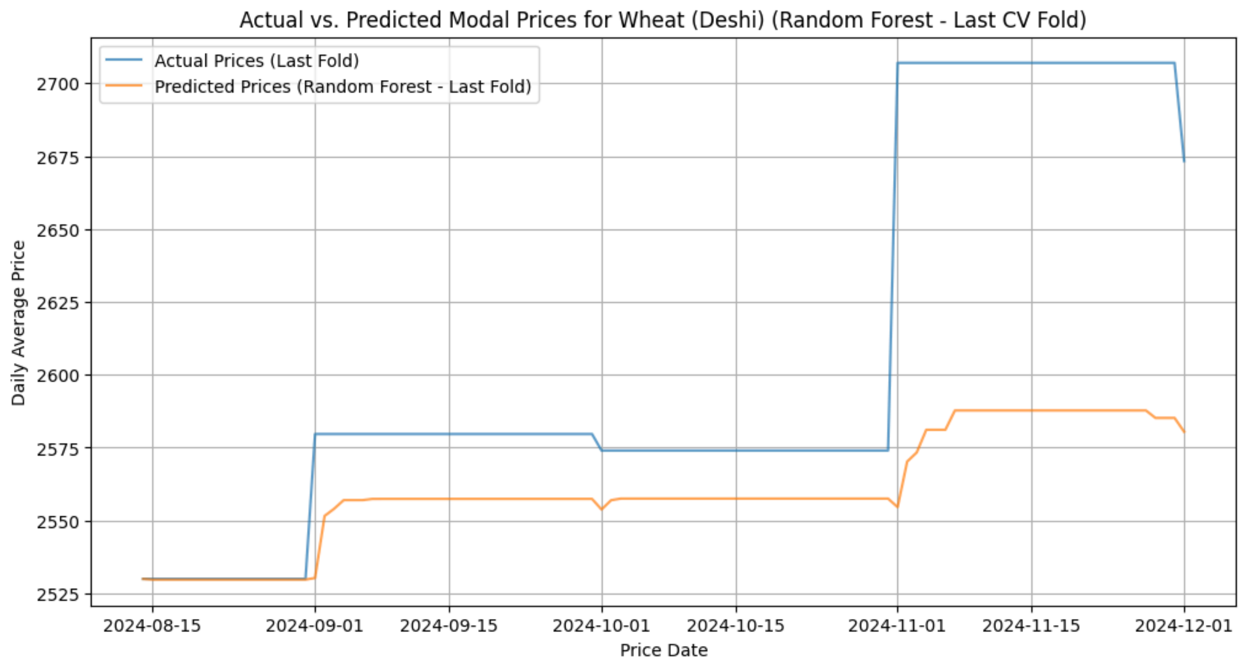
--- Re-evaluating Models with Time-Series Cross-Validation ---

Random Forest (Avg MAE): 55.24

Random Forest (Avg R-squared): 0.12

Random Forest (Avg Correlation R): 0.76

Figure 14.6 - Random Forest Actual vs Predicted Plot for Wheat (Deshi)



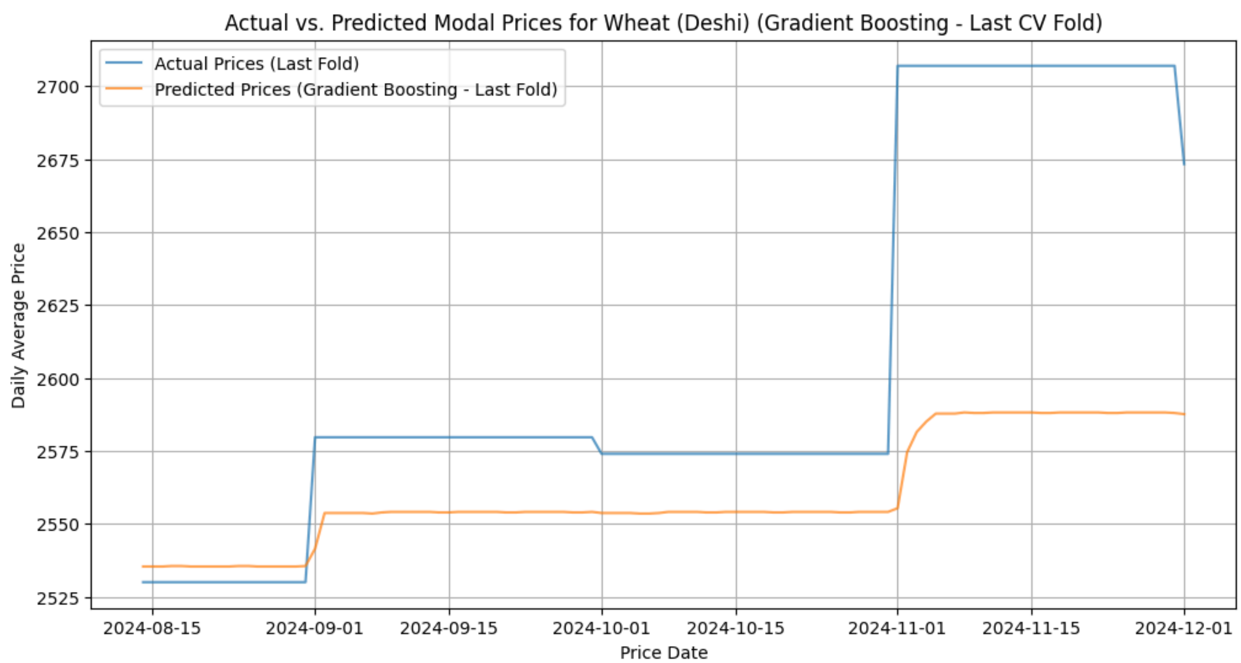
Source(Own analysis, Python Code)

Gradient Boosting (Avg MAE): 59.72

Gradient Boosting (Avg R-squared): 0.02

Gradient Boosting (Avg Correlation R): 0.74

Figure 14.7 - Gradient Boosting Actual vs Predicted Plot for Wheat (Deshi)



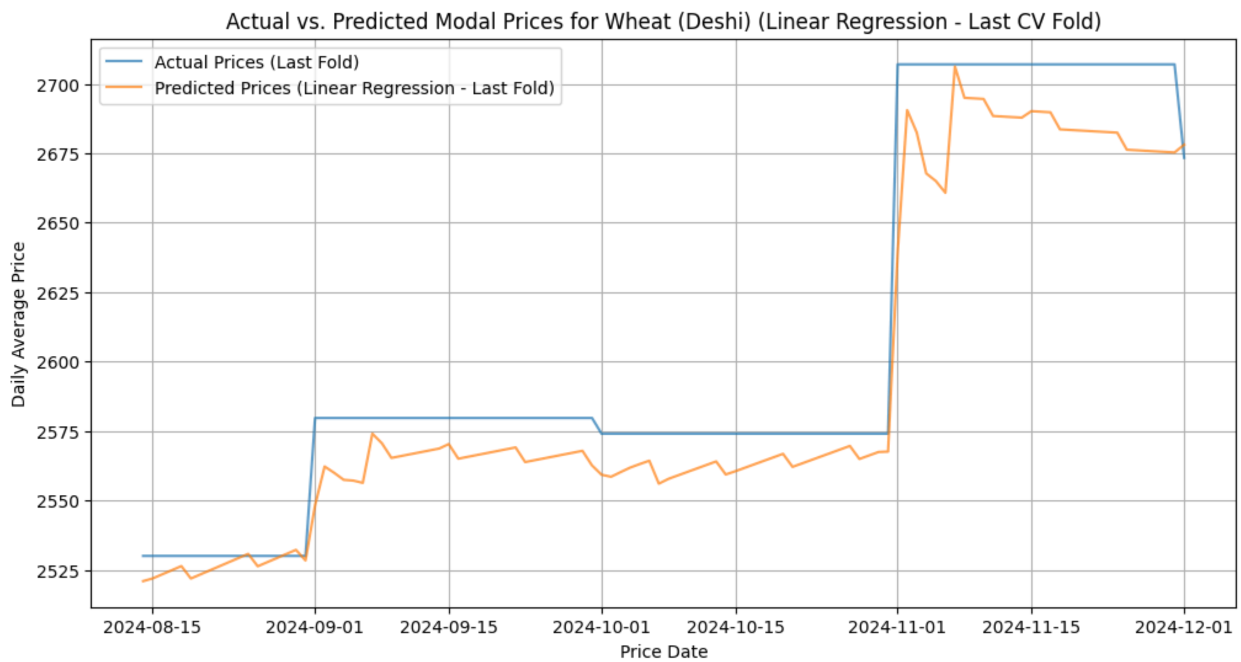
Source(Own analysis, Python Code)

Linear Regression (Avg MAE): 32.07

Linear Regression (Avg R-squared): 0.73

Linear Regression (Avg Correlation R): 0.88

Figure 14.8 - Linear Regression Actual vs Predicted Plot for Wheat (Deshi)



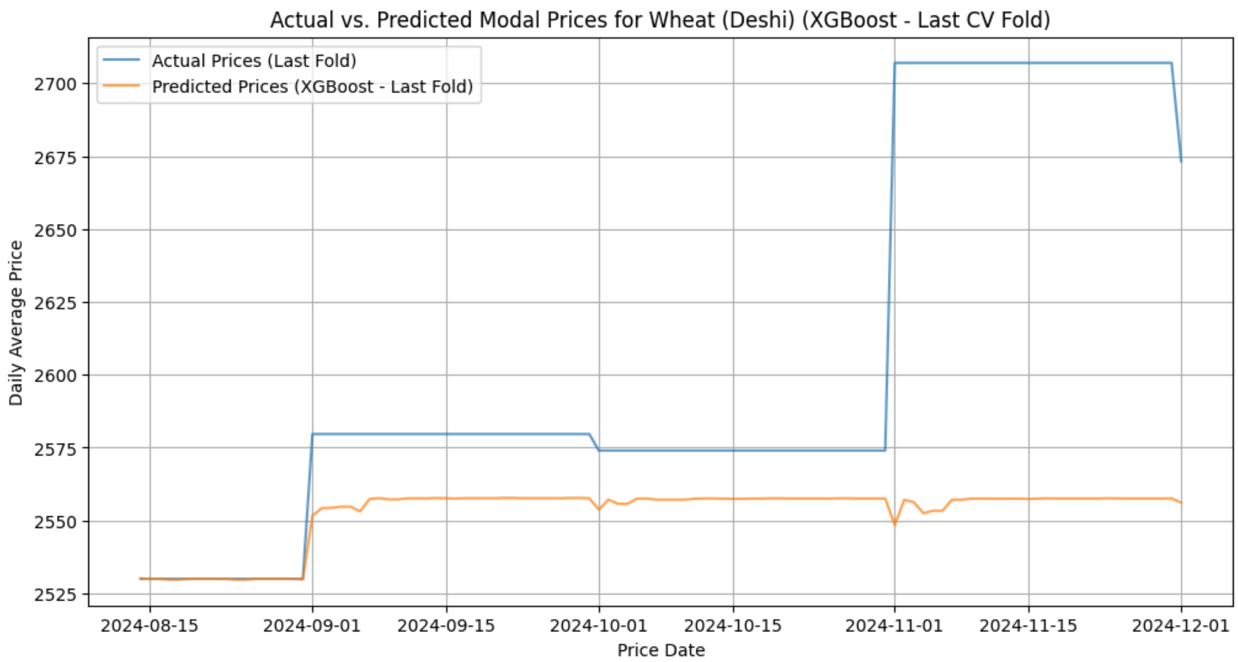
Source(Own analysis, Python Code)

XGBoost (Avg MAE): 61.66

XGBoost (Avg R-squared): -0.07

XGBoost (Avg Correlation R): 0.57

Figure 14.9 - XGBoost Actual vs Predicted Plot for Wheat (Deshi)



Source(Own analysis, Python Code)

ARIMA (MAE): 114.10

ARIMA (R-squared): -3.43

ARIMA (Correlation R): -0.30

Table 14.2 - Model Performance Comparison (After Tuning & TS-CV) ---

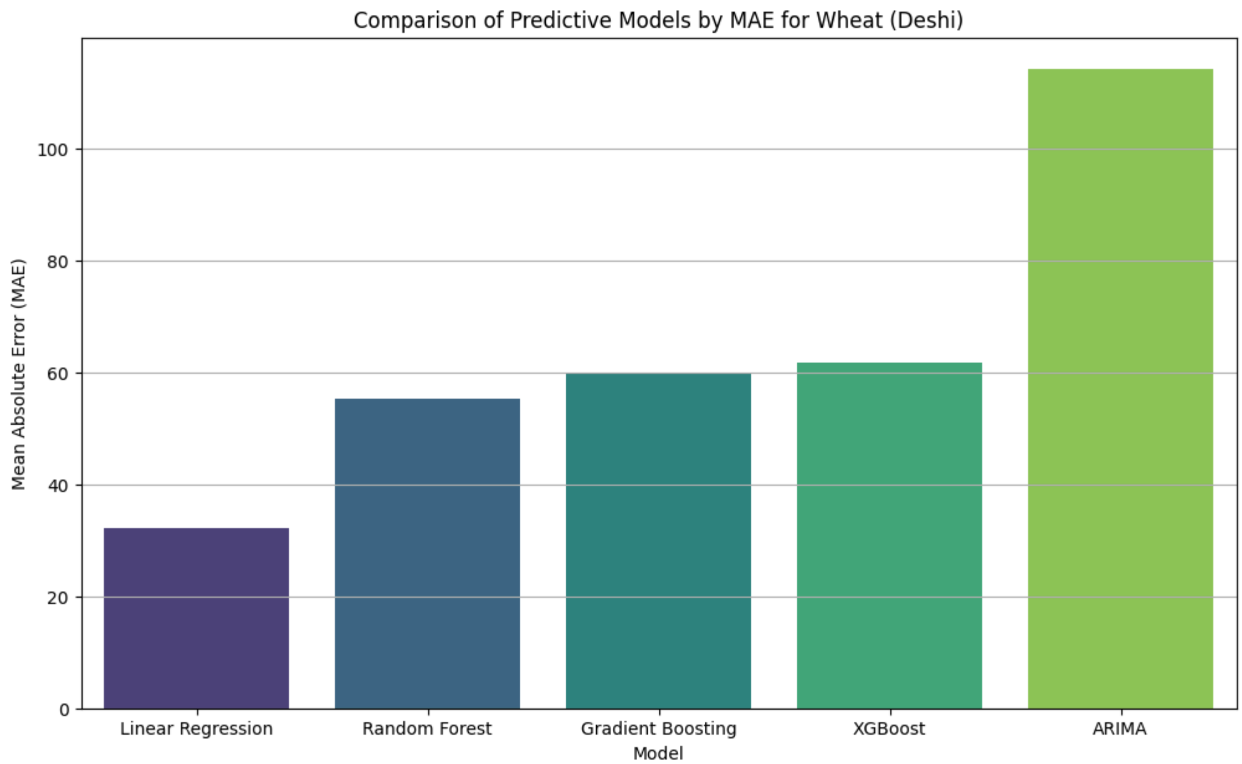
	MAE	R2	R
Linear Regression	32.067335	0.727066	0.884835
Random Forest	55.238825	0.122244	0.755370
Gradient Boosting	59.717963	0.022541	0.736515
XGBoost	61.663759	-0.074599	0.568895
ARIMA	114.100654	-3.428915	-0.296176

Source(Own analysis, Python Code)

Conclusion for Wheat (Deshi): After enhanced feature engineering, hyperparameter tuning, and time-series cross-validation, the **Linear Regression** model currently provides the best prediction accuracy with an average MAE of **32.07** for Wheat (Deshi). It explains **72.71%** of the variance in prices (average R-squared) and has an average correlation coefficient (R) of **0.88** between actual and predicted prices.

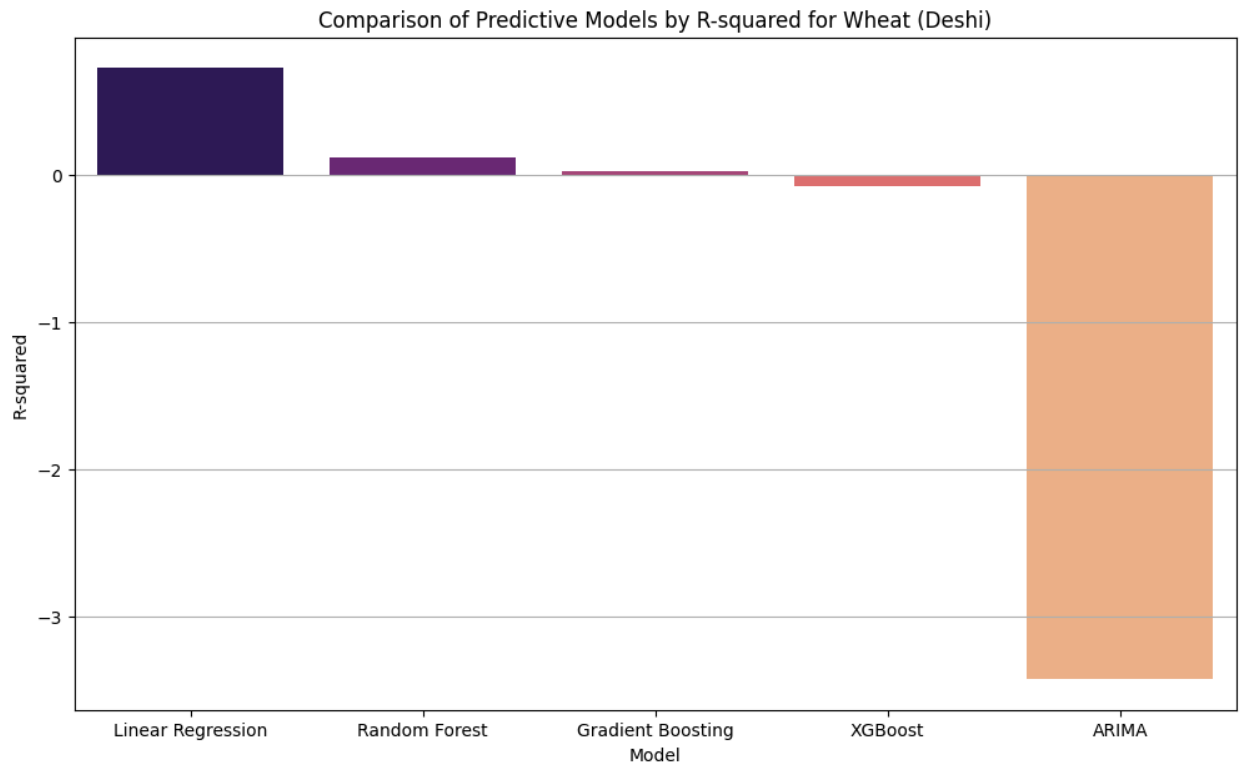
A lower MAE indicates better predictive performance, meaning the model's predictions are closer to the actual prices on average. A higher R-squared (closer to 1) indicates that the model explains more of the variance in the dependent variable. A higher absolute value of R (closer to 1) indicates a stronger linear relationship.

Figure 14.10 - Model Performance Comparison (MAE) Plot for Wheat (Deshi)



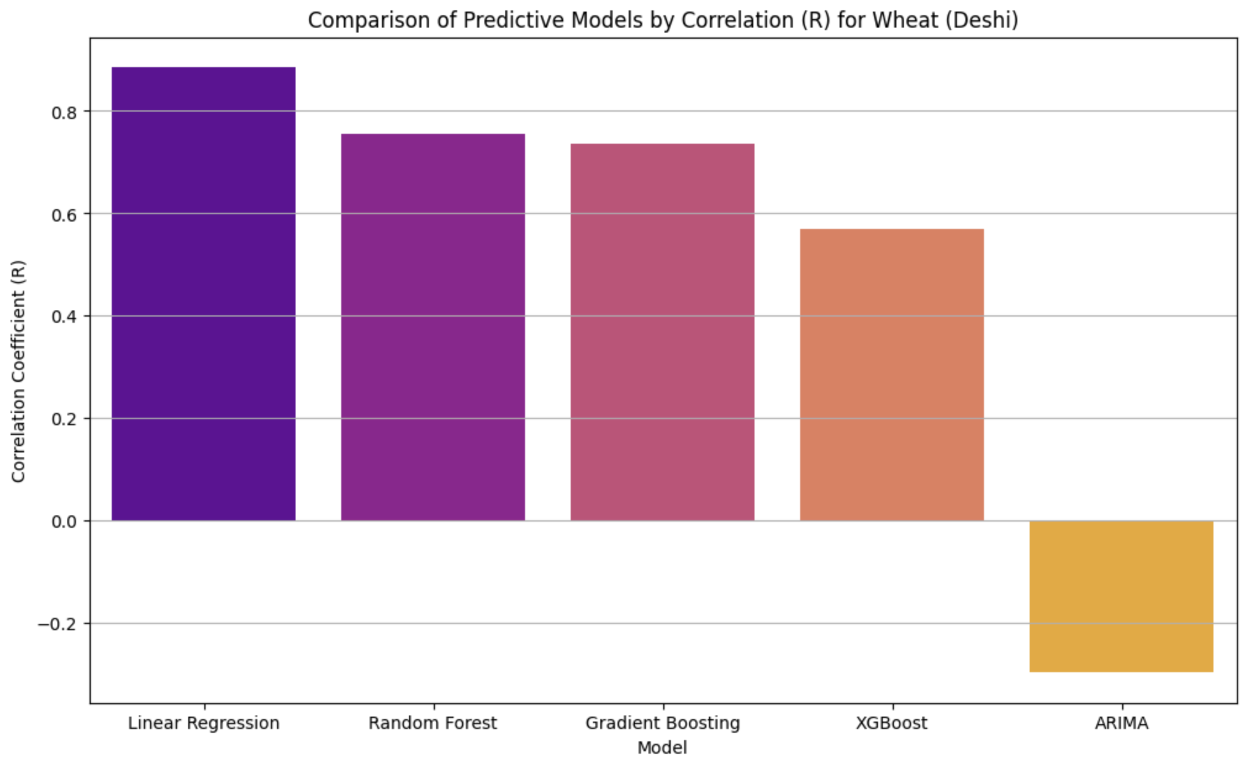
Source(Own analysis, Python Code)

Figure 14.11 - Model Performance Comparison (R-squared) Plot for Wheat (Deshi)



Source(Own analysis, Python Code)

Figure 14.12 - Model Performance Comparison (Correlation R) Plot for Wheat (Deshi)



Source(Own analysis, Python Code)

--- Starting Prescriptive Analysis (Decision System) ---

Sell vs Hold Decision for Wheat (Deshi)

Current Price (last observed): 2673.33

Predicted Price (next period - from ARIMA forecast): 2546.96

Decision based on ARIMA forecast: SELL

Interpretation: If the predicted price is higher than the current price, the recommendation is to HOLD. Otherwise, it's to SELL.

Best Market Recommendation for Wheat (Deshi)

Table 14.3 - Top Markets to Sell (based on average Modal Price):

Market_Name	
Najafgarh	2824.608696
Bavla	2755.333333
Kasrawad	2540.000000
Paatan	2375.818182
Pratapgarh	2373.041667

Name: Modal_Price, dtype: float64

Source(Own analysis, Python Code)

Interpretation: These are the markets where the commodity currently fetches the highest average prices, potentially indicating better selling opportunities.

Best State Recommendation for Wheat (Deshi)

Table 14.4 - Best States to Sell (based on average Modal Price):

STATE	
Delhi	2824.608696
Gujarat	2755.333333
Uttar Pradesh	2373.041667
Rajasthan	2320.313725
Chattisgarh	2307.984375

Name: Modal_Price, dtype: float64

Source(Own analysis, Python Code)

Interpretation: These are the states with the highest average prices for the commodity, suggesting optimal regions for selling.

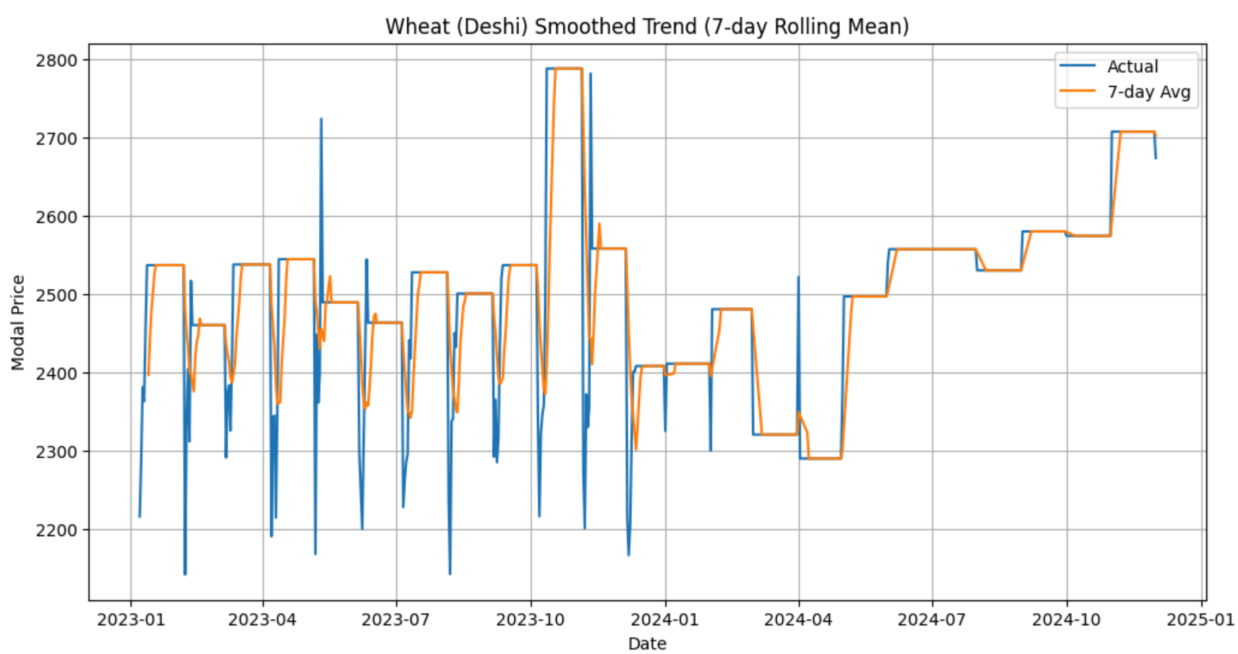
--- Starting Advanced Analysis ---

Volatility Analysis for Wheat (Deshi)

Wheat (Deshi) Price Volatility: 0.0241

Interpretation: Volatility measures the degree of variation of a trading price series over time. A higher value indicates greater price fluctuations.

Figure 14.13 - Rolling Mean (Trend Smoothing) for Wheat (Deshi)

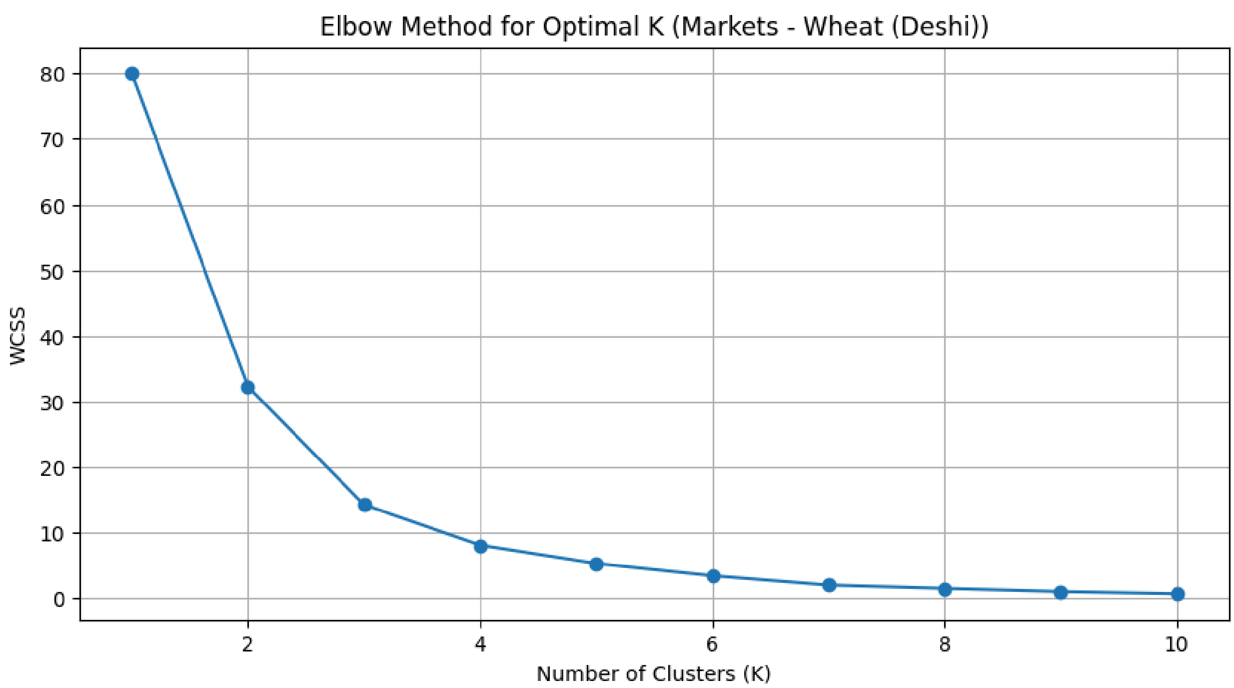


Source(Own analysis, Python Code)

Interpretation: The 7-day rolling mean smooths out short-term fluctuations to highlight longer-term trends or cycles in price movement.

--- Starting Clustering Analysis ---

Figure 14.14 - Clustering Markets by Price Profile for Wheat (Deshi)



Source(Own analysis, Python Code)

Interpretation: The 'elbow' point in the graph suggests the optimal number of clusters (K).

Market Clusters (3 clusters) for Wheat (Deshi):

-- Cluster 0 --

Table 14.5 - Average Prices in Cluster 0:

Avg_Min_Price	2226.68
Avg_Max_Price	2358.11
Avg_Modal_Price	2298.33
Avg_Price_Spread	131.42

Markets in Cluster 0:

['Ajaygarh', 'Anuppur', 'Durg', 'Hanumangarh Town', 'Jora', 'Kareli', 'Kasrawad', 'Khujner', 'Lateri', 'Mandla', 'Paatan', 'Rajnandgaon', 'Sabalgarh', 'Seoni', 'Tendukheda', 'Tonk']

Source(Own analysis, Python Code)

-- Cluster 1 --

Table 14.6 - Average Prices in Cluster 1:

Avg_Min_Price	851.0
Avg_Max_Price	851.0
Avg_Modal_Price	851.0
Avg_Price_Spread	0.0

Markets in Cluster 1:

['Naila']

Source(Own analysis, Python Code)

-- Cluster 2 --

Table 14.7 - Average Prices in Cluster 2:

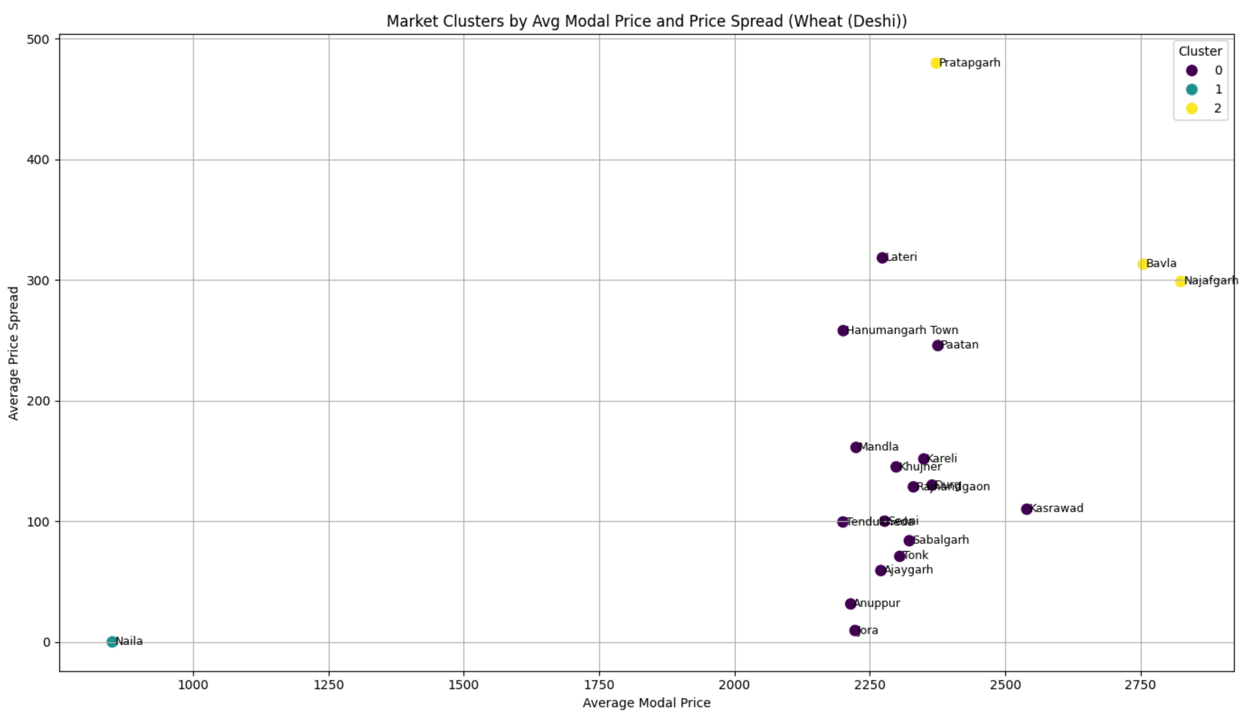
Avg_Min_Price	2530.62
Avg_Max_Price	2894.43
Avg_Modal_Price	2650.99
Avg_Price_Spread	363.81

Markets in Cluster 2:

['Bavla', 'Najafgarh', 'Pratapgarh']

Source(Own analysis, Python Code)

Figure 14.15 - Market Clusters Scatter Plot for Wheat(Deshi)



Source(Own analysis, Python Code)

Interpretation: Markets are grouped into clusters based on their average price characteristics. This can help identify market segments (e.g., high-price/low-spread markets vs. low-price/high-spread markets).

15. References

1. Agriculture price prediction using machine learning.(n.d). Retrieved May 13, 2026, https://showcase.itcarlow.ie/C00290851.pdf
2. Machine Learning Price Forecasting and Welfare-Optimal MSP Determination: Evidence from Indian Wheat and Rice.(n.d) ResearchGate. Retrieved May 13, 2026, https://www.researchgate.net/publication/404372822_Machine_Learning_Price_Forecasting_and_Welfare-Optimal_MSP_Determination_Evidence_from_Indian_Wheat_and_Rice
3. Machine Learning Based Agricultural Price Forecasting for Major Food Crops in India Using Environmental and Economic Factors.(2024). MDPI. Retrieved May 13, 2026, https://www.mdpi.com/2673-9976/54/1/7
4. Agricultural Market Price Prediction Using Machine Learning and ARIMA Time Series Models(n.d.) - IRE Journals, Retrieved May 13, 2026, https://www.irejournals.com/formatedpaper/1711906.pdf
5. Price forecasting models for rice in West Bengal, India.(2024). Agronomy Journal. Retrieved May 13, 2026, https://www.agronomyjournals.com/archives/2024/vol7issue11S/PartC/S-7-11-17-305.pdf
6. Agriculture_price_dataset.csv
7. Food Financialization: Impact of Derivatives and Index Funds on Agri-Food Market Volatility.(2024). MDPI. Retrieved May 13, 2026, https://www.mdpi.com/2227-7072/12/4/121
8. Food Price Volatility in India. (n.d) IDEAS/RePEc. Retrieved May 13, 2026, https://ideas.repec.org/p/ags/icarpp/349209.html
9. Machine Learning Enabled Crop Price Prediction : A Review.(n.d) IJCS. Retrieved May 13, 2026, https://www.indianjournalofcomputerscience.com/index.php/tcsj/article/view/175963
10. The Impact of Commodity Price Volatility on Farmers & Consumers.(n.d). Invade Agro Global. Retrieved May 13, 2026, https://www.invadeagro.com/post/the-impact-of-commodity-price-volatility-on-farmers-consumers
11. Machine Learning for Crop Price Prediction.(n.d). Scribd. Regression Analysis - Scribd,

Retrieved May 13, 2026, https://www.scribd.com/document/985254880/Research-Paper-Crop-Price-Prediction
12. Forecasting of Retail Price Rise in Tomato, Onion and Potato Commodities in India.(n.d). Retrieved May 13, 2026, https://www.isaeindia.com/publication/download-pdf/345
13. Development of AI-ML Based Models for Predicting Prices of Agri-Horticultural Commodities(2025). IJSDR. Retrieved May 13, 2026, https://www.ijsdr.org/papers/IJSDR2511097.pdf
14. Price Dynamics of Tomato, Onion and Potato (TOP)in India.(2024). AWS. Retrieved May 13, 2026, https://sdiopr.s3.ap-south-1.amazonaws.com/2024/Feb/2-Mar-24/2024_AJAEES_112355/Revised-ms_AJAEES_112355_v1.pdf
15. AIML-Based Agro Product Price Prediction and Decision Support System Using XGBoost - IJIRT, Retrieved May 13, 2026, https://ijirt.org/publishedpaper/IJIRT190411_PAPER.pdf
16. Machine learning prediction of agricultural produces for Indian Farmers using LSTM.(2024). Retrieved May 13, 2026, https://www.allmultidisciplinaryjournal.com/uploads/archives/20240719181308_D-22-145.1.pdf
17. Navigating Rice Export Restrictions: The Impact of India's Policy on Domestic and International Markets ASEAN Journal of Agriculture(2022). Retrieved May 13, 2026, http://www.ejournal.bumipublikasinusantara.id/index.php/ajafe/article/download/375/305
18. Outlook of India Rice Production and International Rice Price.(n.d). Choices Magazine. Retrieved May 13, 2026, https://www.choicesmagazine.org/UserFiles/file/cmsarticle_998.pdf
19. The Impact of Meteorological Factors on Crop Price Volatility in India: Case studies of Soybean and Brinjal.(2025). arXiv. Retrieved May 13, 2026, https://arxiv.org/html/2503.11690v1
20. Machine Learning Price Prediction of Agricultural Produces for Indian Farmers using LSTM(2022). ResearchGate. Retrieved May 13, 2026, https://www.researchgate.net/publication/363376035_Machine_Learning_Price_Prediction_of_Agricultural_Produces_for_Indian_Farmers_using_LSTM
21. The Price of Plenty: India's rice surplus and global market effects.(n.d). CGSpace.

Retrieved May 13, 2026,
https://cgspace.cgiar.org/items/92f47565-70a5-43c2-8a3f-bb60b3a9ca06

22. Crop Price Prediction in India dataset.(n.d). - Kaggle, Retrieved May 13, 2026, https://www.kaggle.com/datasets/himanshu9648/crop-price-prediction-dataset-in-india
23. An ARIMA model approach for predicting wheat production in India and China.(n.d). Retrieved May 13, 2026, https://eds.yildiz.edu.tr/AjaxTool/GetArticleByPublishedArticleId?PublishedArticleId=7277
24. Price Fluctuations and its Impact on the Agricultural Sector of India.(2025). Zenodo. Retrieved May 13, 2026, https://zenodo.org/records/15067251
25. Agmark.net (n.d.) Directorate of Marketing and Inspection, Department of Agriculture and Farmers Welfare, Government of India, https://agmarknet.gov.in/home

Tarun Sarvang

Comprehensive Analysis and Machine Learning Frameworks for Agricultural Price Stability in India

Quick Submit

Quick Submit

Delhi Technological University

Document Details

Submission ID

trn:oid:::1:3580160271

Submission Date

May 27, 2026, 1:38 PM GMT+5:30

Download Date

May 27, 2026, 1:44 PM GMT+5:30

File Name

ing_Frameworks_for_Agricultural_Price_Stability_in_India_1.docx

File Size

3.5 MB

82 Pages

14,882 Words

89,986 Characters

27% detected as AI

The percentage indicates the combined amount of likely AI-generated text as well as likely AI-generated text that was also likely AI-paraphrased.

Caution: Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.

69 AI-generated only 26%
Likely AI-generated text from a large-language model.

4 AI-generated text that was AI-paraphrased 1%
Likely AI-generated text that was likely revised using an AI-paraphrase tool or word spinner.

Disclaimer
Our AI writing assessment is designed to help educators identify text that might be prepared by a generative AI tool. Our AI writing assessment may not always be accurate (i.e., our AI models may produce either false positive results or false negative results), so it should not be used as the sole basis for adverse actions against a student. It takes further scrutiny and human judgment in conjunction with an organization's application of its specific academic policies to determine whether any academic misconduct has occurred.

Comprehensive Analysis and Machine Learning Frameworks for Agricultural Price Stability in India

ORIGINALITY REPORT

6%	5%	4%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	agmarknet.gov.in Internet Source	1%
2	agmarknet.dac.gov.in Internet Source	1%
3	www.coursehero.com Internet Source	<1%
4	www.ijarcs.info Internet Source	<1%
5	web.realinfo.tv Internet Source	<1%
6	Vivek Sunnapwar, Hariram Chavan, Namrata F. Ansari. "Advances in Science and Technology", CRC Press, 2026 Publication	<1%
7	Submitted to British University in Egypt Student Paper	<1%
8	ijireeice.com Internet Source	<1%