**Time series modelling for forecasting price and arrival of Jowar crop in**

**Maharashtra markets**

Ambupe P.S

Department of Statistics

Yashavantrao Chavan Institute of Science, Satara

[**ambupepooja@gmail.com**](mailto:ambupepooja@gmail.com)

**Abstract**: In recent years, accurate forecasting of crop prices and arrivals has become increasingly crucial for supporting farmers and stabilizing agricultural markets. This study focuses on predicting the monthly arrival and modal price of jowar (sorghum) in Maharashtra’s markets using dynamic time series models. By employing statistical and machine learning technique specifically Exponential Smoothing, ARIMA, LSTM, hybrid LSTM-ARIMA, Random Forest and XGBoost the research examines the influence of variables such as rainfall, temperature, seasonality, and lagged arrivals on market dynamics. The dataset includes historical data on jowar arrivals and prices from 2010 to 2023 was obtained from Agmarknet website. The data-splitting technique, commonly used in comparative analysis, splits the data into training and testing data sets. Each model's performance is evaluated through metrics like Mean Absolute Percentage Error (MAPE) and ensuring reliable comparisons for forecasting accuracy. At last, Random Forest and XGBoost delivered the most accurate results for predicting prices, whereas ARIMA performed relatively better in forecasting arrivals. This research delivers a forecasting framework that supports Maharashtra’s jowar farmers in making informed market decisions, promotes efficient interventions, and ultimately contributes to the resilience of the agricultural economy. This research delivers a forecasting framework that supports Maharashtra’s jowar farmers in making informed market decisions, promotes efficient interventions, and ultimately contributes to the resilience of the agricultural economy.

**Keywords** :Agricultural Markets, Jowar Crop, Time series, machine learning

**Introduction:**

Time series analysis is a branch of statistical and data analysis that focuses on understanding and modelling data points collected or recorded at specific time intervals. Unlike traditional data analysis, where observations are assumed to be independent, time series data is inherently ordered, and each observation is typically dependent on previous observations. This sequential nature of time series data introduces unique challenges and opportunities for analysis. By recognizing and modelling the inherent patterns in time series data, analysts and researchers can gain valuable insights that support better decision-making and planning in various domains. In time series forecasting, the results are the predicted outputs from the trained models. There are many forecasting models available like LSTM and ARIMA models. ARIMA forecasts temporal dependencies using only historical values. These models help to gain better insights into the data and predict future trends.

Agriculture is a critical sector that plays a significant role in the economy, especially in countries where a large portion of the population depends on farming for their livelihood. Maharashtra is, a state in India that plays a vital role in the country’s Agriculture sector. Agriculture is the backbone of Maharashtra's economy. Maharashtra's agriculture is characterized by a wide range of crops, including cereals, pulses, oilseeds, fruits, vegetables, and cash crops like cotton and sugarcane. Jowar (sorghum) is one of the most important cereal crops in the world and is one of the four major food grains of our country. It is a staple food for millions of poor rural people in Asian and African countries. The origin of sorghum is generally believed to be around present-day Ethiopia or East Central Africa. Sorghum was taken from East Africa to India during the first millennium. India is one of the major producing countries. In the year 2005-06, Maharashtra occupied the highest position in the production of Jawar with 3.90 million tons of production (51.11%).[ https://agmarknet.gov.in/]

Crop prediction in Maharashtra is a critical area of research and policy planning, given the state's diverse agro-climatic zones, varied cropping patterns, and significant reliance on agriculture. Accurate crop predictions are essential for ensuring food security and enhancing the livelihoods of millions of farmers in the state. By leveraging advanced technologies and integrating region-specific data, accurate crop predictions can help address the challenges faced by the agricultural sector in the state. This paper aims to leverage dynamic time series modeling to better understand and predict the behaviors of jawar crop arrivals and prices in Maharashtra's markets.

**Literature Review:**

This literature indicates that while traditional models remain valuable for certain applications, neural network-based approaches like LSTM and CNN often provide better accuracy and adaptability, making them suitable for predicting sales in dynamic and seasonal markets, such as the furniture industry. This study aims to build on these insights by comparing various forecasting techniques to identify the most effective method for predicting future sales in the retail sector[1].The literature on price forecasting for brinjal (eggplant) highlights the importance of accurate predictions for aiding farmers, traders, and consumers in decision-making. Traditional statistical models like ARIMA have been commonly used but are limited by their assumptions and inability to capture non-linear patterns. These studies indicate that ML models outperform traditional methods in handling complex time series data, capturing sudden price fluctuations, and providing more accurate forecasts.[2].

The literature on time series forecasting in agricultural markets, indicates the importance of accurate prediction models. Studies have applied models such as the Autoregressive Integrated Moving Average (ARIMA) and Seasonal ARIMA (SARIMA) for stationary time series with clear seasonality and trends. Advanced models like Prophet and Simple Recurrent Neural Networks (RNNs) are also used for non-stationary or complex nonlinear time series. These models help predict market dynamics, which is crucial for stakeholders such as farmers and traders[3].This study indicates that price prediction and forecasting in agriculture highlight the importance of selecting the right algorithm for accurate predictions. Traditional models are used, particularly when seasonality influences price variations. Advanced machine learning models, including ensemble methods like stacking, have shown promise in improving prediction accuracy by combining multiple models. Recent research suggests that hybrid and ensemble models often outperform individual models, particularly for complex and nonlinear datasets[4].

The literature on time series forecasting highlights its critical role in various domains such as electricity demand, cloud workload, weather, sales, and business costs. Research has evolved from traditional statistical methods like ARIMA to advanced machine learning and deep learning approaches, including CNNs, LSTMs, and their hybrid forms like CBLSTM. These modern techniques offer improved accuracy and flexibility compared to older methods. The paper surveys these forecasting methods, comparing their effectiveness across different types of time series and applications[5].

**Methodology:**

Secondary data on Jowar crop price and arrival was obtained from the Agmarknet website, a reliable source of agricultural market data in India. The dataset included monthly data from multiple agricultural markets in Maharashtra, covering price (in INR per quintal) and arrival (in tons) of Jowar. The data period spans 2010 to 2024, allowing for comprehensive trend analysis over recent years.

For predicting Jowar prices and arrivals, we focused target variables as modal price (INR/quintal) and arrival (in tons) and independent variables as rainfall, temperature, rain days, and time-based variables (month and year). Seasonal categories (e.g., summer, monsoon, winter) were generated from the month data to capture any cyclical effects in Jowar arrivals and prices. The preliminary aim of this study is to develop time series forecasting models to predict the modal price and arrival of jowar crop, leveraging historical data and multivariate features such as rainfall, temperature, rainy days, and lagged arrival.

1. **Exponential Smoothing :**

Exponential smoothing is a baseline method of time series forecasting that assigns exponentially decreasing weights to past observations. It handles both trend and seasonality, either in additive or multiplicative form. In this project ,The Holt-Winters seasonal method was used due to the observed trend and seasonal patterns in the price and arrival data.

Exponential smoothing refers to the use of an exponentially weighted moving average (EWMA) to “smooth” a time series. Holt’s exponential smoothing is a little more complicated. It consists of two EWMAs: one for the smoothed values of xt, and another for its slope. The terms level and trend are also used.

st=αxt+(1−α)(st−1+bt−1)

bt=β(st−st−1)+(1−β)bt−1

1. **ARIMA:**

The ARIMA (Auto-Regressive Integrated Moving Average) model is a time series forecasting technique that combines three components: Auto-Regressive (AR), Differencing (I), and Moving Average (MA).  The AR component captures the relationship between an observation and a specified number of previous observations (lags). The Integrated (I) part refers to differencing, which is used to make the time series data stationary by removing trends or seasonality. The MA component models the relationship between an observation and a residual error from a moving average of previous error.

The AR(*p*) model is written as,

*Xt* = *c* + *ϕ*1*Xt-1*+ *ϕ*2*Xt-2*+ *· · ·* + *ϕ*n*Xt-p*+ *ϵt*

where *ϕ*1, *ϕ*2…..ϕp are [parameters](https://en.wikipedia.org/wiki/Parameter) and the random variable εtis [white noise](https://en.wikipedia.org/wiki/White_noise).

The Moving Average (MA) model of order q is given by:

Xt=μ+ϵt+θ1ϵt−1+ θ2ϵt−2+⋯+ θqϵt−q

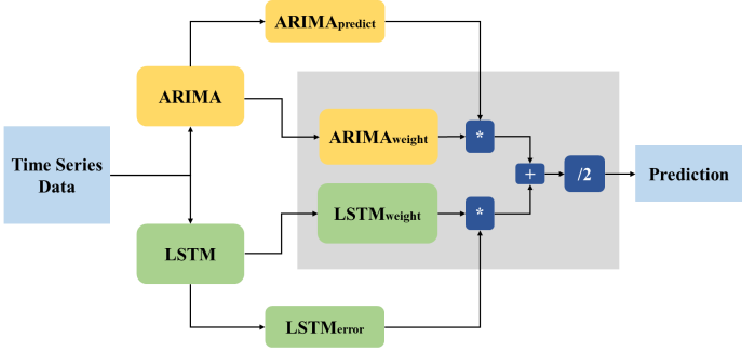
where ϵt​,ϵt−1​, …,ϵt−q​ white noise error terms.

The Autoregressive Moving Average (ARMA) model of order (p,q) combines both AR and MA components:

*Xt* = *c* + *ϕ*1*Xt-1*+ *ϕ*2*Xt-2*+ *· · ·* + *ϕ*n*Xt-p*+ *ϵt* +θ1ϵt−1+ θ2ϵt−2+⋯+ θqϵt−q

1. **ARIMAX-LSTM Hybrid Model:**

The hybrid model is used for two approaches as ARIMAX for linear patterns with external factors and LSTM for non-linear relationship to improve the model performance. Residuals are calculated as the difference between actual values and ARIMAX predictions, capturing nonlinear patterns. Residuals from the ARIMAX model, which represent nonlinear components, are used as input for the LSTM model. LSTM learns the complex nonlinear particularly those not captured by ARIMAX.



**Machine Learning Approach :**

1. **Random Forest and Xgboost:**

Random Forest was used to predict two target variables: price and arrival of the jowar crop. It handles both numerical and categorical variables. It builds multiple decision trees, each learning from different parts of the data. By combining the predictions from all these trees, the model becomes more accurate. XGBoost can handle both non-linear relationships and interactions between these variables, which traditional models like ARIMA may struggle with. Additionally, XGBoost is robust to overfitting.

**Metrics used for Model Performance :**

**MAPE:**

MAPE is a used accuracy metric for regression models, particularly in forecasting. It measures the percentage deviation between the predicted and actual values, making it interpretable and easy to understand. MAPE is calculated as:

MAPE=

Where:

Yi​: Actual value at time

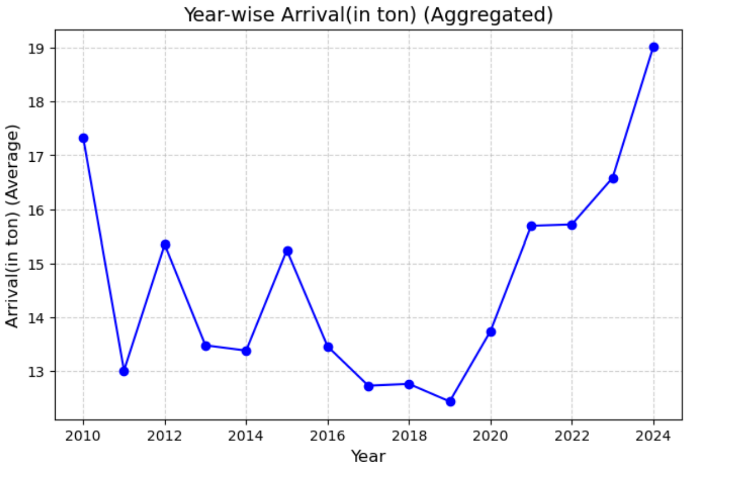
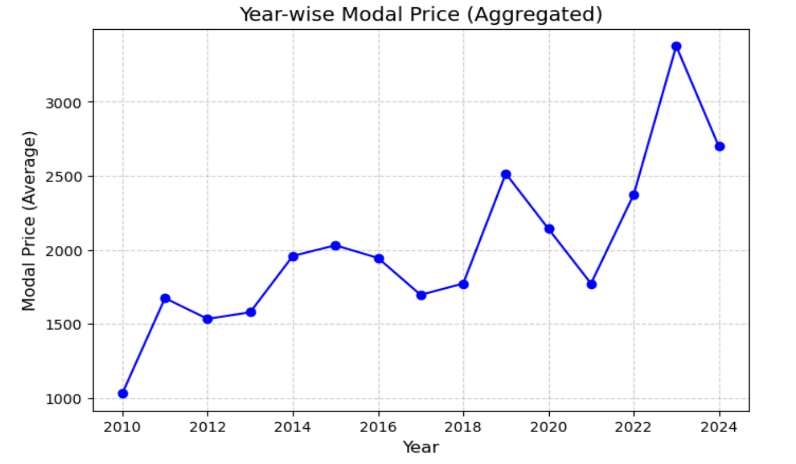
Yi\_hat: Predicted value at time

n: Total number of observations

**Statistical Analysis:**

* **Graphical Representation:**

1. **Year wise Price(per quintal) and Arrival(in ton)**

****

Year wise Trend in Average Year wise Trend in Average

Arrival (in ton) Modal Price(Per quintal)

**Fig 1**

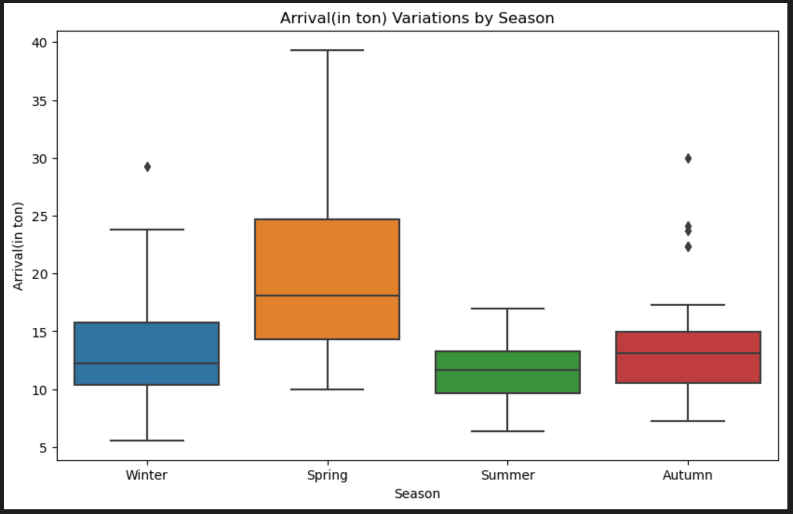
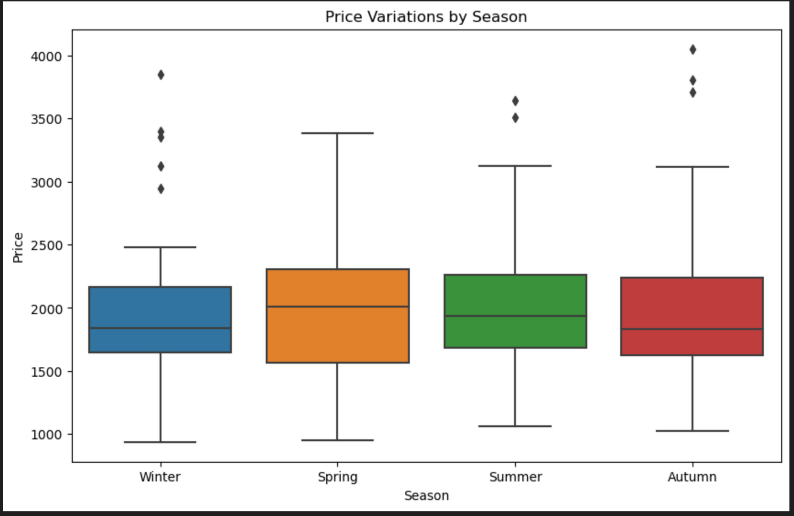
**Arrival (in tons):**

The graph shows a noticeable fluctuation in the average arrival of produce (in tons) over the years. There's a sharp decline in arrival between 2010 and 2011-2012.Then arrival gradually recovers and shows a general upward trend, especially towards the later years 2020-2024.The line graph exhibits volatility, indicating that the amount of produce arriving in the market varies significantly from year to year.

**Price (Per quintal):**

Similar to arrival, the modal price (in quintal) also shows a fluctuating trend. The modal prices start relatively low around 2010. The price fluctuations appear more volatile than the arrival fluctuations. The graph shows distinct peaks and valleys, indicating periods of rapid price increases and decreases. There appears to be a significant peak in modal price in the later years around 2023-2024

1. **Variation of Arrival and Price by Season:**

****

**Fig 2 Fig 3**

Variation of Arrival (in ton) by SeasonVariation of Price (per quintal)

by Season

The boxplots explain the seasonal variations in arrival (in tons) and price (per quintal) of the commodity:

In Fig 2, arrivals fluctuate across seasons, with Spring showing the highest median and variability, indicating greater supply during this period. Winter and Autumn have moderate arrivals, while summer records the lowest median arrival, suggesting reduced production.

In Fig 3, price variations across seasons are evident. While the median price remains fairly stable, Winter and Spring show slightly higher median prices than Summer and Autumn. The presence of outliers suggests price volatility, possibly due to seasonal demand.

Overall, higher arrivals in Spring may lead to slight price reductions, whereas lower arrivals in Summer could result in price stability or increases. Understanding these seasonal trends can aid in strategic pricing and inventory management.

* **Modelling:**

1. **Exponential Smoothing :**

**For Arrival(in ton):**

 **Level Equation**:

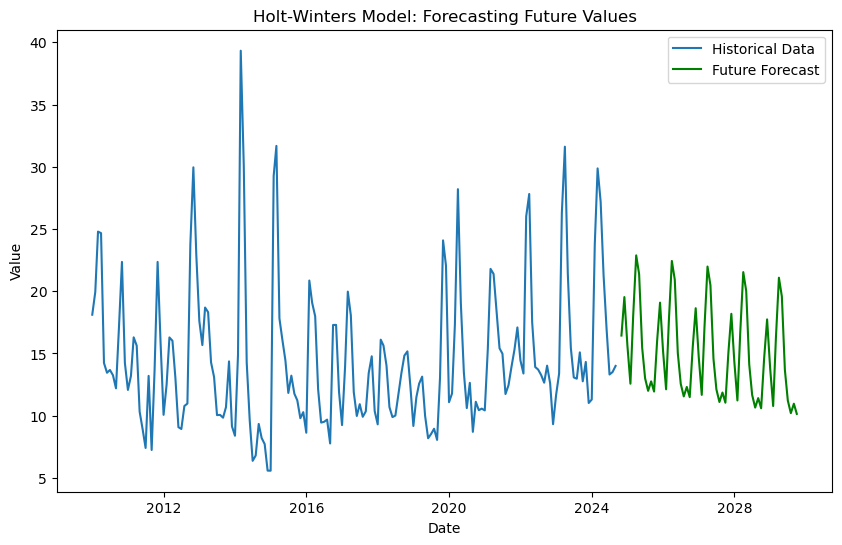
ℓt=0.9550⋅(yt−st−m) +(1−0.9550) ⋅(ℓt−1+bt−1)

 **Trend Equation**:

bt=3.5448×10-7⋅(ℓt−ℓt−1) +(1−3.5448×10-7) ⋅bt

 **Seasonal Equation**:

st=3.6816×10−8⋅(yt−ℓt) +(1−3.6816×10−8) ⋅st



**Fig 5**

**For Price (per quintal):**

 **Level Equation**:

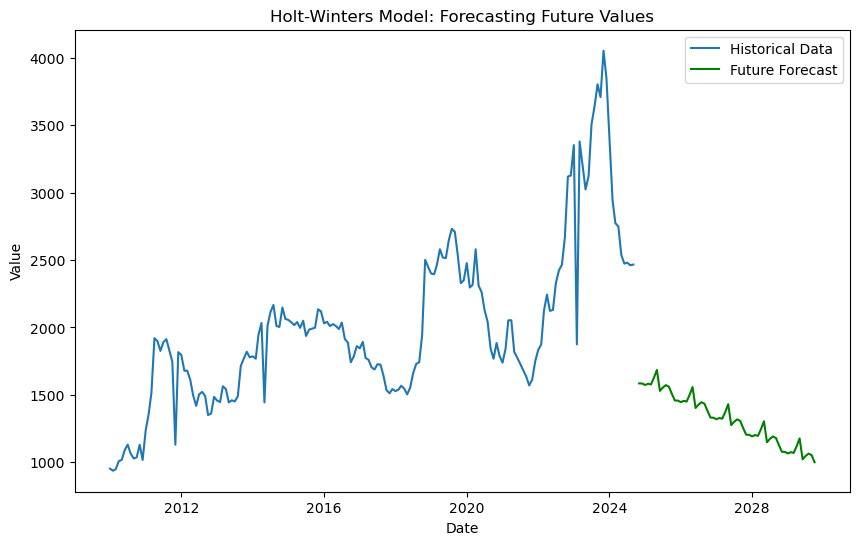
ℓt= 0.8279⋅(yt−st−m) +(1−0.8279) ⋅(ℓt−1+bt−1)

 **Trend Equation**:

bt= 0.06550⋅(ℓt−ℓt−1) +(1−0.06550) ⋅bt

 **Seasonal Equation**:

st= 0.007385⋅(yt−ℓt) +(1−0.007385) ⋅st



**Fig 6**

1. **ARIMA:**

**For Arrival (in ton)**:

By ADF test, p-value = 0.040745 < 0.05

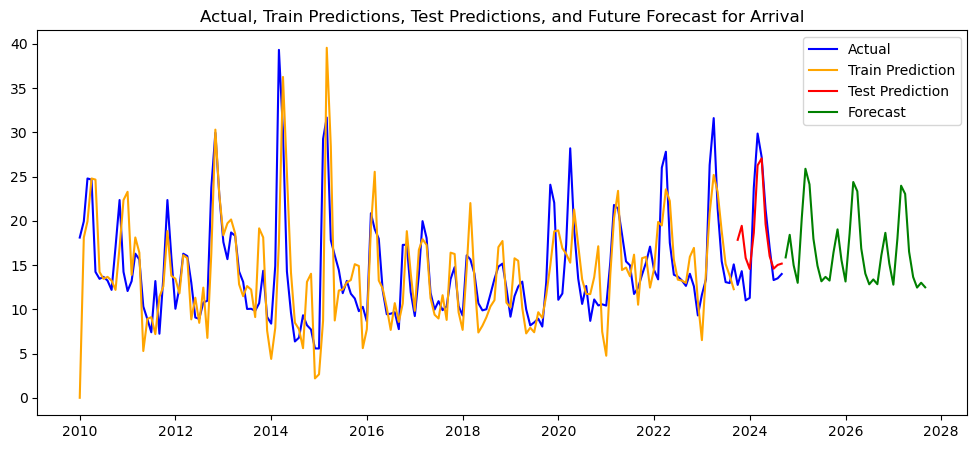
Hence, It is concluded that data is Stationary.

The ARIMA model is characterized by three key parameters: p (autoregressive order), d (degree of differencing), and q (moving average order). we determined the optimal parameters to be p=1, d=0 and q=1.

The Auto ARIMA function was used to automate the order selection process by iteratively testing multiple combinations of ARIMA parameters (p, d, q). The optimal order was determined based on the lowest Akaike Information Criterion (AIC) score, ensuring the best balance between model complexity and goodness of fit.

For ARIMA(1, 0, 1), the model is written as:

*Yt= ϕ*1*Yt-1+ϵt* +θ1*ϵt-1*



**Fig 7**

**For Price (in quintal):**

By ADF test, , p-value = 0.1941> 0.05

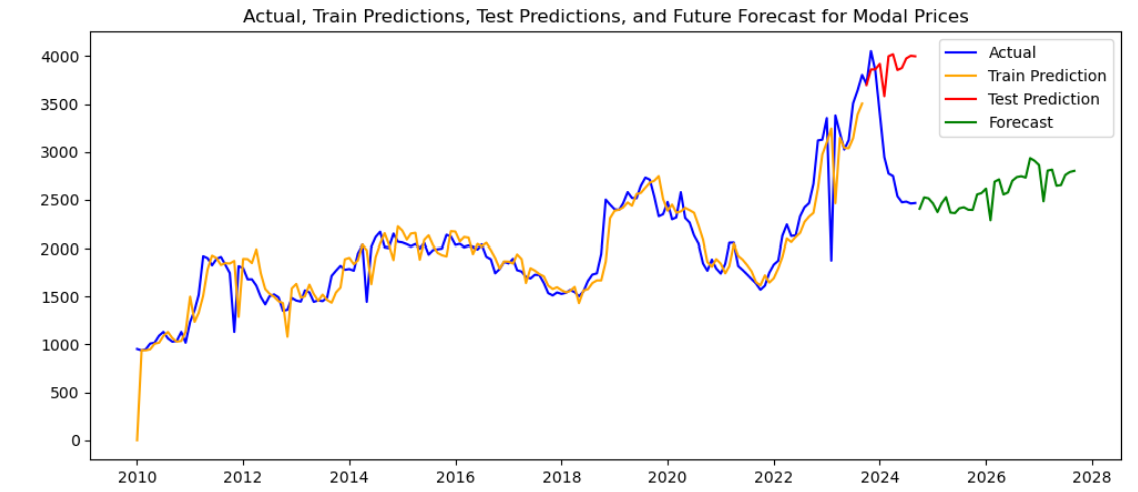
It is concluded that data is non stationary. Hence ,we use differencing to make the data stationary. The ARIMA model is fitted to the data using the selected p, d and q values, combining past values and errors to predict future points (p=0,d=1,q=1).

The ARIMA model is characterized by three key parameters: p (autoregressive order), d (degree of differencing), and q (moving average order). we determined the optimal parameters to be p=1, d=0 and q=1.

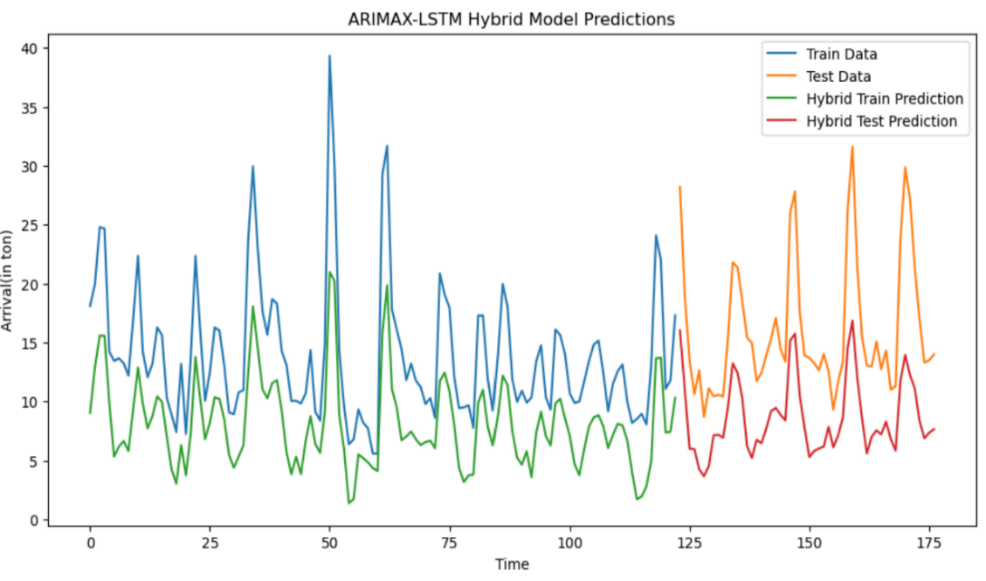
The Auto ARIMA function was used to automate the order selection process by iteratively testing multiple combinations of ARIMA parameters (p, d, q). The optimal order was determined based on the lowest Akaike Information Criterion (AIC) score, ensuring the best balance between model complexity and goodness of fit.

For ARIMA(0, 0, 1), the model is written as:

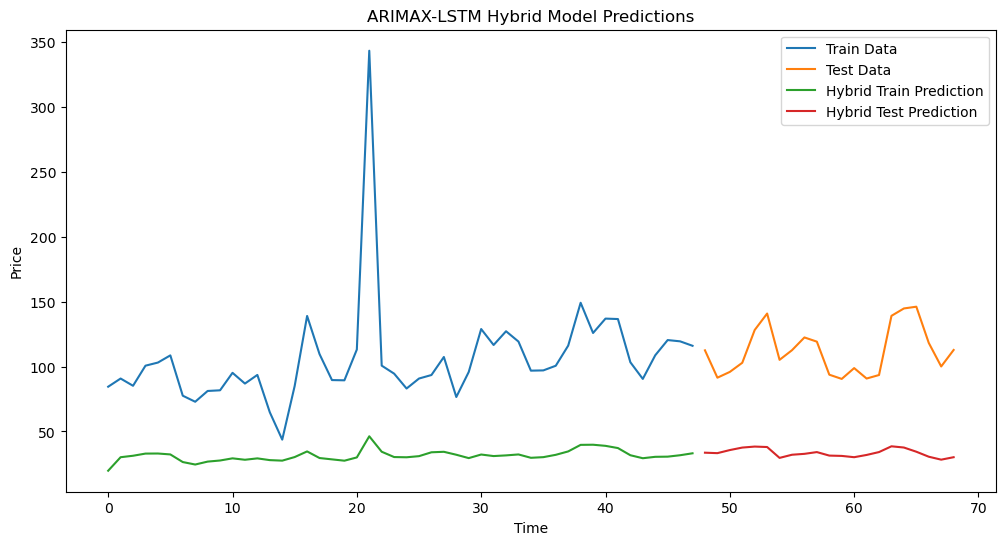
*Yt - Yt-1= ϕ*1*+ϵt* +*θ1ϵt-1*

****

1. **ARIMAX-LSTM Hybrid Model:**

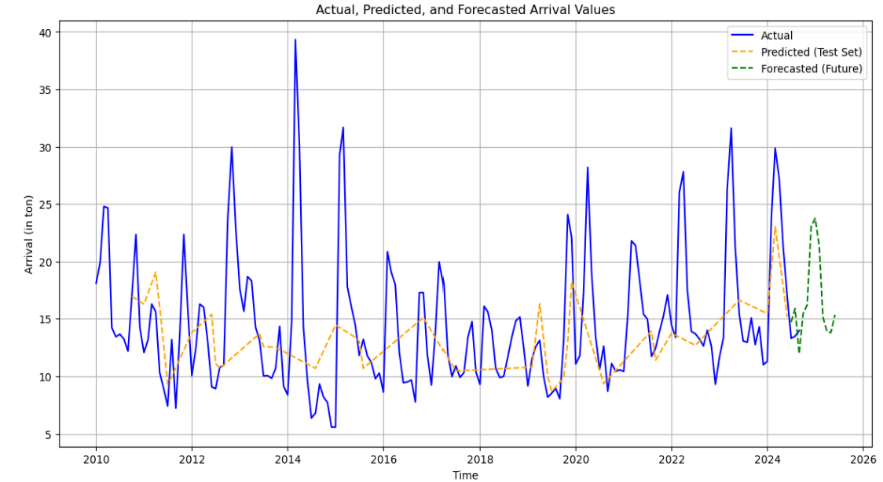
**For Arrival(in ton:**

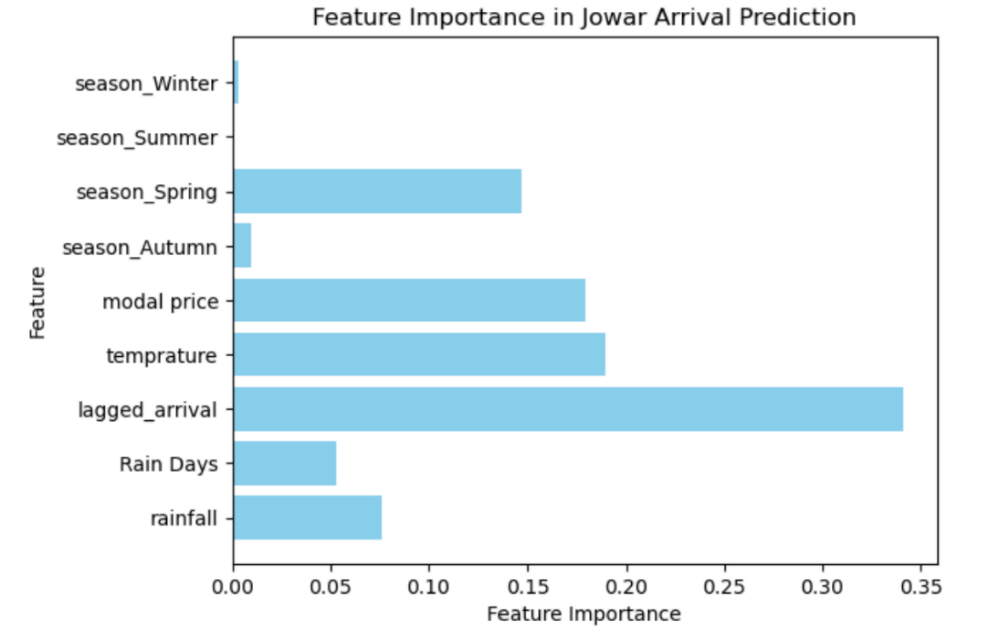
**Fig 9**

**For Price (in quintal):**

**Fig 10**

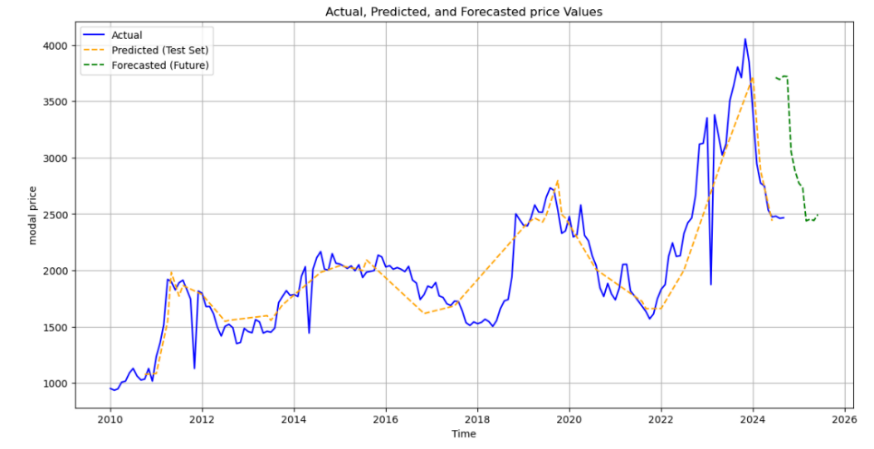
1. **Random forest:**

**For Arrival (in ton):**

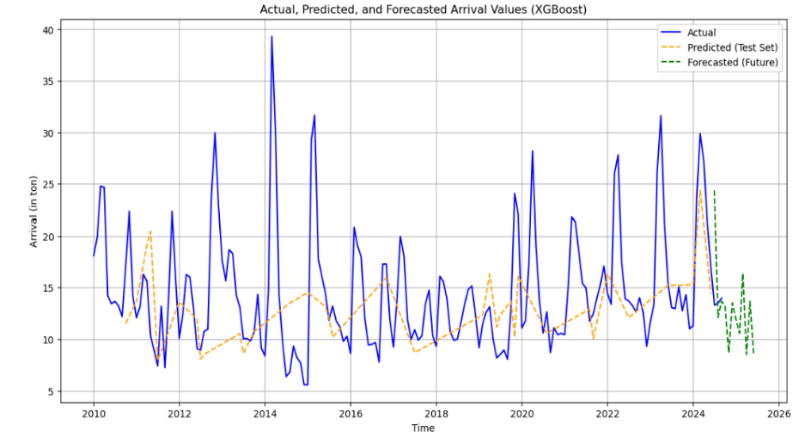


**Fig 12**

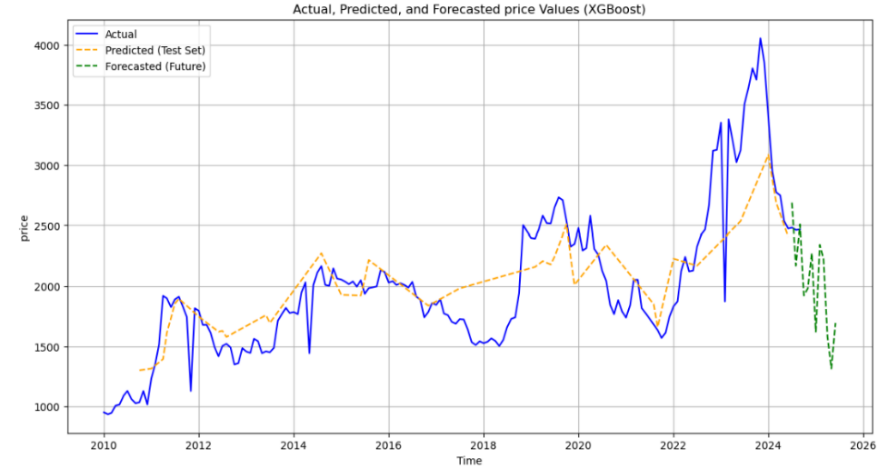
**Conclusion:** The featurelagged arrival has the highest importance means that past value of arrival significantly impacts the current arrival. Also, the temperature, spring season, modal price plays crucial role in determining future jowar arrival.

**For Price (in quintal):**

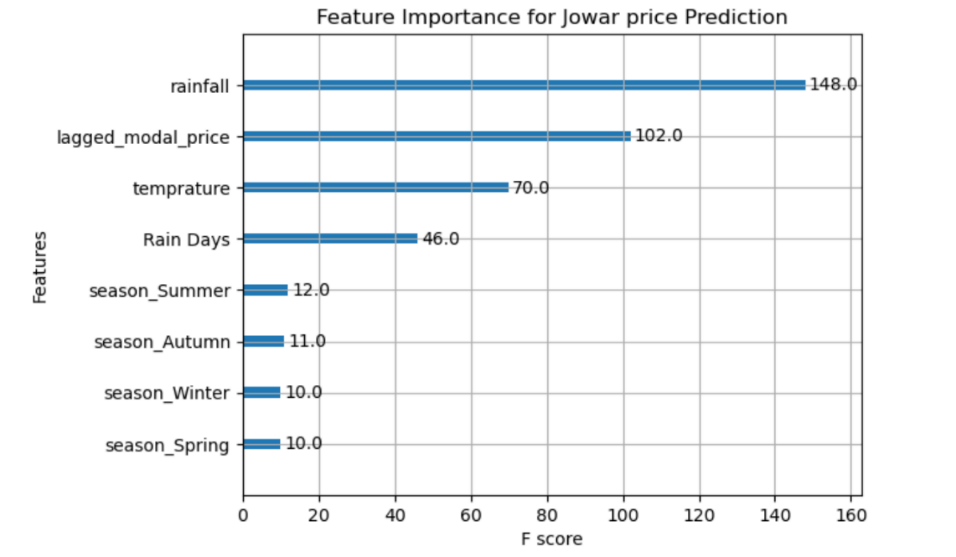
1. **XGBoost:**

**For Arrival(in ton):**

**For Price(in quintal):**

****

**Fig 15**

****

**Fig 16**

**Conclusion:** Rainfall, lagged modal price, and temperature are the dominant features affecting Jowar price prediction. Seasonal variations play a comparatively minor role in this model.

**Result:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model Name** | **Price** | | **Arrival** | |
|  | Train MAPE | Test MAPE | Train MAPE | Test MAPE |
| **Exponential smoothing** | 18.93% | 21.02% | 5.63% | 42.59% |
| **Arimma** | 7% | 35% | 23% | 18% |
| **Random forest** | 4.38% | 5.61% | 11% | 25% |
| **Xgboost** | 9.78% | 10.02% | 18.27% | 27.88% |
| **Hybrid (ARIMAX+LSTM)** | 28% | 19% | 44% | 46% |

Exponential Smoothing showed moderate accuracy for price prediction, with a small difference between train and test MAPE but may not suitable for arrival prediction producing high overfitting. ARIMA performed well for price prediction in the training data but showed significant overfitting with small test accuracy may not suitable for price forecasting. the discrepancy indicates that model may not be the most suitable for arrival prediction, as it performs well on the training data hence it is reasonably good for arrival.

Random Forest demonstrated excellent performance for price prediction, with low MAPEs, but showed signs of overfitting for arrival prediction. XGBoost become effective for price forecasting with low MAPE and it displayed overfitting for arrival prediction. The Hybrid model (ARIMAX+LSTM) had high MAPEs for both price and arrival predictions indicating less suitable for both arrival and price.

**Conclusion:**

The average arrival (in tons) shows fluctuations over the years. After a decline around 2017-2018, there has been a consistent increase in arrivals, which is high in recent years (2023-2024). The modal price (per quintal) has shown an upward trend over the years, with significant growth after 2020, indicating increased demand. Spring shows the highest variation and median values for arrivals, indicating a better season for crop supply.

Lagged arrival is the most critical feature, indicating past arrivals significantly influence current arrivals while Rainfall, lagged modal price, and temperature are key determinants for price predictions .Overall, **Random Forest** and **XGBoost** proved to be the most effective for price prediction, while **ARIMA** showed somehow better for arrival prediction. When viewed from the graphical point of view, there is very little difference between actual values and predicted values in Fig. 7,13 and 15. The **Hybrid model** (ARIMAX+LSTM) was generally less effective across both targets.

**Limitations:**

The limitations of the study were,

1. The study is limited to Jowar crops, making it less generalizable to other crops without significant modifications.
2. Complex seasonal patterns or irregular trends may not be fully captured by the time series models used.
3. Expanding the model to larger datasets or real-time systems may require additional computational optimization.

**Future Scope:**

This study can help market participants to plan for price and arrival. This could reduce the impact of price shocks on both farmers and consumers. It can provide ongoing updates to the agricultural community, enhancing the ability to manage risk effectively. It can help to farmers to predict future prices so they can plan properly to sell their crop. The scope and advantage of this study lie in its ability to provide actionable insights for farmers, traders, policymakers, and other stakeholders in the agricultural sector. It can help to enhance agricultural decision-making by providing accurate, data-driven forecasts for crop arrivals and prices. Its decision-support tool for farmers, advising them on the best time to plant, harvest, and sell their crops based on predicted arrival times and pricing trends.

**Reference:**

[1] Girish Hegde, Vishwanath R. Hulipalled, Jay B. Simha. Data driven algorithm selection to predict agriculture commodities price. *International Journal of Electrical and Computer Engineering (IJECE)*.August 2023, DOI: 10.11591/ijece.v13i4.pp4671-4682

[2] Khulood Albeladi, Bassam Zafar*,* Ahmed Mueen. Time Series Forecasting using LSTM and ARIMA. IJACSA), *International Journal of Advanced Computer Science and Applications,*2023.

[3] Samapriya Trivedi, Shalini Jaiswal, Ashok Kumar, and Shambhavi Mishra. Time series forecasting for market prices and arrival quantities of potatoes in Lucknow. *Journal of Applied Statistics & Machine Learning*, Number 1; 2023.

[4] Ranjit Kumar PaulID, Md. YeasinID, Pramod Kumar, Prabhakar Kumar. Machine learning techniques for forecasting agricultural prices: A case of brinjal in Odisha, India. *ICAR-Indian Agricultural Statistics Research Institute*,2022.

[5] Sourav Kumar Purohit, Sibarama Panigrahi ,Prabira Kumar Sethy and Santi Kumari Behera. Time Series Forecasting of Price of Agricultural Products Using Hybrid Methods. *Applied artificial intelligence*, 2021,<https://doi.org/10.1080/08839514.2021.1981659>

[6] Chu, C.W., Zhang, G.P.: A comparative study of linear and Nonlinear models aggregate retail sales forecasting,*International Journal of production economics*,2003.

[7] H V Ramachandra, Balaraju G, Rajashekhar A,Harish Patil ,Machine Learning application for Black Friday Sales Prediction Framework,*International Conference on Emerging Smart Computing and Informatics (ESCI),2023.*

[8] Ravikant Suryawanshi, Sandeep Musale, Dr. Snehal Bhosale. Comparative Analysis of use of Machine Learning Algorithm for Prediction of Sales, *J. Electrical Systems 20-3s* (2024).

[9] Tanujit Chakraborty, Ashis Kumar Chakraborty, Zubia Mansoor, hybrid regression model for water quality prediction, *OPSEARCH* ,<https://doi.org/10.1007/s12597-019-00386-z>.

[10] Salvatore Mancha Gonzales, Hasnain Iftikhar, and Javier Linkolk Lopez-Gonzales. Analysis and forecasting of electricity prices using an improved time series ensemble approach an application to the Peruvian electricity market. *AIMS Mathematics*, DOI: 10.3934/math.20241067.