Predictive Modelling for Crop Yield in Maharashtra: Analysing Trends and Future Insights

Anuja Shirke[a], Rohit Bhor[a], Mohini Kulkarni[a] and Tukaram Navghare[a*]
[a] Department of Statistics, B. N. Bandodkar College of Science, Thane, 400601, Maharashtra, India.
E-mail: 2023420367@vpmthane.org, 2023420368@vpmthane.org, mpkulkarni@vpmthane.org, tvnavghare@vpmthane.org*(Corresponding Author)*

Abstract :

Farming is the primary sector which is foremost contributor in Indian economy. India is the second country in farm produce globally. So, season wise predictions of farm yield is a major issue that needs to be solved based on available data to help farmers for their increased income. This study focuses on predicting crop yield trends in Maharashtra for Rabi and Kharif season to assist farmers to know about future trends and demands of the crops that can contribute to their increased income. Also, a study has been conducted on India's import and export of major crops to analyse and understand the agricultural market globally. Using secondary data on area, production and yield for crop, the statistical tools like ARIMA model, LSTM were applied for analysis. Results indicate that maize yield has a consistent upward trend and it is estimated to grow in future too about 24% from the year 2023 to 2032. While, Bajra and Jowar show declining trends due to reduced cultivation and abnormal rainfall. Jowar has decreased about 40%. Recommendations include promoting maize for industrial use, reviving the production for Bajra and Jowar, improving markets access for stable crops. The study emphasizes data-driven decisions to address yield variability and gives a marketing approach to farmers by knowing the demand of significant crops in the future. Also, the study on the import and export of crops provides valuable insights for developing effective marketing strategies for the farmers.

Keywords: ARIMA, Forecasting, GDP, Nutri-cereals, Global market

Introduction :

Agriculture is an important primary sector and any developing economy dependent primarily on agriculture has to undergo the process of agricultural development for accelerating the economic growth. The process of economic development of a country mainly depends upon as to how the agricultural sector is advancing. The performance of agricultural sector is therefore crucial in the process of economic development of a country [1,5]. The Crop yield plays a vital role in the agrarian economy. Major crops include Maize, Wheat, Rice, and Pulses with productivity influenced by monsoon variability and irrigation practices. Agriculture employs a large portion of the population & hence, crop yield significantly impacts farmers' livelihoods and the state economy, as fluctuations in yield directly affect rural incomes and the overall contribution of agriculture to GDP [3]. Prediction of yield is one of the most significant and well-known themes in accuracy farming as it characterizes yield planning and assessment, coordinating harvest supply with the request, and crop management [4,6,13].

Agriculture plays a vital role in food production, yet many nations face hunger due to population growth, climate variability, and soil degradation[17]. Ensuring sustainable crop growth and yield prediction is crucial for food security. Accurate yield forecasts help policymakers make informed decisions on imports and exports to enhance national food stability [8,19].Meteorologicalparameters, especially temperature, significantly impact climate and agriculture, making accurate forecasting crucial. Various studies compare SARIMA, ARIMA, and neural networks for temperature prediction, demonstrating their effectiveness in different regions[18]. Onion is a major spice crop, with India being the second-largest producer after China. A study using time series data (1978-2008) found a cubic trend in area, production, and productivity, with high instability in certain periods. Forecasts using the ARIMA model predicted India's onion production to reach 23.02 million tonnes by 2020, aiding policy decisions [10,13,15]. Agriculture producing a major portion of the world's food. However, many countries still face hunger due to food shortages and a growing population. Increasing food production is crucial to eliminating famine. Enhancing food security and reducing hunger by 2030 are key UN objectives. Thus, crop protection, land assessment, and yield prediction are essential for global food production. Accurate forecasts help policymakers make informed export and import decisions to strengthen national food security [7,11,12].

In 2018, the Government of India declared this year as "Nutri-Cereals" in recognition of the high nutritional value and anti-diabetic properties of millets. As one of the earliest domesticated crops, millets were first cultivated in India. This policy aimed to encourage farmers to expand millet cultivation, while many foreign countries are also shifting towards millet-based foods.

In India almost 60% of the population is engaged in agriculture and allied activities. Also, it contributes to almost 19% of India’s GDP. Maharashtra contributes about 13% to India’s GDP Still the farmers don’t get proper value for the crops they produce. Kharif and Rabi are two crops grown in India in two cycles each year. So, the aim is to aid farmers, policymakers & stakeholders to know about future trends and demands by forecasting the area, production and yield for various crops (season wise). Also, to study about the import and export of significant crops to understand the global market.

Related work :

Kumari, et.al.[14] analysed trends in India's food grain production (1995-96 to 2015-16) using official data, showing an annual growth rate of 1.73%, driven by minimal increases in area and productivity. Major cereals like rice, wheat, and maize showed consistent growth, while pulse production fluctuated. Maize yields improved significantly, and overall production and productivity trends were positive and statistically significant ($P<0.01$). Choudhari et.al.[5] estimates trends and analyzes growth in area, production, and productivity of selected crops in Western Maharashtra, revealing wide fluctuations and inconsistent trends across divisions. Lower productivity in 1994 was linked to significant land coverage by less productive crops. Results indicate lower productivity in Nasik and Pune divisions due to scarcity zones, while Kolhapur division showed moderate growth trends. Padhan [16] applies the ARIMA model to forecast annual productivity of 34 crops using data from 1950 to 2010, predicting values for the next five years. The model's accuracy is validated using criteria like Adj R^2, AIC, and MAPE, with tea showing the lowest MAPE and cardamom the lowest AIC values, highlighting their forecasting accuracy. Verma et.al.[21] evaluates crop yield trends in Marienborn, Germany using ARIMA models to predict yields for various crops, including wheat, barley and maize. Historical yield data spanning four to five decades were analyzed, and ARIMA models with specific parameters were fitted for each crop. Different ARIMA configurations (e.g., ARIMA (0,1,1) for wheat, ARIMA (1,1,0) for barley) were applied and results showed low deviations between predicted and actual yields, supporting ARIMA's suitability for short-term crop yield forecasting, except for oat and rapeseed. Badmus, et.al [2] studied the forecasts of maize cultivation area and production in Nigeria using the ARIMA model, based on 1970–2005 data. By 2020, maize production is projected at 9952.72 thousand tons and cultivated area at 9229.74 thousand hectares, highlighting the need for land conservation policies.). Gautam and Singh [9] analyzes the trends, contributions, and instability in area, production, and productivity of sorghum in Maharashtra. Sorghum, a key staple crop in semi-arid regions, saw a declining area under cultivation in Maharashtra (-1.18% per year) from 1965-66 to 2014-15, showing a decline in area but positive growth in production and yield. The findings highlight significant variability in area (19.18%), production (32.44%), and productivity (27.19%) during the study period. Rani, et.al.[19] analyzed cereals and commercial crops showed positive growth in productivity, while maize and pulses saw increased production due to supportive policies from 1997 to 2019 in Karnataka. In contrast, oilseeds experienced a decline in area (-5.87%) due to high costs and low prices.

Methodology :

In this paper statistical tools used like ARIMA (p, d, q) model – AutoRegressive Integrated Moving Average which is popularly used for time series forecasting, AIC & BIC Criteria, ACF/PACF, Box-Ljung test for the comprehensive data analysis. The well fitted ARIMA models for the parameters such as Area under Cultivation, Production and Yield are obtained from the significant spikes of their respective Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) plots.

ARIMA Model:

An autoregressive integrated moving average (ARIMA) is a generalization of an autoregressive moving average (ARMA) model. ARIMA model is classified as an ARIMA (p, d, q) model, where,

AR (p): Number of lag terms (past values), I (d): Number of times differencing is applied, MA (q): Number of past error terms used.

The general formula for an ARIMA (p, d, q) model is:

$$y_t = \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} + \cdots + \emptyset_p y_{t-p} + \emptyset_1 \varepsilon_{t-1} + \emptyset_2 \varepsilon_{t-2} + \cdots + \emptyset_q \varepsilon_{t-q} + \varepsilon_t$$

Where:

y_t: The actual value of the time series at time t,

$\emptyset_i$: Coefficients of the autoregressive terms (AR terms),

$\emptyset_j$: Coefficients of the moving average terms (MA terms),

ε_t: Error term at time t.

where, p is the number of autoregressive terms, d is the number of non seasonal differences needed for stationarity and q is the number of moving average terms. The model can be written as: (1 −B)dϕ(B)Xt = θ(B)Zt Where, ϕ(B) = 1−ϕ1B−⋯−ϕpBp is termed as autoregressive polynomial, θ(B) = 1+θ1B+⋯+θqBq is termed as moving average polynomial, and Zt is the white noise process.

AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) are both criteria used for model selection in statistics and machine learning. They help to compare different models by balancing goodness-of-fit and model complexity. In both AIC/BIC, lower values indicate a better model.

Figure 1 shows the overall method of analysis outlining key steps right from deciding the objective till the analysis and it's conclusion.This includes specifying the research question, hypothesis, or problem that needs to be addressed by gathering the secondary data.Process the collected data by handling missing values and ensuring accuracy.Perform statistical techniques to the data and analyzing it. This structured process ensures a systematic approach to data analysis, making it more efficient and effective.

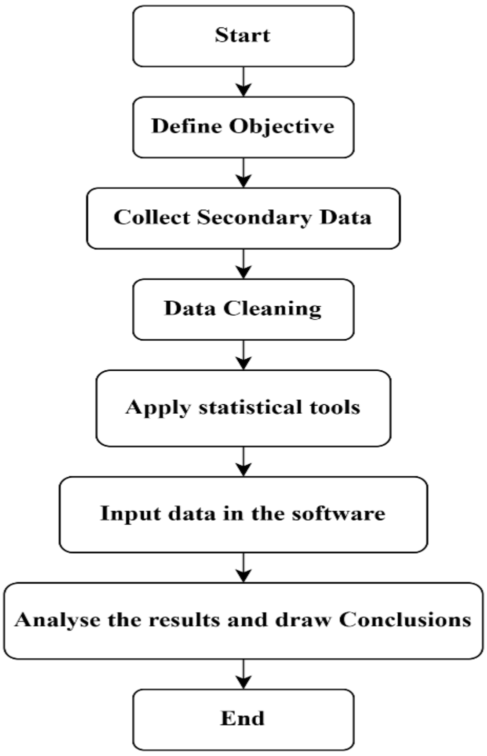

Fig 1. Flowchart of the study

Dataset and Software:

In this study, secondary data is extracted from Department of Economics and Statistics attached with the office of Agriculture (https://data.desagri.gov.in/website/crops-apy-report-web). Dataset includes parameters like Area, Production & Yield of crops in various states of India from the year 1998 – 2023. Crops in rabi season include Maize, Jowar, Gram & Wheat while crops in kharif season include Bajra, Rice, Sesamum, Niger seeds, Sunflower. Moreover, data for India's Export & Import of Maize and Jowar is taken from https://tradestat.commerce.gov.in&https://agriexchange.apeda.gov.in/product_profile/exp_f_india.aspx?categorycode=0604. Also Software like Python, R Studio, Tableau and MS Excel are used and advanced libraries and packages such as *tseries*&*Forecastlibraries*, *auto.arima* are applied to develop the time series forecasting. This combination of statistical methods, software tools, and visualizations provided an easier way of understanding the analysis of the data collected (for Laymans' too).

Result and discussion:

In this section we explored the result of forecasted area, production and yield for multiple crops. Out of the various crops analyzed, the ones listed below have demonstrated the best results and forecasting accuracy.

The results for Rabi season (crops include Maize, Jowar, Gram, Wheat and some other pulses) are shown below:

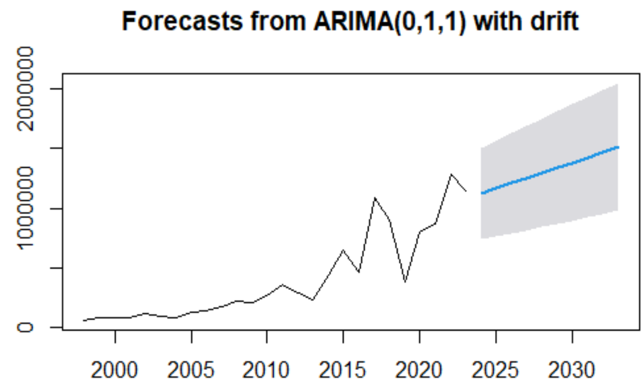

Fig 2. Maize Production Forecast

Figure 2 shows the forecast for maize production with ARIMA (0,1,1) model. The model suggests a steady increase in maize production implying improved agricultural output.

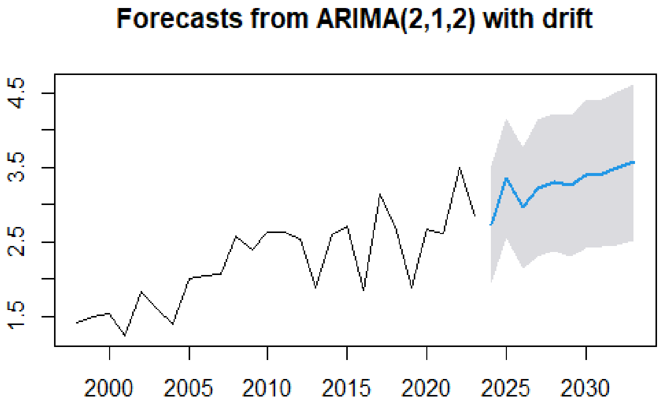

Fig 3. Maize Yield Forecast

Figure 3 shows the forecast for maize yield with ARIMA (2,1,2). The model indicates a consistent rise in maize production, suggesting enhanced agricultural productivity. If this trend continues as forecasted, maize yield is expected to grow consistently until 2032. This trend may be driven by advancements in farming techniques, better seed varieties, improved irrigation, and favourable government policies.

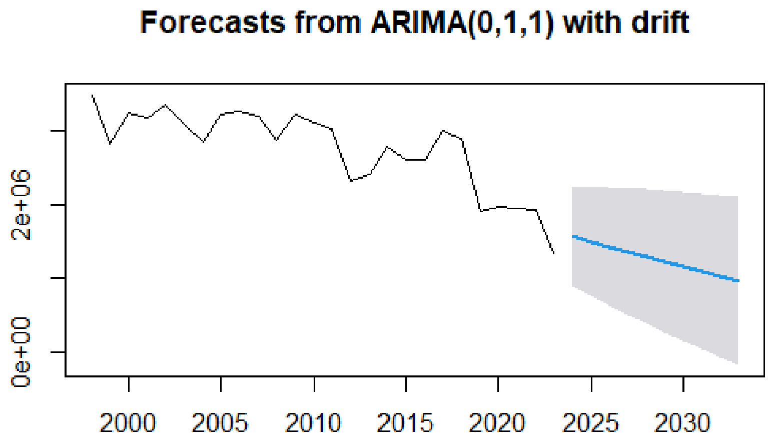

Fig 4. Jowar Area Forecast

Figure 4 shows the forecast for jowar area with ARIMA (0,1,1) model. From the year 1998 to 2024 the cultivation area for jowar is decreasing and the forecast predicts that the area for jowar will be declining in future. The decreasing cultivation area has directly affected to a decline in jowar yield over time. This trend raises concerns about food security and the sustainability of jowar production.

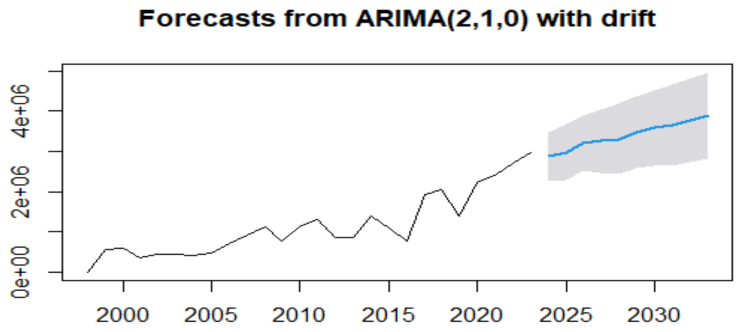

Fig 5. Gram Production Forecast

Figure 5 is the forecast for gram yield with ARIMA (2,1,0) model which is estimated to have a upward trend. The forecasted increase in gram yield suggests that this positive trend is crucial for food security and economic benefits as Gram is an important pulse crop that provides protein in our diet and helps improve soil health by adding nitrogen.

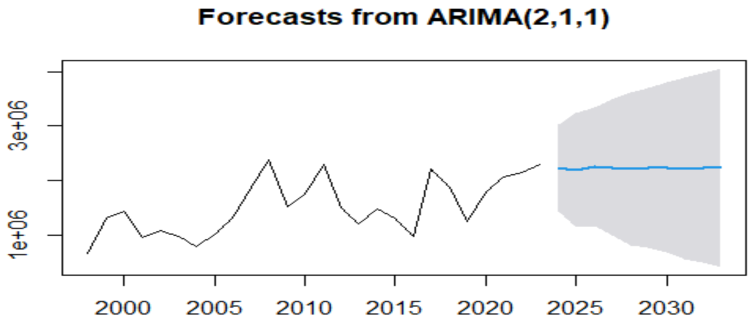

Fig 6. Wheat Yield Forecast

Figure 6 shows the forecast graph for wheat yield with ARIMA (2,1,1) model. It has a consistent trend. Jowar production has declined due to low consumption and demand and hence the export is also decreased. Also due to higher rates, people prefer wheat as it is a sustainable crop and has a consistent trend.

For Kharif season (crops include Bajra, Rice, Sesamum and other pulses). Following are the results for these crops forecasted by ARIMA model up to year 2032.

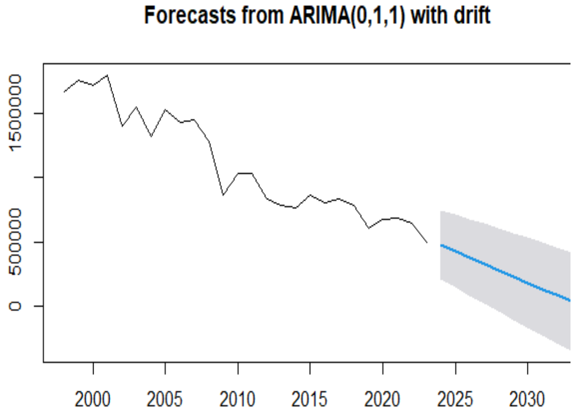

Fig 7. Bajra Area Forecast

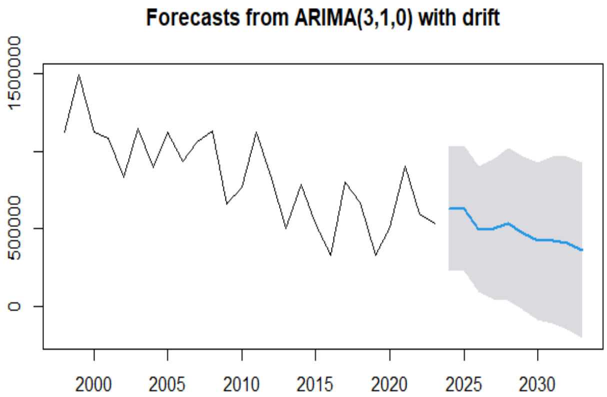

Fig 8. Bajra Production Forecast

Figure 7 and Figure 8 shows the forecast for Bajra area and production respectively. Both the forecast for Bajra cultivation area and production, shows a significant decline over time. The decline in Bajra area (Figure 7) directly impacts production levels (Figure 8), as less land is allocated for its cultivation. This reduction may also be influenced by soil degradation, urbanization, and limited government support for millet farming.

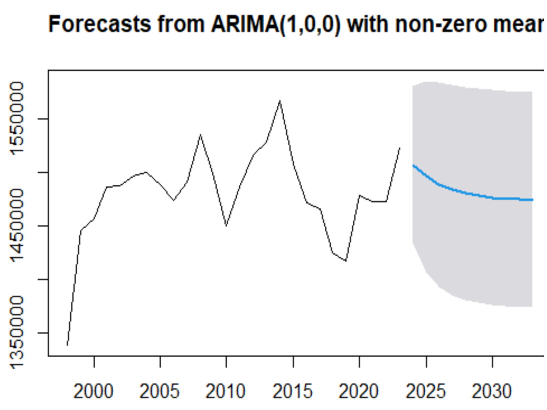

Fig 9. Rice Area forecast

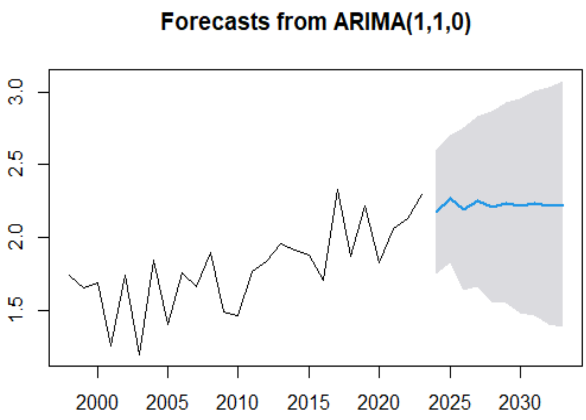

Fig 10. Rice Yield forecast

Figure 9 illustrates the forecast for rice cultivation area, showing a declining trend. However, Figure 10 shows that despite the declining cultivation area, rice yield has maintained a consistent horizontal trend, indicating that production efficiency has remained stable. If the declining area trend continues, then the yield may affect in future days.

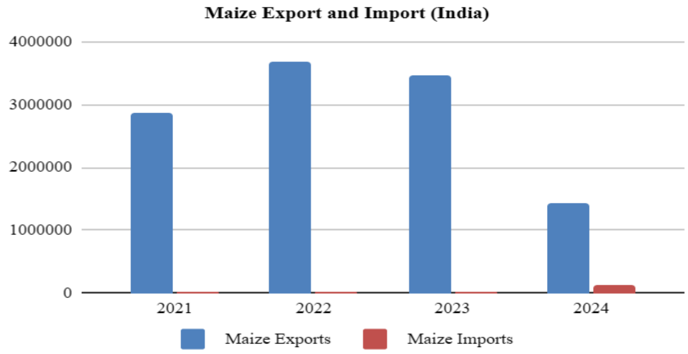

Fig 11. India's Maize export and import

In Figure 11, the import and export of maize in India from the years 2021 to 2024 are represented. Maharashtra is one of the major maize producing states in India which produces 12% of maize of India's export. One the main reason of good yield is suitable climatic patterns. However, India's maize exports fell to a four-year low in the 2023-24 period, experiencing a 58% decline in volume. This drop was due to higher domestic prices with increasing demand from ethanol, poultry, and starch manufacturers, fuel, etc. Also, Maize is a significant raw material in food industries like food sweeteners, beverages, colour pigmentation etc.

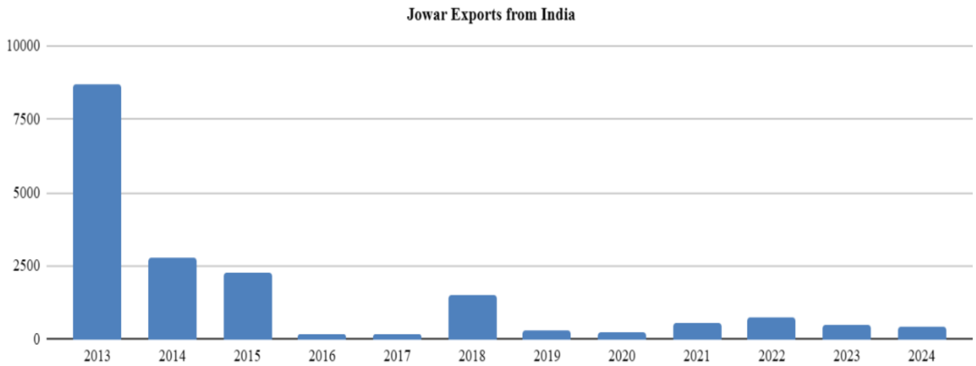

Fig 12. Jowar export from India

Figure 12. Represents Jowar export from India from the year 2013-2024. The declined jowar cultivation resulted into the decrease of India's jowar export from the year 2013 to 2017. However, as per the India's Government policy in 2018, they have been encouraging farmers with various schemes to elevate the cultivation and consumption of millets, the jowar export also increased in 2018. The UN declared 2023 as the International Year of the Millets, and the government has undertaken programs to boost its production and marketing.

Table 1. AIC/BIC values for ARIMA model

	ARIMA Model		
		AIC	BIC

Maize	Area	607.54	609.98
	Production	**684.14**	687.79
	Yield	**33.58**	40.9
Jowar	Area	**712.83**	716.48
	Production	754.75	758.52
	Yield	-17.23	-13.46
Gram	Area	703.21	706.87
	Production	**709.09**	713.97
	Yield	-9.09	-7.87
Wheat	Area	677.58	681.24
	Production	722.14	727.01
	Yield	722.14	727.01
Bajra	Area	667.06	670.72
	Production	**689.5**	695.59
	Yield	-6.44	-3.92
Rice	Area	625.97	629.74
	Production	**712.46**	714.9
	Yield	**-1.73**	0.71
Sesamum	Area	513.63	515.98
	Production	476.46	479.99
	Yield	-63.43	-62.25
Niger Seed	Area	484.13	486.49
	Production	452.43	453.6
	Yield	-62.76	-59.1
Sunflower	Area	533.95	539.84
	Production	522.79	525.15
	Yield	-32.52	-30.08

Table 1gives AIC/BIC values for the ARIMA method which evaluates the model in their complexity and good fit. The values in bold represent the significant results for the crops that have future scope to be analyzed further using the AIC and BIC criteria.

Conclusion :

In this paper, it is concluded that in Maharashtra, the Maize yield will increase about 24% from the year 2023 to 2032 & simultaneously Jowar yield has decreased by 40%. The area for Gram has a slight increase. Rice shows a good consistent trend. While all other crops are declining. Due to less production, cultivation area of crops will be declining. Cultivation area of Bajra has decreased from 18 Lakh (Hectare) to 33k (Hectare) over the period of 25 Years. Bajra yield rate is drastically declining due to it's low consumption and abnormal rainfall patterns. Jowar is a gluten-free and protein-rich millet. So, the production of Bajra & Jowar needs to be remarked as it is a staple food with nutritional benefits and can grow into industries.

In coming years, Farmers can increase the cultivation of maize as it has high demand in various industrial sectors (like feed, starch manufacturing, ethanol production) as future demand of maize is increasing which will contribute to the increased income of farmers. Moreover**,** Global export of maize is a good market planning that will contribute to India's GDP.

References :

1. Andreoni, A., & Postorino, M. N. (2006). A multivariate ARIMA model to forecast air transport demand. *Proceedings of the Association for European Transport and Contributors*, 1-14.
2. Badmus, M. A., & Ariyo, O. S. (2011). Forecasting cultivated areas and production of maize in Nigerian using ARIMA Model. *Asian Journal of Agricultural Sciences*, *3*(3), 171-176.
3. Bhimavarapu, U., Battineni, G., &Chintalapudi, N. (2023). Improved optimization algorithm in LSTM to predict crop yield. Computers, 12(1), 10.https://doi.org/10.3390/computers12010010
4. Chen, P., Yuan, H., & Shu, X. (2008, October). Forecasting crime using the ARIMA model. In *2008 fifth international conference on fuzzy systems and knowledge discovery* (Vol. 5, pp. 627-630). IEEE.https://doi.org/10.1109/FSKD.2008.222
5. Choudhari, S. W. A. T. I., Kumbhar, J. S., Amale, A. J., & Shinde, H. R. (2013). Study of growth analysis and trend of area, production and yield of selected crops grown in Western Maharashtra. *International Research Journal of Agricultural Economics and Statistics*, *4*(1), 10-17.
6. Dhal, S. B., Kalafatis, S., Braga-Neto, U., Gadepally, K. C., Landivar-Scott, J. L., Zhao, L., ... & Bhandari, M. (2024). Testing the Performance of LSTM and ARIMA Models for In-Season Forecasting of Canopy Cover (CC) in Cotton Crops. Remote Sensing, 16(11), 1906.https://doi.org/10.3390/rs16111906
7. Elavarasan, D., & Vincent, P. D. (2020). Crop yield prediction using deep reinforcement learning model for sustainable agrarian applications. *IEEE access*, *8*, 86886-86901.
8. Fattah, J., Ezzine, L., Aman, Z., El Moussami, H., & Lachhab, A. (2018). Forecasting of demand using ARIMA model. *International Journal of Engineering Business Management*, *10*, 1847979018808673. https://doi.org/10.1177/1847979018808673
9. Gautam, Y., & Singh, P. K. (2018). Economic analysis of sorghum in Maharashtra, India. *International Journal of Agricultural and Statistical Sciences*, *14*(2), 601-606.
10. Hamel, A., & Abdulwahhab, B. I. (2022). Time series Forecasting Using ARIMA model. *autoimmunity*, *9*, 1.
11. Jadhav, V., Chinnappa, R. B., & Gaddi, G. M. (2017). Application of ARIMA model for forecasting agricultural prices.
12. Kale, V. C., Rajmane, K. D., & Waghmare, R. R. (1998). Trends in Agricultural growth in Parabhani district of Maharashtra State. *J. Maharashtra Agric. Univ*, *12*(3), 361-363.
13. Khanna, A., & Kaur, S. (2022). Time Series Analysis of Environmental Data by using ARIMA and LSTM models for Precision Agriculture.https://doi.org/10.21203/rs.3.rs-2157455/v1
14. Kumari, N., Mehta, V. P., & Bhatia, J. K. (2020). Foodgrains production in India:

Trend and decompositions analysis. *Economic Affairs*, *65*(3), 333-342. http://dx.doi.org/10.46852/0424-2513.3.2020.3
15. Mishra, P., Sarkar, C., Vishwajith, K. P., Dhekale, B. S., & Sahu, P. K. (2013). Instability and forecasting using ARIMA model in area, production and productivity of onion in India. *Journal of Crop and Weed*, *9*(2), 96-101.
16. Padhan, P. C. (2012). Application of ARIMA model for forecasting agricultural productivity in India. *Journal of Agriculture and Social Sciences*, *8*(2), 50-56. 11–017/AWB/2012/8–2–50–56
17. Paidipati, K., & Banik, A. (2019). Forecasting of rice cultivation in India–a comparative analysis with ARIMA and LSTM-NN models. EAI Endorsed Transactions on Scalable Information Systems, 7(24).http://dx.doi.org/10.4108/eai.13-7-2018.161409
18. Phad, G. S., Muley, A. A. (2022). A Systematic Approach for Minimum and Maximum Temperature Prediction Using SARIMA Model, *Int. J. of Emerging Technologies and Innovative Research*, 9(7), b405-b413.
19. Rani, S. U., Kumar, P., Singh, N. P., Paul, R. K., Padaria, R. N., &Tadigiri, S. (2022). Trend and growth rate estimation of principal crops in Karnataka State in India. *Int. J. Plant Soil Sci*, *34*(5), 72-80. 10.9734/IJPSS/2022/v34i530867
20. Rashid, M., Bari, B. S., Yusup, Y., Kamaruddin, M. A., & Khan, N. (2021). A comprehensive review of crop yield prediction using machine learning approaches with special emphasis on palm oil yield prediction. *IEEE access*, *9*, 63406-63439.
21. Verma, U., Koehler, W., & Goyal, M. (2012). A study on yield trends of different crops using ARIMA analysis. *Environment and Ecology*, *30*(4A), 1459-1463.

~*~*~*~*~*~*~