Journal of Animal & Plant Sciences, 35(2): 2025, Page: 522-530
ISSN (print): 1018-7081; ISSN (online): 2309-8694
https://doi.org/10.36899/JAPS.2025.2.0043

FISH PRODUCTION MODELING AND FORECASTING IN INDIA USING THE XGBOOST ALGORITHM

S. Yadav[1], B. Kumari[2], A. M. G. Al khatib[3], Y. S. Raghav[4], D. Sharma[5], B. M. Alshaib[3], H. Nayak[6], S. Ray[7,*], T. Biswas[8], N. Mishra[9] and P. Mishra[10]

[1]Miranda House, Department of Geography, University of Delhi, India
[2]Department of Agricultural Economics, Rashtriya Kisan Post Graduate College, Shamli, Uttar Pradesh, India
[3]Department of Finance and Banking, Faculty of Economics, Damascus University, Damascus, Syrian Arab Republic
[4]Department of Mathematics, College of Science, Jazan University, P.O. Box. 114, Jazan 45142, Kingdom of Saudi Arabia
[5]Central Institute of Coastal Engineering for Fishery, Department of Fisheries, Ministry of Fisheries Animal Husbandry and Dairying. Govt. of India, 560013, India; Divya.ds38@gmail.com
[6]Department of Silviculture and Agroforestry, Odisha University of Agriculture and Technology, Bhubaneswar
[7]Department of Agricultural Economics and Statistics, Centurion University of Technology and Management, Paralakhemundi, Odisha, India
[8]Symbiosis Statistical Institute, Symbiosis International (Deemed University), Pune-411004, India
[9] Department of Mathematics and Statistics School of Physical Sciences, Amity University, Mohali, Punjab, India
[10]College of Agriculture, Rewa, Jawaharlal Nehru Krishi Vishwa Vidyalaya, Madhya Pradesh, India
Corresponding Author's email: raysoumik4@gmail.com

ABSTRACT

Time series analysis using machine learning is vital for forecasting in commodity sciences. This research leverages advanced machine learning models for time series forecasting of fish production at both state and national levels in India. The study developed and compared traditional models, like the autoregressive integrated moving average (ARIMA) and state space models, with the advanced machine learning model, extreme gradient boosting (XGBoost), using training and test data sets. Results showed that XGBoost and state space models significantly outperformed the ARIMA model. Specifically, XGBoost had the highest accuracy in eight of eighteen series, followed by state space models (seven out of eighteen), and ARIMA models (three out of eighteen). This confirms that applying diverse machine learning models can enhance forecasting accuracy for fish production. After identifying the best-performing models, forecasts for fish production were extended to 2030, indicating that India's total and marine fish production would likely continue to grow, with minimal change expected in key producing states. This data-driven analysis offers valuable insights for food security planning and policy-making in the region.

Keywords: Fish production, Time series, machine learning, forecasting, ARIMA, XGBoost.

Published first online March 12, 2025 Published final April 28, 2025

INTRODUCTION

Fish production plays a pivotal role in global food systems, offering essential protein, nutrients, and economic benefits to millions worldwide. As one of the fastest-expanding food sectors, it is crucial for meeting the nutritional demands of a growing global population. Fish and fishery products are not only a primary nutrition source for many but also provide essential nutrients like vitamins, minerals, and omega-3 fatty acids, all beneficial for human health. India ranks second globally in fish production, with the southern and much of the eastern population being the primary consumers. In 2020, the Indian fish market generated about INR 1.232 billion and is projected to grow by 10.5% to reach INR 2.243 billion by 2026 (Sharma *et al*., 2018).

India's aquaculture production encompasses both inland and marine sources, with inland production contributing over twice as much as marine production. Andhra Pradesh leads in inland fish production, while Gujarat ranks highest in marine fish production. This sector supports livelihoods and employment for vast rural communities in India. Fish provides consumers with a wealth of protein, healthy fats, long-chain omega-3 fatty acids, vitamin D, iodine, and various micronutrients. In a rapidly growing population like India, fish is considered the most affordable source of iron and animal protein, addressing nutritional insecurity. Fish accounts for about 20% of animal-derived protein in low-income, fooddeficient nations and 13% in industrialized countries (Delgado *et al*., 2002). Coastal aquaculture has emerged as a promising sector with significant potential for boosting exports and foreign exchange earnings (Krishnan and Birthal, 2002).

The demand for fish as a dietary staple has surged due to its high nutritional value. As a renewable natural resource, fish can replenish itself, but unsustainable harvesting practices could lead to depletion and even extinction. Sustainable harvesting strategies are essential to prevent this and ensure consistent production across years (Aanes *et al*., 2002). Consequently, accurate forecasting of fish production is crucial for policy-making to secure a surplus supply. Numerous researchers have explored methods for predicting fish production.

Yadav *et al*. (2013) conducted a comparative study on fish production forecasting in India, using artificial neural networks (ANN) and fuzzy time series models, concluding that ANN yielded better accuracy. In Tamil Nadu, Anuja *et al*. (2017) used ARIMA (Auto-Regressive Integrated Moving Average) and regression analysis for marine fish production forecasting, finding ARIMA slightly more precise. Raman *et al*. (2017) used the ARIMA model to forecast marine fish production in Odisha from 2015 to 2018. Similarly, Boruah *et al*. (2020) analyzed data from 1978 to 2018, concluding that total fish production increased at a higher rate than marine fish production in India, with ARIMA (0,2,1) and ARIMA (2,1,4) being the best models for marine and total fish production, respectively.

In another study, Yadav *et al*. (2020) applied time series data from 1980-81 to 2014-15 to predict fish production in Assam using an ARIMA model validated with data from 2015-16 to 2018-19, where ARIMA (1,1,0) was identified as the best fit, showing a rising production trend. Mishra *et al*. (2021) evaluated fish production in India using ARIMA, Holt's Linear, BATS, and TBATS models, concluding that Holt's Linear model was ideal for marine fish forecasting, while ARIMA (2,2,1) and ARIMA (3,2,0) were best for inland and total fish production forecasts. In Andhra Pradesh, Stephen *et al*. (2022) employed Non-linear Autoregressive (NAR) ANN, ARIMA, and Empirical Mode Decomposition-based ANN (EMD-ANN) models for fish catch prediction.

Having the importance of this context, we tried to build the time series model using statistical and machine learning models to predict the fish production nature in marine and inland condition of all major states in India. We firmly believe that the results of this study will be useful to stakeholders, policymakers, and researchers in making decisions on fish production policies.

MATERIALS AND METHODS

The time series information from 1993 to 2022 collected from Handbook on fisheries statistics, published by department of fisheries, Ministry of fisheries, animal husbandry and dairying were used to analysis in this study. The data was then divided into a training set consisting of 90% of observations from 1993-2019, and a validation set comprising the remaining 10% for 2020-2022. The training set was used to fit and estimate parameters of different forecasting models, while the validation s*et al*lowed for an unbiased evaluation of their predictive performance.

Coming to the methodology context, the following modeling approaches were utilized,

Autoregressive Integrated Moving Average (ARIMA) models: ARIMA is a well-established statistical modeling framework for univariate time series data. It incorporates autoregressive and moving average components applied to differenced data to capture autocorrelation patterns and achieve stationarity. Model identification and parameter estimation were conducted using maximum likelihood method (Abotaleb *et al*., 2021; Al khatib *et al*., 2021; Mishra *et al*, 2020; Mishra *et al*., 2021; Mishra *et al*,.2023; Mohammed *et al*,.2021; Raghav *et al*., 2022; Rahman *et al*,.2022; Ray *et al*., 2023a; Ray *et al*., 2023b; Yonar *et al*., 2022;).Generally, ARIMA model is followed by three parameters; *p* (order of Auto regressive (AR) term) *d* (order of differencing) and *q* (order of Moving average (MA) term). The simple equation of ARIMA model defined as;

$$\Delta^d y_t = c + \sum_{i=1}^{p} \phi_i \Delta^d y_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t,$$

where, y_t is the time series information at time *t*; *c* is the constant; $\Delta^d y_t$ is the stationary series after differencing; ϕ_i is the AR coefficients; θ_j is the MA coefficients; ϵ_t is the white noise error followed as $\epsilon_t \sim WN(0, \sigma^2)$ at time *t*.

State Space Models (SSM): SSMs provide a flexible structural approach to decompose time series into various unobserved components like trend, seasonality and noise. The Kalman filter was employed for efficient parameter estimation in the SSM framework (Yadav *et al*., 2022) and (Niranjan *et al*., 2022). The simple state space model equation is followed as state equation and observation equation denoted as,

$$x_t = Ax_{t-1} + Bu_t + w_t \text{(State equation)},$$

where, x_t is the state vector at time *t*; *A* is the state transition matrix; *B* is the control metrix; u_t is the control input vector and w_t is the residual covariance.

$$y_t = Cx_t + Du_t + v_t \text{ (Observed equation)},$$

where, y_t is the observed vector at time *t*; *C* is the observed matrix; *D* is the input effect matrix and v_t is the observed residual covariance.

Combinedly the state estimation followed as,

$$\hat{x}_t = A\hat{x}_{t-1} + k_t(y_t - C\hat{x}_t)$$

where, k_t is the Kalman filter.

Extreme Gradient Boosting (XGBoost): As an ensemble machine learning technique, XGBoost builds a collection of decision trees from the training data in an iterative manner. It can inherently capture nonlinear relationships and handle outliers.

Mathematically, the algorithm can be expressed as by the following steps;

$X_t = [y_{t-1}, y_{t-2}, \dots\dots.., y_{t-n}]$; estimating the lagged feature to the target point y_t in the time series.

Train the series with normalized the function;

$L(\theta) = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \sum_{k=1}^{K}\Omega(f_k)$; where $\Omega(f_k) = \gamma T + \frac{1}{2}\lambda\sum_{j=1}^{T} w_j^2$ (regularization term)

Then the prediction equation; $\hat{y}_t = \sum_{k=1}^{K} f_k(X_t)$;

Next, the model should be updated with gradients and Hessians;

$$g_i = \frac{dL(y_i, \hat{y}_i)}{d\hat{y}_i}$$

$$h_i = \frac{d^2L(y_i, \hat{y}_i)}{d\hat{y}_i^2}$$

Now, the best gain should be determined by splitting

$$Gain = \frac{1}{2}\left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} + \frac{(G_L + G_R)^2}{H_L + H_R + \lambda}\right] - \gamma$$

The most suitable model configurations from each approach were selected based on metrics like Akaike information criterion (AIC), mean absolute error (MAE) and mean square error (MSE) evaluated on the training set. Forecasts from the best performing ARIMA, SSM and XGBoost models were then compared based on their accuracy on the validation samples. This allowed for a comprehensive evaluation of various forecasting techniques for agricultural production time series (Chen & Guestrin, 2016). The total analysis was performed in R open source platform (https://cran.r-project.org/bin/windows/base/).

RESULTS AND DISCUSSION

Before modeling the time series information, it is crucial to check the data's behavior. So, first, summary statistics were evaluated to describe the series (Table 1).

The table 1 presents summary statistics for fish production data across various states and regions in India from 1993 to 2022. The table is divided into two sections. The first section reports central tendency measures: mean fish production, indicating the average fish production for each variable over 30 years, the median which is the middle value when production data is arranged in ascending order, and the minimum and maximum show the lowest and highest production levels recorded for each variable. This information provides the enough evidence of the central measurement of the data series. The second section describes the data's distribution and variability: standard deviation shows how much production values fluctuate from the mean. Coefficient of variation expresses variability relative to the mean. Skewness indicates the asymmetry in the distribution of production values, with positive skewness indicating a longer right tail (more high values) and negative skewness indicating a longer left tail (more low values), and excess kurtosis shows the "peakedness" of the distribution; a positive value means the distribution has a higher peak and fatter tails than normal (more extreme values), while a negative value suggests a flatter, less extreme distribution. One can determine about the variation of the data series from these measurements. The table 1 highlights production patterns and variability across different regions. Andhra Pradesh, for instance, showed the highest mean 416.25 and maximum 1393.7 in inland fish production, but also the highest standard deviation 401.57, indicating significant fluctuation around the average. In contrast, India Inland had a much lower mean 31.467 and standard deviation 4.1623, reflecting a smaller scale and more stable production over time. West Bengal Marine showed negative skewness (-0.80129), indicating a longer left tail (more low production values), while India Marine had positive skewness (0.9783), showing a longer right tail (more high production values). Excess kurtosis values were negative for all series, suggesting flatter distributions with fewer extreme values than a normal distribution. These descriptive statistics provide a comprehensive overview of the central tendency, dispersion, and shape of the fish production data for each region. Andhra Pradesh has the highest mean, median, minimum, and maximum fish production among all states, with West Bengal and Gujarat following. India as a region has the lowest mean, median, minimum, and maximum values, with Karnataka and Odisha just above. Andhra Pradesh inland shows the highest coefficient of variation, indicating substantial fluctuation over time. India inland has the lowest standard deviation and coefficient of variation, showing minimal fluctuation over time. India marine displays the highest skewness and excess kurtosis, reflecting a highly asymmetric and peaked production distribution. West Bengal marine has the lowest skewness and excess kurtosis, suggesting a more symmetric and flat distribution.

Table 2 displays the most appropriate ARIMA models for the training data based on the following evaluation criteria: Akaike information criterion (AIC), root mean squared error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), in-sample mean squared error (MSE), and the number of significant coefficients. These criteria assess the fit and prediction performance of the models.

After building the ARIMA model, state space model also employed of each training series (Table 3).

The most suitable model for all the time series data was selected based on the same goodness of fit criteria.Table 4 showed that XGBoost outperformed the other models on the training dataset, based on the evaluation criteria used in this study.

Table 1: Summary Statistics for fish production lakh tonnes (LT)*.

Variable	Mean	Median	Minimum	Maximum
Andhra Pradesh Inland	416.25	243.57	15.03	1393.7
Andhra Pradesh Marine	159.3	153.19	4.38	433.28
Andhra Pradesh Total	575.55	391.85	19.78	1808.1
West Bengal Inland	695.11	804.76	13.92	1338
West Bengal Marine	121.13	162	1.55	197.11
West Bengal Total	816.24	972.51	15.8	1517
Karnataka Inland	81.919	86.805	1.59	199.05
Karnataka Marine	156.4	183.58	3.47	357.32
Karnataka Total	238.32	282.38	5.55	546.43
Odisha Inland	131.15	141.03	2.94	291.83
Odisha Marine	87.465	118.85	1.2	156.08
Odisha Total	218.62	269.1	4.14	410.14
Gujarat Inland	49.57	55.385	0.98	142.85
Gujarat Marine	457.22	620.15	6.83	745.71
Gujarat Total	506.79	672.8	7.94	816.51
India Inland	31.467	29.84	25.76	41.27
India Marine	50.712	40.26	17.89	121.21
India Total	82.179	69.98	43.65	162.48
Variables	Std. Dev.	C.V.	Skewness	Ex. kurtosis
Andhra Pradesh Inland	401.57	0.96474	0.8258	-0.40242
Andhra Pradesh Marine	124.73	0.78295	0.26856	-0.62009
Andhra Pradesh Total	521.06	0.90532	0.70729	-0.48619
West Bengal Inland	495.84	0.71333	-0.35555	-1.3346
West Bengal Marine	80.403	0.66375	-0.80129	-1.2633
West Bengal Total	571.85	0.70059	-0.45498	-1.3329
Karnataka Inland	62.338	0.76098	0.0345	-1.0636
Karnataka Marine	113.35	0.72478	-0.15828	-0.99759
Karnataka Total	171.99	0.72168	-0.15135	-0.93889
Odisha Inland	94.971	0.72412	-0.17318	-1.2375
Odisha Marine	57.879	0.66173	-0.79401	-1.2454
Odisha Total	148.43	0.67895	-0.55996	-1.2926
Gujarat Inland	37.613	0.75877	0.11034	-0.54548
Gujarat Marine	302.69	0.66202	-0.81101	-1.2545
Gujarat Total	335.3	0.66161	-0.80812	-1.2581
India Inland	4.1623	0.13227	0.61688	-0.70496
India Marine	29.547	0.58265	0.9783	-0.13947
India Total	33.468	0.40725	0.9215	-0.26961

* 1.0 lakh tonne = 100,000 tonnes

Table 2: ARIMA Model fitted for State-wise Fish Production and India on training data set (1993-2019).

State	MODEL	Akaike Information Criterion (AIC)	Bayesian Information Criterion (BIC)	MAE	RMSE	MAPE	in-sample MSE
Andhra Pradesh Inland	ARIMA(2,0,2)	351.86	359.63	146.24	330.68	624.00	1.0935e+05
Andhra Pradesh Marine	ARIMA(1,0,0)	321.83	324.42	45.18	85.55	337.77	7318.77
Andhra Pradesh Total	ARIMA(1,0,0)	398.48	401.07	192.19	356.16	318.45	1.2685e+05
West Bengal Inland	ARIMA(1,0,0)	382.59	385.19	121.13	259.49	352.91	67335.59
West Bengal Marine	ARIMA(1,0,0)	268.95	271.55	14.13	30.87	304.36	952.71
West Bengal Total	ARIMA(4,0,3)	351.41	361.78	167.95	343.16	843.24	1.1776e+05
Karnataka Inland	ARIMA(1,0,0)	275.73	278.33	20.46	36.13	313.25	1305.61
Karnataka Marine	ARIMA(1,0,0)	313.85	317.74	43.44	72.08	532.98	5195.84

Karnataka Total	ARIMA(1,0,0)	334.59	337.18	51.98	106.10	344.45	11257.95
Odisha Inland	ARIMA(1,0,0)	300.58	303.17	29.71	57.36	362.47	3289.93
Odisha Marine	ARIMA(0,1,0)	243.64	244.94	10.81	25.54	396.36	652.45
Odisha Total	ARIMA(1,0,0)	319.14	321.73	37.01	79.85	365.53	6375.6
Gujarat Inland	ARIMA(1,0,0)	257.2	261.08	17.8	25.5	558.84	650.45
Gujarat Marine	ARIMA(0,1,0)	336.17	337.47	70.24	152.36	401.36	23214.19
Gujarat Total	ARIMA(1,0,0)	358.39	360.98	82.6	162.82	370.05	26509.82
India Inland	ARIMA(0,2,1)	85.155	87.746	1.051	1.28	3.486	1.639
India Marine	ARIMA(0,2,1)	104.27	108.157	1.416	1.781	3.086	3.174
India Total	ARIMA(0,2,1)	114.562	117.154	1.765	2.273	2.434	5.165

Table 3: State Space Models fitted for State-wise Fish Production and India on training data set (1993-2019).

State	Akaike Information Criterion (AIC)	Bayesian Information Criterion (BIC)	MAE	RMSE	MAPE	in-sample MSE
Andhra Pradesh Inland	372.94	376.82	117.83	280.69	383.14	78788.55
Andhra Pradesh Marine	310.33	312.92	34.05	87.53	368.37	7661.29
Andhra Pradesh Total	386.45	390.34	146.68	364.06	375.57	1.3254e+05
West Bengal Inland	369.6	373.49	87.47	263.34	376.77	69346.92
West Bengal Marine	256.99	259.58	12.95	31.38	312.94	984.75
West Bengal Total	373.07	375.66	90.64	292.5	362.24	85558.44
Karnataka Inland	265.21	267.8	18.16	36.76	335.94	1351.18
Karnataka Marine	305.03	308.92	33.61	76.07	396.16	5786.54
Karnataka Total	321.67	324.26	45.65	108.86	369.09	11851.28
Odisha Inland	291.09	294.97	21.88	58.17	406.17	3384.15
Odisha Marine	247.49	251.38	11.16	25.15	403.5	632.72
Odisha Total	308.34	312.22	28.66	81.06	399.15	6570.37
Gujarat Inland	249.95	253.84	15.65	26.36	473.5	694.68
Gujarat Marine	339.93	343.82	70.64	148.79	427.02	22139.89
Gujarat Total	345.47	349.36	75.37	165.54	416.99	27404.18
India Inland	88.515	91.107	1.013	1.229	3.314	1.511
India Marine	149.321	151.913	3.05	3.958	6.259	15.664
India Total	156.932	159.524	3.698	4.582	4.505	20.992

Table 4: XGBoost Machine learning Method fitted for State-wise Fish Production and India on training data set (1993-2019).

State	XGBoost Method			
	MAE	RMSE	MAPE	in-sample MSE
Andhra Pradesh Inland	4.43	8.407	2.618	70.670
Andhra Pradesh Marine	11.635	18.206	5.134	331.453
Andhra Pradesh Total	7.901	14.584	3.025	212.694
West Bengal Inland	0.6123	0.9530	0.3523	0.9082
West Bengal Marine	0.5414	0.7962	0.5996	0.6339
West Bengal Total	0.9253	1.7715	0.2818	3.1382
Karnataka Inland	3.561	4.844	16.802	23.463
Karnataka Marine	14.63	21.39	7.977	457.517
Karnataka Total	7.479	9.12	8.53	83.182
Odisha Inland	1.01	1.752	3.134	3.071
Odisha Marine	1.472	2.229	7.896	4.967
Odisha Total	2.556	4.326	4.894	18.711
Gujarat Inland	4.798	8.461	11.54	71.586
Gujarat Marine	16.21	22.569	6.038	509.351
Gujarat Total	21.413	29.575	5.873	874.692
India Inland	0.272	0.3481	0.8885	0.1211
India Marine	0.041	0.0991	0.0739	0.0098
India Total	0.1739	0.2379	0.24	0.0566

After developing the all three models, based on our objectives a comparison performance was employed. Table 5 compares the performance of ARIMA models, state space models, and XGBoost model. The evaluation of XGBoost on the validation sample was based on four criteria: root mean squared error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and mean absolute scaled error (MASE). Lower values across these metrics indicate better predictive accuracy. Results showed that the ARIMA model outperformed the state space model and XGBoost for predicting three specific time series: marine fish production in Odisha, marine fish production in India, and total fish production in India, with notably lower error values across all criteria, indicating high prediction accuracy.

However, Al Khatib *et al*. (2021) warned that newer forecasting methods do not necessarily outperform traditional methods in every case, as several factors (data frequency, complexity, number of observations, seasonality, cyclic variations, stationarity, trend, forecast horizon, and randomness) can impact forecasting accuracy.

XGBoost showed superior performance over ARIMA and state space models in predicting eight time series, including marine fish production in Andhra Pradesh, West Bengal, and Karnataka, as well as various inland and total fish production series in states like Gujarat and Odisha. However, while XGBoost performed well on the training data for all series, it showed signs of potential overfitting, which may limit its generalizability to the validation sample for some series.

Conversely, the state space model performed better than both ARIMA and XGBoost in predicting seven time series: inland and total fish production in Andhra Pradesh, West Bengal, Karnataka, and Odisha. This model may have captured complex, nonlinear aspects of the data, including outliers, structural changes, and time-varying factors, which ARIMA cannot handle as effectively due to its requirement for stationary data. State space models are advantageous for dynamic time series with unobserved components and are more adaptable to structural breaks and missing data.

Mishra *et al*. (2023b) argued that the limitations of ARIMA stem from its linear assumptions, which can hinder accuracy when data exhibit nonlinear interactions. They recommended nonlinear modeling approaches for cases where linearity is insufficient, as enforcing a linear model on nonlinear data can reduce prediction accuracy. This suggests that model selection should be based on the data's underlying features rather than a strict preference for linear models.

Ultimately, the best model is the one that generalizes well to new data, as seen in the validation sample, rather than merely fitting the training data. In this study, XGBoost emerged as the top performer in eight of the eighteen series, followed by state space models in seven series, and ARIMA in three series. These findings underscore the value of evaluating multiple models on both training and validation sets to select the most appropriate approach for a given dataset.

Table 6 shows the forecasted fish production (lakh tonnes (LT)). for five major fish producing states and India from 2023 to 2030, using the best performing predictive models.

The forecasted fish production values for major fish-producing states in India—Andhra Pradesh, West Bengal, Karnataka, Odisha, and Gujarat—as well as for India as a whole, from 2023 to 2030, are presented in lakh tonnes. These forecasts, generated using the best-performing predictive models (ARIMA, state space, or XGBoost) identified in the study, offer insights for food security planning and policy-making. In Andhra Pradesh, inland fish production is expected to remain stable at 41.9 LT annually, while marine production holds steady at 5.8 LT each year, resulting in a consistent total of 47.9 LT per year from 2023 to 2030. Similarly, West Bengal's inland production is forecasted to stay at 16.5 LT annually, with marine production remaining constant at 1.64 LT, leading to a stable total of 18.4 LT per year over the same period. Karnataka's inland production is projected to hold steady at 4.8 LT, marine production at 4.386 LT, and total production at 10.74 LT annually from 2023 to 2030, showing no significant variation. In Odisha, inland fish production is expected to remain at 7.82 LT, marine production at 2.01 LT, and total production at 8.37 LT per year throughout the forecast period. Gujarat's inland production is forecasted to stay constant at 1.65 LT, marine production at 6.83 LT, and total production at 8.58 LT annually from 2023 to 2030. At the national level, India's inland fish production is projected to remain stable at 37.75 LT each year, while marine production shows steady growth, increasing from 129.5 LT in 2023 to 188 LT in 2030. Consequently, India's total fish production is expected to rise consistently, starting at 172.9 LT in 2023 and reaching 246.5 LT by 2030, reflecting moderate growth in national fish production. These forecasts indicate that major fish-producing states are likely to maintain relatively stable inland and total production levels through 2030, with minimal year-to-year changes, while national marine and total production are poised for moderate growth, providing valuable data for stakeholders and policymakers to support sustainable fisheries management and food security in India over the next eight years.

Table 5: RMSE, MAE, MAPE and MASE for validation sample (2020-2022).

	ARIMA Models				State Space Models				XGBoost			
	RMSE	MAE	MAPE	MASE	RMSE	MAE	MAPE	MASE	RMSE	MAE	MAPE	MASE
Andhra Pradesh Inland	32.43	30.47	78.95	0.2801	3.5113	3.2972	8.34	0.0303	5.55	4.9054	11.98	0.0451
Andhra Pradesh Marine	0.6508	0.5789	10.02	0.0170	0.316	0.2867	4.974	0.0084	0.1956	0.1878	3.3172	0.0055
Andhra Pradesh Total	8.04	7.95	17.49	0.0573	3.2776	3.0753	6.77	0.0222	5.03	4.2622	9.07	0.0308
West Bengal Inland	1.1131	1.0786	6.54	0.0125	0.3032	0.2243	1.3482	0.0026	0.3961	0.3374	2.0334	0.0039
West Bengal Marine	0.237	0.1621	8.76	0.0125	0.2129	0.1467	8	0.0113	0.1622	0.1256	7.02	0.0097
West Bengal Total	163.23	151.75	831.03	1.6741	0.2661	0.2033	1.1112	0.0022	0.4212	0.3407	1.8567	0.0038
Karnataka Inland	1.4606	1.1308	29.76	0.0623	1.3186	0.9567	23.99	0.0527	1.1815	0.9969	28.52	0.0549
Karnataka Marine	40.25	40.23	952.91	1.2072	1.4307	1.0317	20.05	0.031	1.0372	0.9255	20.26	0.0278
Karnataka Total	2.9957	2.0704	21.72	0.0454	2.706	1.78	18.1	0.039	2.1424	2.019	25.72	0.0442
Odisha Inland	1.1568	1.1374	15.79	0.0547	0.6977	0.6721	9.3	0.0323	1.043	0.8934	11.98	0.043
Odisha Marine	0.186	0.1467	7.73	0.0133	0.1907	0.1479	7.77	0.0135	0.1791	0.1589	8.88	0.0145
Odisha Total	1.2761	1.239	13.7	0.0451	0.8535	0.8062	8.85	0.0293	0.9085	0.6903	7.27	0.0251
Gujarat Inland	14.17	14.17	854.17	0.9432	0.1973	0.1641	9.45	0.0109	0.1356	0.1206	6.98	0.008
Gujarat Marine	0.1085	0.0833	1.2158	0.0012	0.1051	0.0761	1.1112	0.0011	0.1061	0.0758	1.0856	0.0011
Gujarat Total	0.5245	0.4769	5.52	0.0065	0.2382	0.2279	2.6467	0.0031	0.1359	0.1143	1.3364	0.0016
India Inland	4.0819	3.6253	9.38	3.5799	4.0934	3.4267	8.79	3.3837	2.6808	2.3321	6.15	2.3029
India Marine	0.814	0.7795	0.6997	0.2555	8.03	8	7.09	2.6237	17.38	15.96	13.85	5.23
India Total	5.34	4.4737	2.8828	1.2099	9.97	8.92	5.79	2.4114	17.69	15.34	9.89	4.148

Table 6: Forecasting from 2023 to 2030 using best forecasting models for State-wise Fish Production and India.

	Andhra Pradesh			West Bengal			Karnataka			Odisha			Gujarat			India		
Year	Inland	Marine	Total	Inland	Marine	Total	Inland	Marine	Total	Inland	Marine	Total	Inland	Marine	Total	Inland	Marine	Total
2023	41.9	5.8	47.9	16.5	1.64	18.4	4.8	4.386	10.74	7.82	2.01	8.37	1.65	6.83	8.58	37.75	129.5	172.9
2024	41.9	5.8	47.9	16.5	1.64	18.4	4.8	4.386	10.74	7.82	2.01	8.37	1.65	6.83	8.58	37.75	137.91	183.5
2025	41.9	5.8	47.9	16.5	1.64	18.4	4.8	4.386	10.74	7.82	2.01	8.37	1.65	6.83	8.58	37.75	146.26	194
2026	41.9	5.8	47.9	16.5	1.64	18.4	4.8	4.386	10.74	7.82	2.01	8.37	1.65	6.83	8.58	37.75	154.61	204.5
2027	41.9	5.8	47.9	16.5	1.64	18.4	4.8	4.386	10.74	7.82	2.01	8.37	1.65	6.83	8.58	37.75	162.9	215
2028	41.9	5.8	47.9	16.5	1.64	18.4	4.8	4.386	10.74	7.82	2.01	8.37	1.65	6.83	8.58	37.75	171.3	225.5
2029	41.9	5.8	47.9	16.5	1.64	18.4	4.8	4.386	10.74	7.82	2.01	8.37	1.65	6.83	8.58	37.75	179.6	236
2030	41.9	5.8	47.9	16.5	1.64	18.4	4.8	4.386	10.74	7.82	2.01	8.37	1.65	6.83	8.58	37.75	188	246.5

Conclusions: This study applied advanced time series modeling and machine learning techniques, specifically ARIMA, state space, and XGBoost models, to analyze historical trends and forecast future fish production at both state and national levels in India. Annual fish production data from 1993 to 2019 were used for model fitting, and the forecasting accuracy of each model was rigorously evaluated using a validation sample from 2020 to 2022. While XGBoost initially outperformed ARIMA and state space models on the training data, based on evaluation metrics such as RMSE, MAE, and MAPE, the validation sample revealed a different picture. XGBoost and state space models generally exhibited superior forecasting accuracy compared to ARIMA models for the majority of the time series. A key observation was that no single modeling approach proved universally best across all 18 state-level and national datasets. This highlights the importance for practitioners to consider and evaluate multiple techniques when forecasting diverse agricultural production time series. The forecasts generated from the best-performing models suggest that major fish-producing states like Andhra Pradesh and West Bengal are likely to maintain relatively stable fish production levels through 2030. In contrast, marine and total fish production at the national level are projected to experience moderate growth over the next eight years. The analysis underscored that while XGBoost can effectively capture complex patterns in training data, its ability to generalize to new data can vary, and, state-space models and even ARIMA models in some instances demonstrated greater accuracy on the validation data. This comprehensive analysis provided valuable, data-driven insights into India's future fish supply, information crucial for supporting effective food security planning and informing policy-making decisions by relevant stakeholders and authorities in the region. The established forecasting framework can, through regular updates with new data, be used for the long-term monitoring of fish production patterns.

Institutional Review Board Statement: Not applicable.

Funding: No Funding

Declaration of Interest Statement: The authors declare no conflict of interest.

Authors' Contribution

Shikha Yadav: Conceptualization, Formal analysis, Writing- Review & Editing. Binita Kumari: Conceptualization, Writing-original draft. Divya Sharma: Data Curation, Visualization, Resources. Abdullah Mohammad Ghazi Al khatib: Software, Validation. Bayan Mohamad Alshaib: Software, Validation, Yashpal Singh Raghav: Investigation, Supervision. Hiranmayee Nayak: Investigation, Supervision. Tufleuddin Biswas: Writing- Review and Editing, Visualization. Soumik Ray: Methodology, Writing-original draft, Supervision. Neha Mishra: Writing- Review and Editing, Data Curation. Pradeep Mishra: Investigation, Supervision, Writing-Review and Editing.

REFERENCES

Aanes, S., S. Engen, B.E. Sæther, T. Willebrand and V. Marcström (2002). Sustainable harvesting strategies of willow ptarmigan in a fluctuat-ing environment. Ecol. Appl. 12: 281–290 https://doi.org/10.2307/3061152

Abotaleb, M., S. Ray, P. Mishra, K. Karakaya, C. Shoko, A.M.G. Al Khatib and R. Balloo (2021). Modelling and forecasting of rice production in south Asian countries. Ama, Agril. Mechaniz in Asia, Africa and L. America. 51: 1611-1627.

AL Khatib, A.M.G., H. Yonar, M. Abotaleb, P. Mishra, A. Yonar, K. Karakaya, A. Badr and V. Dhaka (2021). Modeling and forecasting of egg production in India using time series models. Eurasian J. Vet. Sci. 37(4): 265-273

Anuja, A., V.K. Yadav, V.S. Bharti and N.R. Kumar (2017). Trends in marine fish production in Tamil Nadu using regression and autoregressive integrated moving average (ARIMA) model. JANS. 9(2):653–657 https://doi.org/10.31018/jans.v9i2.1252

Boruah, B.B., P. Roy, A. Dutta and B.B. Hazarika (2020). Forecasting marine and total fish production in India using ARIMA model. Indian J. Econ. Bus. 19 (2): 161–174.

Chen, T. and C. Guestrin (2016). XGBoost: a scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16). Association for Computing Machinery, New York, NY, USA. 785–794. https://doi.org/10.1145/2939672.2939785

Delgado, C.L., N. Wada, M.W. Rosegrant, S. Meijer and M. Ahmed (2002). Sustainable solutions for ending hunger and poverty. 2020 Vision Food Policy Report, IFPRI and World Fish Centre.

Krishnan, M. and P.S. Birthal(2002). Aquaculture development in India: an economic overview with special reference to coastal aquaculture. Aquac. Econ. Manag. 6 (1):81–96.https://doi.org/10.1080/13657300209380305 .

Mishra, P., A.M.G. Al Khatib, P. Lal, A. Anwar, K. Nganvongpanit, M. Abotaleb and V. Punyapornwithaya (2023a). An overview of pulses production in India: retrospect and prospects of the future food with an application of hybrid models. Natl. Acad. Sci. Lett. 46: 367–

374. https://doi.org/10.1007/s40009-023-01267-2
Mishra, P., A.M.G. Al Khatib, I. Sardar, J. Mohammed, K. Karakaya, A. Dash and A. Dubey (2021). Modeling and forecasting of sugarcane production in India. Sugar Tech. 23(6): 1317-1324.
Mishra, P., A.M.G. Al Khatib, I. Sardar, J. Mohammed, M. Ray, K. Manish and K. Rono (2020). Modelling and forecasting of COVID-19 in India. J. of Infect, Dis. Epid. 6(5): 1-11.
Mishra, P., K.M. Alakkari, A. Lama, S. Ray, M. Singh, C. Shoko and K. Karakaya (2023b). Modeling and forecasting of sugarcane production in South Asian countries. Cur. App. Sci. Tech., 23(1): 1-15. DOI: 10.55003/cast.2022.01.23.002.
Mishra, P., S. Ray, A.M.G. Al Khatib, M. Abotaleb, S. Tiwari, A. Badr and R. Balloo (2021). Estimation of fish production in India using ARIMA, Holt's linear, BATS and TBATS models. Ind.J. of Eco. 48: 1154-1161
Mohammed, J., A.M.G. Al Khatib, P. Mishra, P. Adjei, P.K. Singh, S.K. Priya, and S.S. Das (2021). Modeling and forecasting of Covid-19 from the context of Ghana. Afri. Rev.of Econ. Fin. 13(2): 124-141.
Niranjan, H.K., B. Kumari, Y.S. Raghav, P. Mishra, A.M.G. Al Khatib and M. Abotaleb (2022). Modeling and forecasting of tea production in India. J. Anim. Plant. Sci. 32(6): 1598 doi. 10.36899/JAPS.2022.6.0569
Raghav, Y.S., P. Mishra, K.M. Alakkari, M. Singh, A.M.G. Al Khatib and R. Balloo (2022). Modelling and forecasting of pulses production in south asian countries and its role in nutritional security. Legume Res.-An Interntl. J. 45(4): 454-461.
Rahman, U.H., S. Ray, A.M.G. Al Khatib, P. Lal, P. Mishra, C. Fatih and K. Alakkari (2022). State of Art of SARIMA model in second wave on COVID-19 in India. Interntl. J. of Agril. Stat. Sci. 18(1): 141-152.
Raman, R.K., T.V. Sathianandan, A.P. Sharma and B.P. Mohanty (2017). Modelling and forecasting marine fish production in Odisha using seasonal ARIMA model. Natl. Acad. Sci. Lett. 40(6):393–397. https://doi.org/10.1007/s40009-017-0581-2
Ray, S., A.M.G. Al Khatib, B. Kumari, T. Biswas, A.C. Nuta and P. Mishra (2023a). Forecasting cash crop production with statistical and neural network model. J. Crop and Weed. 19(1): 194-201. https://doi.org/10.22271/09746315.2023.v19.i1.1679
Ray, S., P. Mishra, H. Ayad, P. Kumari, R. Sharma, B. Kumari, A.M.G. Al Khatib, A. Tamang, and T. Biswas (2023b). Prediction of fruit production in India: an econometric approach. J. Hort. Res. 31(1): 25-34. https://doi.org/10.2478/johr-2023-0005
Sharma, M., N. Jasrotia, B. Kumar, A. Bhat and S. Mahajan (2018). Modeling of monthly arrival of rohu fish using ARIMA in Jammu Region of J. and K. State. J. Anil. Res. 8(2): 259-262.
Stephen, S.K., V.K. Yadav and R.R. Kumar (2022). Comparative study of statistical and machine learning techniques for fish production forecasting in Andhra Pradesh under climatic change scenario. Ind. J. Geo Mar. Sci. 51(09): 776-784.
Tiwari, R.K., P. Mishra, B. Kumari, A.M.G. Al Khatib, S. Yadav, M. Abotaleb, S. Ray and M. Kumari (2022). Modelling and forecasting of electricity consumption used in agriculture purpose in India. J. Crop and Weed **18** (3): 200-205. https://doi.org/10.22271/09746315.2022.v18.i3.1636
Yadav, A.K, K.K. Das, P. Das, R.K. Raman, J. Kumar and B.K. Das (2020). Growth trends and forecasting of fish production in Assam, India using ARIMA model. J. Appl. Natl. Sci. 12(3):415–421https://doi.org/10.31018/jans.v12i3.2353
Yadav, S., P. Mishra, B. Kumari, I.A. Shah, K. Karakaya, S. Shrivastri and A.M.G. Al Khatib (2022). Modelling and Forecasting of maize production in South Asian countries. Eco. Aff. 67(4): 519-531.
Yadav, V.K., M. Krishnan, R.S. Biradar, N.R. Kumar and V.S. Bharti (2013). A comparative study of neural-network and fuzzy time series forecasting techniques. Case study: Marine fish production forecasting. Ind. J Geo-Mar. Sci. 42(6): 707–716.
Yonar, H., A. Yonar, P. Mishra, M. Abotaleb, A.M.G. Al Khatib, T. Makarovskikh and M. Cam (2022). Modeling and forecasting of milk production in different breeds in Turkey. The Ind. J. Anil. Sci. 92(1): 105-111.