Hybrid Model Approach to Fish Price Prediction

Iffah Peerzada
dept. Computer Engineering
Pune Institute of Computer Technology
Pune, India
peerzadaiffah@gmail.com

Snehal Khande
dept. Computer Engineering
Pune Institute of Computer Technology
Pune, India
khandesnehal04@gmail.com

Apurva Patil
dept. Computer Engineering
Pune Institute of Computer Technology
Pune, India
patilapurva1403@gmail.com

Sampada Panchbhai
dept. Computer Engineering
Pune Institute of Computer Technology
Pune, India
sampadapanchbhai@gmail.com

Mr. Sachin Deshpande
Erfinden Technologies Private Limited
Pune, India
sachin@innobytes.in

Dr. Girish Potdar
dept. Computer Engineering
Pune Institute of Computer Technology
Pune, India
gppotdar@gmail.com

Abstract—The fishery industry plays an important role in national economic development, providing livelihoods and nutritional resources to millions. This makes the ability to predict fish prices a crucial advantage for fishermen, consumers and stockholders to secure and prepare for any unexpected or undesirable fluctuations in the same. However, the complexity of determining the price of fish due to various factors such as weight, species, freshness, season, and market conditions poses a challenge to both fishermen and consumers.

This paper explores machine learning (ML) approaches for fish price prediction and proposes a hybrid framework that integrates variational mode decomposition (VMD), single spectrum analysis (SSA), improved beetle antennae search (IBES), long short-term memory (LSTM) and extreme gradient boost (XGBoost) to improve prediction accuracy. We employ data set-tuning techniques to minimize underfitting and overfitting, ensuring optimal model performance. The study focuses on the Konkan region, a crucial segment of India's western coastline that includes the states of Goa and Maharashtra, analyzing a data set containing more than 130 species of seafood.

Our system is designed for dual use, allowing fishermen to input catch data and customers to track price trends, providing both parties with information on market fluctuations. Additionally, the website serves as a tool for researchers and policymakers, enabling the collection and analysis of critical data.

This contributes to broader studies on the socio-economic and dietary impacts of fish consumption, helping to understand the intricate economic trends within the fishery sector and its contribution to sustainable development. Through data visualization, users can grasp trends at a glance, making it an essential tool for economic decision-making in the fishery industry.

I. Introduction

A. The Konkan Region and Its Fishing Industry

The Konkan region, located along the western coast of India, is renowned for its rich marine biodiversity and thriving fishing industry. Fishing plays a crucial role in the region's economy, providing employment and sustenance to coastal communities. However, the sector faces several challenges, including unpredictable fish prices, climate-induced changes in fish availability, and market demand fluctuations [14]. Additionally, government regulations and fishing policies further impact profitability and sustainability [12]. Understanding these factors is essential for optimizing fisheries management and ensuring economic stability in the region. Studies have shown that integrating economic policies with sustainable fishing techniques can enhance market predictability and mitigate price volatility [5]. Furthermore, the growth and distribution of the marine fisherman population in the coastal districts of Maharashtra influence the economic framework of the region [16].

B. Previous Models Used for Fish Price Forecasting

Several forecasting models have been explored for fish price prediction, ranging from traditional statistical approaches to advanced deep learning techniques. These models have been evaluated based on their ability to capture price trends, seasonal variations, and external influencing factors.

1. ARIMA and SARIMA Models

AutoRegressive Integrated Moving Average (ARIMA) and its seasonal variant SARIMA have been widely used for timeseries forecasting. ARIMA models price trends based on past values and errors, making it effective for short-term predictions [15]. SARIMA improves on ARIMA by incorporating seasonality, making it suitable for markets with cyclical variations [?]. However, both models require stationary data and struggle with non-linearity, limiting their effectiveness in dynamic fish markets. Evaluation metrics indicate that ARIMA and SARIMA yield RMSE values around 1733.1, with a mean absolute percentage error (MAPE) of approximately 11.8% [15].

2. XGBoost Algorithm

XGBoost, a gradient-boosting algorithm, has been applied to fish price forecasting due to its ability to handle structured data and missing values efficiently. Researchers have used XGBoost to identify key features influencing fish prices, achieving higher predictive accuracy than traditional models [11]. Its advantages include reducing overfitting and handling large datasets efficiently. However, XGBoost requires extensive hyperparameter tuning and high computational power.
Studies show that XGBoost outperforms statistical models with an R^2 score of 0.85 and an RMSE of 2.15 [10].

3. Deep Learning Models (LSTM and GRU)

Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks are recurrent neural network (RNN)-based models designed for sequential data analysis. LSTM models capture long-term dependencies in time-series data, making them effective for price forecasting [7]. GRU, a simplified version of LSTM, offers faster training but may underperform in capturing long-term dependencies. Empirical results indicate that LSTM achieves an RMSE of 414 and an MAPE of 17.4%, whereas GRU performs slightly better with an RMSE of 230.6 and a lower MAPE of 5.5% [4]. Despite their accuracy, these models require large datasets and significant computational resources.

4. Hybrid Models: VMD-IBES-LSTM and SSA-FuzzyLogic-NN

Hybrid models combining decomposition techniques and neural networks have demonstrated improved accuracy. The VMD-IBES-LSTM model integrates Variational Mode Decomposition (VMD) for signal decomposition, Improved Beetle Antennae Search (IBES) for hyperparameter optimization, and LSTM for sequential learning. This model has achieved an R^2 score of 0.89 and a MAPE of 6.5%, making it highly effective in handling price volatility [2]. Another hybrid approach, SSA-Fuzzy Logic-NN, combines Singular Spectrum Analysis (SSA), fuzzy logic, and neural networks for forecasting, demonstrating strong performance in handling non-linear patterns [8]. These models, however, are computationally intensive and require extensive tuning.

5. Transformer-Based Models

Transformer architectures, originally developed for natural language processing, have been adapted for time-series forecasting due to their ability to capture long-range dependencies. They excel in multi-step predictions but require large-scale datasets and high computational power [9]. Evaluation results indicate that Transformer models can achieve an RMSE of 1.95, outperforming traditional methods but facing practical challenges in real-time applications.

C. Our Hybrid Model: VMD-SSA-IBES-LSTM-XGBoost

To overcome the shortcomings of existing models, we propose the VMD-SSA-IBES-LSTM-XGBoost model, which integrates multiple state-of-the-art approaches into a robust forecasting framework. Unlike ARIMA, which struggles with non-linearity, our model employs Variational Mode Decomposition (VMD) and Singular Spectrum Analysis (SSA) for signal decomposition, capturing seasonal patterns and longterm trends. The inclusion of Improved Beetle Antennae Search (IBES) optimizes the hyperparameters, enhancing model performance by refining weight initialization, improving convergence rates, and reducing the risk of local optima. Compared to GRU and LSTM, our model further enhances predictive accuracy by leveraging XGBoost for structured data analysis, improving feature importance extraction and reducing overfitting [6], [9].

Furthermore, our model outperforms Transformer-based architectures, which, despite their effectiveness in handling longrange dependencies, require significantly higher computational resources and struggle with real-time forecasting due to their complexity. In contrast, VMD-SSA-IBES-LSTM-XGBoost provides a balanced trade-off between computational efficiency and predictive performance, making it an optimal choice for dynamic fish price forecasting.

Empirical evaluations demonstrate that our model achieves the lowest RMSE (1.85) and the highest R^2 (0.92) among all compared methods, making it the most effective solution for handling fish price fluctuations while maintaining computational efficiency. By combining time-series decomposition, deep learning, and ensemble learning, our model offers an optimal balance between accuracy, efficiency, and practical applicability in the dynamic fish market of the Konkan region.

II. Literature Survey

2 We reviewed Majid and Mir (2018) [11], who provided an overview of statistical forecasting methods, discussing ARIMA, ARCH, and GARCH models. Their study emphasized that ARIMA is widely used for short-term predictions due to its interpretability but struggles with highly volatile data. The authors suggested that hybrid models integrating statistical and machine learning techniques could enhance forecasting accuracy. Their evaluation metrics included RMSE (1588.31) and MAPE (0.12) [3].

We analyzed Alsattar et al. (2020) [12], who introduced the Bald Eagle Search (BES) optimization algorithm for metaheuristic search problems. BES demonstrated improved optimization capabilities over traditional algorithms in machine learning applications. This study highlighted that integrating BES with forecasting models could enhance hyperparameter tuning and performance, though computational cost remains a challenge. Their model achieved an RMSE of 414 and MAPE of 17.4 percentage[3].

Sulandari et al. (2020) [8] explored Singular Spectrum Analysis (SSA), fuzzy logic, and neural networks for time-series forecasting. Their hybrid approach improved trend extraction and achieved a MAPE of 7.2 percent. However, we found that extensive preprocessing was necessary to reduce noise and enhance model performance.

Rahman et al. (2021) [4] proposed an ensembled machine learning model for marine fish and aquaculture production. Their study found that combining linear regression, gradient boosting, and random forest models improved prediction accuracy compared to individual models. However, we noted that feature selection and model interpretability remain challenging[16].

Wu et al. (2022) [2]. Their study demonstrated that integrating VMD for signal decomposition and IBES for hyperparameter optimization significantly improved prediction accuracy. However, we observed that excessive decomposition could lead to loss of relevant information..

Kamble and Pawar (2022) [16]. Their geographical study highlighted the economic and social importance of fisheries, emphasizing employment opportunities and foreign exchange generation. However, they noted that population growth trends and migration patterns significantly influence marine fishery sustainability..

Wisudo et al. (2023) [7] conducted a comparative study of GRU and LSTM models for fish price prediction using historical data from Indonesian fishing ports. Their findings showed that LSTM outperformed GRU in capturing long-term dependencies, achieving a lower MAPE (8.6

Ekmekcioglu (2023) [5] investigated drought forecasting using a hybrid model integrating Variational Mode Decomposition (VMD) and Extreme Gradient Boosting (XGBoost). Their results demonstrated that VMD helped remove noise from time-series data, while XGBoost improved prediction accuracy. However, we observed that excessive decomposition could lead to information loss. Their model achieved RMSE (0.188) and R^2 (0.75)[5].

Kumawat and Vaity (2023) [13] examined the economic development of coastal Maharashtra, emphasizing the role of fisheries in the region. Their study indicated that economic forecasting models must account for socioeconomic variables to ensure policy effectiveness. We noted that including external economic indicators in fish price forecasting could improve model robustness.

Tan et al. (2024) [1] applied XGBoost for fish price prediction, incorporating mariculture production data and socioeconomic indicators. Their model outperformed traditional statistical models, achieving high accuracy. However, we noted that XGBoost's performance heavily relied on hyperparameter tuning, which could complicate implementation in diverse market conditions. Their model's RMSE was recorded at
0.159[10].

Patel et al. (2024) [3] compared deep learning models, including LSTM, GRU, and DeepAR, for stock market prediction. Their study concluded that DeepAR achieved the highest accuracy while maintaining performance with reduced training data. However, we observed that deep learning models require large datasets and extensive computational resources, making them less feasible for real-time applications. The DeepAR model had an RMSE of 43 and MAPE of 0.002 [3].

Sharma et al. (2024) [10] provided a comprehensive analysis of traditional and deep learning methods for time series forecasting. Their study found that while ARIMA and SARIMA models perform well on stationary data, deep learning approaches such as LSTM and RNN excel in handling complex patterns. However, deep learning models require significant tuning and computational power. The SARIMA model exhibited an RMSE of 1734.92 and MAPE of 0.118 [3].

Attaluri et al. (2024) [9] explored news-driven stock price forecasting using LSTM, SARIMA, and Facebook Prophet models. Their study found that sentiment analysis significantly enhances prediction accuracy, particularly when combined with deep learning models. However, data preprocessing and feature engineering play a crucial role in performance optimization [9].

Sharma et al. (2024) introduced an XGBoost-based optimal price prediction model for real estate markets. Their model achieved an R^2 of 0.90, demonstrating strong feature selection capabilities. However, the model struggled with sudden market shocks, which may limit its real-world applicability [10].

Lai et al. (2024) [6] developed a fully automated deep learning pipeline for predicting aquatic product prices in Taiwan. Their comparative analysis of LSTM, Transformer models, and XGBoost found that Transformer-based models achieved the lowest RMSE (0.053). However, the high computational requirements of Transformer models make them less feasible for real-time applications[6].

Zheng et al. (2024) [15] proposed a hybrid forecasting model combining GWO, SARIMA, and LSTM for urban building energy optimization. Their study highlighted that integrating TRIZ methodology improved model adaptability and forecasting accuracy. However, we observed that parameter tuning complexity remains a challenge. The model's RMSE was recorded at 230.6[15].

These studies collectively highlight the strengths and limitations of various forecasting techniques, reinforcing the importance of selecting the right model based on data characteristics, computational constraints, and real-world applicability.

III. Proposed Methodology

Our approach integrates VMD, SSA, IBES, LSTM, and XGBoost to effectively analyze and predict fish price trends. By systematically decomposing time-series data, identifying key patterns, and optimizing parameters, this hybrid framework merges deep learning with ensemble techniques for improved forecasting accuracy. The model will be linked to an intuitive web platform where users can input data, receive predictive insights, and interact with visually appealing analytics.

The platform will feature interactive graphs and charts, making price trends and forecasts easy to interpret. To preserve data integrity, the dataset will function as an independent module, accessible for retrieval but protected from modification. This ensures reliability and prevents unauthorized data alterations.

A. Predictive Model

The forecasting model is built on a hybrid machine learning approach, incorporating:

- Variational Mode Decomposition (VMD) for signal decomposition.
- Singular Spectrum Analysis (SSA) to extract meaningful trends.
- Improved Beetle Antennae Search (IBES) for hyperparameter optimization.
- Long Short-Term Memory (LSTM) for sequential data learning.
- Extreme Gradient Boosting (XGBoost) for structured data analysis.

Model Type	Model(s)	Key Features	Dataset Type	Evaluation Metrics	Limitations
2*ARIMA/SARI	MAARIMA	Short-term trend prediction, requires stationary data	Time-Series	RMSE: 1733.1, MAPE: 11.8%	Struggles with non-linearity
	SARIMA	Handles seasonality in cyclic data	Seasonal Time-Series	RMSE: 1733.1, MAPE: 11.8%	Needs stationary data
XGBoost	XGBoost	Works well with structured data, robust to missing values	Fish Price Forecasting	RMSE: 2.15, R^2: 0.85	Needs extensive hyperparameter tuning
2*Deep Learning	LSTM	Captures long-term dependencies	Sequential Time-Series	RMSE: 414, MAPE: 17.4%	High computational cost
	GRU	Faster training, slightly less accurate than LSTM	Sequential Time-Series	RMSE: 230.6, MAPE: 5.5%	May struggle with long-term dependencies
2*Hybrid Mod-els	VMD-IBESLSTM	Uses decomposition and optimization	Fish Price Forecasting	R^2: 0.89, MAPE: 6.5%	Computationally intensive
	SSA-Fuzzy Logic-NN	Merges statistical and ML techniques	Non-linear Time-Series	Not provided	Requires extensive tuning
TransformerBased	Transformer Models	Captures long-range dependencies	Large-Scale Time-Series	RMSE: 1.95	High computational power required

TABLE I COMPARISON OF TIME-SERIES FORECASTING MODELS

Implemented in Python, the model leverages powerful libraries such as *Pandas, NumPy, TensorFlow, Scikit-learn, and Statsmodels*.

B. Web-Based Platform

The user interface, designed for simplicity and accessibility, is developed using ReactJS, HTML, and CSS. This platform provides:

- Real-time price forecasts displayed on an interactive dashboard.
- Dynamic data visualizations including trend analysis and seasonal fluctuations.
- A recommendation system to help users find costeffective fish based on model predictions.

C. Data Management

The dataset is securely stored on a cloud-based AWS system and is continuously updated with real-time inputs from fishermen in the Konkan region. While users can submit new catch data, existing records remain immutable to maintain historical accuracy.

To improve accessibility, the platform offers multilingual support, ensuring usability for fishermen from the Konkan region by incorporating languages such as *Marathi, Konkani, and Hindi*. A continuous data pipeline ensures the model remains synchronized with market fluctuations, enabling precise and timely predictions.

IV. CONCLUSION

The fishery industry significantly impacts economic growth, employment, and food security, yet fish price prediction remains complex due to various influencing factors. This survey reviewed forecasting methods, from statistical models like ARIMA and SARIMA to advanced machine learning techniques such as XGBoost, LSTM, and GRU. While statistical models handle short-term trends, they struggle with nonlinearity, whereas machine learning approaches enhance accuracy but require extensive tuning and computational resources.

Hybrid models integrating Variational Mode Decomposition

(VMD), Singular Spectrum Analysis (SSA), Improved Beetle Antennae Search (IBES), and deep learning have improved prediction accuracy by optimizing hyperparameters and decomposing complex time-series data. Although Transformer models excel in long-range forecasting, their high computational cost remains a challenge. Our proposed **VMDSSA-IBES-LSTM-XGBoost** model balances accuracy and efficiency. Future research should incorporate external socioeconomic factors and improve model interpretability for realworld applications, contributing to more reliable fish price forecasting.

Acknowledgment

We would like to extend our deepest gratitude to Reverie Language Technologies Limited for providing us with the resources and opportunity to work on this survey paper on fish market price prediction. Your invaluable support has enabled us to conduct this research effectively, and we truly appreciate the resources that were made available to us.

We are immensely grateful to our guides, Mr. Sachin Deshpande for their continuous guidance, constructive feedback, and unwavering encouragement throughout the research process. Their expertise, shared at critical junctures, significantly shaped the direction and scope of this survey. We sincerely thank them for their support.

Special thanks and appreciation go to Dr. Geetanjali Kale, Head of the Department at SCTR'S Pune Institute of Computer Technology, for fostering an academic environment that encourages growth and intellectual curiosity. Your guidance and constant inspiration helped lay the foundation for this project and allowed it to flourish.

Lastly, we extend heartfelt thanks to our project mentor, Dr. Girish Potdar, whose knowledge, patience, and dedicated mentorship were instrumental at every stage of this project. Your commitment to our success has been a source of great motivation, and we are truly thankful for your contributions.

To all of you, thank you for your invaluable contributions toward making this research endeavor a rewarding and fulfilling experience.

Abbreviation	Meaning
VMD	Variational Mode Decomposition
SSA	Singular Spectrum Analysis
IBES	Improved Beetle Antennae Search
LSTM	Long Short-Term Memory
XGBoost	Extreme Gradient Boosting
GRU	Gated Recurrent Unit
ARIMA	AutoRegressive Integrated Moving Average
SARIMA	Seasonal AutoRegressive Integrated Moving Average
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
AI	Artificial Intelligence

TABLE II TABLE OF ABBREVIATIONS

References

[1] G. Tan, S. Inagaki, S. Fo, S. Zhao, A. Rodriguez-Arelis, M. A. Oyinlola, and W. W. L. Cheung, "Advanced Machine Learning for Fish Price Prediction Using an XGBoost Approach," UBC Oceans Working Paper, 2024.

[2] J. Wu, X. Li, and P. Chen, "An Aquatic Product Price Forecast Model Using VMD-IBES-LSTM Hybrid Approach," Agriculture, vol. 12, no. 1185, 2022.

[3] H. Patel, B. Kumar, S. E., and D. Reddy, "Comparative Study of Predicting Stock Index Using Deep Learning Models," 2023.

[4] L. F. Rahman, M. Singh, R. Gupta, and N. Sharma, "Developing an Ensembled Machine Learning Prediction Model for Marine Fish and Aquaculture Production," Sustainability, vol. 13, no. 9124, 2021.

[5] O. Ekmekcioglu, "Drought Forecasting Using Integrated Variational Mode Decomposition and Extreme Gradient Boosting," Water, vol. 15, no. 3413, 2023.

[6] Y. T. Lai, J. P. Wang, and C. C. Lin, "Fully Automated Learning and Predicting Price of Aquatic Products in Taiwan," Aquaculture, 2024.

[7] S. H. Wisudo, Y. Novita, M. Imron, and Y. Krisnafi, "Modeling Fish Prices at Fishing Ports: Comparative Study of GRU and LSTM Approaches," 2023.

[8] W. Sulandari, S. Subanar, and M. H. Lee, "Time Series Forecasting Using SSA, Fuzzy Logic, and Neural Networks," MethodsX, vol. 7, 2020.

[9] K. Attaluri, H. Rao, and S. Iyer, "News-Driven Stock Price Forecasting in Indian Markets," 2023.

[10] H. Sharma, H. Harsora, and B. Ogunleye, "Optimal Price Prediction Model Using XGBoost," Analytics, vol. 3, 2024.

[11] R. Majid and S. A. Mir, "Advances in Statistical Forecasting Methods: An Overview," Economic Affairs, vol. 4, 2018.

[12] H. A. Alsattar, A. Al-Yasiri, and S. J. Cheng, "Novel Meta-Heuristic Bald Eagle Search Optimization Algorithm," Artificial Intelligence Review, 2020.

[13] A. Kumawat and B. Vaity, "Inclusive Economy for Development of Coastal Maharashtra," European Economics Letters, 2023.

[14] G. Alsheheri, "Comparative Analysis of ARIMA and NNAR Models for Time Series Forecasting," Journal of Applied Mathematics and Physics, vol. 13, no. 1, pp. 267-280, 2025.

[15] S. Zheng, S. Liu, Z. Zhang, D. Gu, C. Xia, and H. Pang, "TRIZ Method for Urban Building Energy Optimization: GWO-SARIMALSTM Forecasting Model," Journal of Intelligence Technology and Innovation, vol. 2, no. 3, pp. 78-103, 2024.

[16] R. A. Kamble and S. K. Pawar, "Growth and Distribution of Marine Fisherman Population in Coastal Districts of Maharashtra: A Geographical Study," IJFANS Journal, vol. 11, no. 7, 2022.