Seafood Market Price Prediction

Problem Statement

The seafood market is characterized by significant price volatility driven by complex factors including seasonal demand, varying catch volumes, and broader economic conditions. For businesses, distributors, and consumers, this unpredictable fluctuation creates substantial challenges:

- **Financial Risk:** Distributors face inventory management risks due to the difficulty of optimizing purchasing strategies without reliable future price signals.
- **Operational Inefficiency:** Businesses struggle with inaccurate budgeting

Objective

The primary challenge is the lack of a robust and reliable forecasting mechanism that can accurately model both the historical linear (trend/seasonality) and non-linear (market shocks/complex interactions) components of seafood price movements.

This project addresses this by developing and comparing a suite of advanced machine learning and statistical models (ARIMA, Random Forest, and XGBoost) to establish the most accurate and actionable prediction of future seafood prices, thereby mitigating financial uncertainty and enabling data-driven optimization of supply chain and pricing strategies.

Approach

1. Data Loading & Conversion

•Imports the seafood dataset from CSV.

•Converts prices from USD to INR using a fixed conversion rate.

•Parses dates and organizes the data for time series analysis.

2. Exploratory Data Analysis (EDA)

•Visualizes seafood price fluctuations over time.

•Identifies trends, seasonality, and outliers.

•Calculates rolling averages to understand long-term changes.

3. Data Preprocessing

•Handles missing values and inconsistent data points.

•Sets the `Date` column as the time index.
•Prepares the data for model training and forecasting.

4. Model Building

Multiple forecasting models are applied and compared to identify the best performer:

•**ARIMA Model (AutoRegressive Integrated Moving Average):** — This model is a classic time series forecasting method that captures linear trends and seasonality in the price data. It's a foundational model that provides a baseline for evaluating more complex algorithms.

•**Random Forest (RF)** — As an ensemble learning method, Random Forest builds multiple decision trees to improve prediction accuracy. It is effective at handling non-linear relationships and interactions between features, which can be useful for capturing complex patterns in price fluctuations..

•**XGBoost (eXtreme Gradient Boosting):** — This is another powerful and popular gradient-boosted ensemble model. It is known for its speed and performance in machine learning competitions. XGBoost can effectively handle large datasets and complex relationships, making it a strong candidate for price forecasting.

5. Model Evaluation

•Evaluates models using **MAE**, **RMSE**, and **R^2 Score**.

•Compares prediction accuracy across models.

•Selects the model with the lowest forecast error as the best predictor.

6.Forecasting Future Prices

•Forecasts seafood prices for upcoming months/years.

•Visualizes forecasted trends alongside historical data.

•Provides insights into expected future demand and market prices.

Results & Insights

•The dataset showed a gradual increase in seafood prices, influenced by seasonal demand patterns.

•XGBoost and Random Forest (RF) models, which are strong at capturing complex non-linear relationships, typically provided the most accurate forecasts. They were superior in handling the sudden spikes and dips not fully explained by linear time factors alone.

•Forecast results show a steady upward trend, indicating potential price increases in the near future.

•The ARIMA model served as a reliable baseline, successfully capturing the underlying linear trends and macro-level seasonality.

Challenges Faced

Data Scarcity and Access Restrictions

The primary and most severe challenge faced by this project was the **profound difficulty in securing adequate, reliable data**, transforming the standard modeling task into a foundational data acquisition struggle.

Specifically, the limitations were:

- Extreme Data Scarcity and Fragmentation
- Governmental Access Restrictions
- Absence of Critical Exogenous Variables
- Verification and Trust Issues

Model Selection and Interpretation

- Overfitting Complex Models

Deployment and Actionability

- Timely Forecasting-To run the model and generate the output graph lot of mathematical calculations are required so the time used increases.