Development Of AI-MI Based Models for Predicting Prices of Agri-Horticultural Commodities

Puneeth N
Department of Computer Science and Engineering,
Presidency University, Bangalore, Karnataka, India
puneethpunee104@gmail.com

Thanu Shree M
Department of Computer Science and Engineering,
Presidency University, Bangalore, Karnataka, India
thanushreemanjunath03@gmail.com

Dr. Shanthi S
Associate Professor
Department of Computer Science and Engineering,
Presidency University, Bangalore, Karnataka, India
shanthi.s@presidencyuniversity.in

Rohan Gowda A
Department of Computer Science and Engineering,
Presidency University, Bangalore, Karnataka, India
a.rohan1998@gmail.com

Ananya B R
Department of Computer Science and Engineering,
Presidency University, Bangalore, Karnataka, India
ananyabrbr@gmail.com

Ajin V Joseph
Department of Computer Science and Engineering,
Presidency University, Bangalore, Karnataka, India
ajinvarghese188@gmail.com

Abstract : The Vegetable Price Prediction project aims to develop a reliable predictive model to predict vegetable prices based on crucial seasonal, weather, and crop health data. As vegetable price is a top concern for farmers, suppliers, and consumers with volatile vegetable prices, accurate price predictions can facilitate informed decision-making for stakeholders. This project makes use of historical data like vegetable kind, season of growth, month, temperature, recent calamity occurrences, state of vegetable, and other variables that affect the price per kilogram to predict the price per kilogram. Utilizing advanced machine learning algorithms ARIMA, LSTM, XGBoost, Linear Regression and Hybrid Algorithms we aim to identify complex patterns and time dependencies in the data to make the prediction. Each algorithm has a strength: ARIMA for time series, LSTM for long-range dependencies, XGBoost for performance, and Linear Regression for simplicity and interpretability. The results of the model will be verified for effectiveness and accuracy, making strong predictions in different scenarios.

***Keywords*:** Machine Learning Models, ARIMA, LSTM (Long Short-Term Memory), XGBoost, Linear Regression, Hybrid Algorithms.

1 INTRODUCTION

Agriculture and horticulture form the cornerstones of economies in general, supplying important food and raw material. Prices instability of agri-horticultural commodities are perennial problems striking the farmers, the traders, as well as the policymakers alike. Prices of agri-horticultural commodities fluctuate on the basis of various factors like supply and demand, weather, government actions, market conditions, and other external trade factors. Proper prediction of these prices is necessary to guarantee economic stability, promote farmer income, and plan the market accordingly.

With the development of artificial intelligence (AI) and machine learning (ML), data-driven solutions have become key to solving complex economic problems. AI-ML models offer advanced ways of analyzing large sets of data, detecting patterns, and making accurate predictions, which can be useful inputs for stakeholders in the agricultural supply chain.

This project aims to develop predictive models of agri-horticultural commodity prices on the basis of AI-ML with data-driven techniques. Agricultural prices are influenced by numerous factors such as weather, supply-demand, government policy, and global trade, so an accurate forecast is required for market stability and farmers' profitability. Based on historical price data, satellite images, climatic conditions, and market trends, the project will leverage sophisticated machine learning models—involving time-series forecasting, deep learning, and regression models—to analyze trends and make precise projections. The project intends to provide actionable insights to farmers, traders, and policy-makers for optimal decision-making, minimizing risks, and enhancing farm planning. Lastly, this initiative seeks to increase price transparency, remove uncertainties, and create a more effective market system for the agricultural sector.

1.1 Feature Selection

Predicting agri-horticultural commodity prices using AI and machine learning includes careful selection of important features to ensure reliable predictions. Agri-horticultural commodities are driven by numerous factors including past price movements, supply-demand changes, climatic situations, economic policies, and consumers' moods. Past prices are required to identify long-term trends and seasonality, while market forces—production levels, trade volumes, and inflation levels—dictate price movements. Weather conditions, including rainfall, temperature changes, and dryness patterns, directly influence crop production, affecting supply chains and prices. In addition, crop-specific factors like seasons of harvest, storage requirements, and perishability also lead to price volatility. External global conditions, including export-import policies, currency movements, and geopolitical issues, also impact market stability. Additionally, social aspects, including consumer dietary behavior, festive seasons, and sentiment in the market, also play important roles in creating price volatility. Technical innovations in agriculture, including mechanization, irrigation techniques, and availability of fertilizers, also affect market prices and efficiency in production. To pre-process these variables for machine learning models, feature selection techniques like correlation analysis, tree-based model feature importance ranking, and PCA can be utilized to eliminate redundant information and enhance predictive accuracy. Incorporating expert knowledge and statistical testing further adds to model reliability to ensure that price forecasting using AI can deliver decision-supportable results for farmers, policymakers, and market players. With these attributes tuned, predictive models can support better decision-making, reduce uncertainties, and enable more efficient agricultural planning, leading to enhanced price transparency and stability for the agri-horticultural sector.

1.2 Contribution

The Vegetable Price Forecasting project aims to establish a solid forecast model that can precisely predict vegetable prices based on crucial seasonal, environmental, and crop condition factors. Because vegetable prices are uncertain, this project empowers farmers, suppliers, and consumers by providing data-driven insights for better decision-making. Utilizing advanced machine learning methods such as ARIMA for forecasting from time series data, LSTM for finding long dependencies, XGBoost for generating high-speed forecasts, and Linear Regression for readability, the model extracts sophisticated trends from historical data to produce high-precision predictions. Important features like the growth season, month, temperature, occurrence of disasters, and vegetable condition are used to refine the model's predictions so that it considers actual-world factors influencing price movements. By incorporating feature selection techniques like correlation analysis, XGBoost feature importance ranking, and Principal Component Analysis (PCA), the model optimizes data input, thus becoming more precise and robust. The information gained from this project will enable farmers to plan production cycles, suppliers to optimize supply chains, and consumers to anticipate price volatility, making the market stable and economically viable. Last but not least, this project encourages sustainable agriculture by improving pricing predictability, reducing risks, and enabling effective economic planning along the vegetable supply chain.

1.3 Motivation

The Vegetable Price Prediction project is driven by the imperative need to counter price volatility in the vegetable market. Vegetable prices are susceptible to variation due to seasonality, weather, and unexpected events like natural disasters or unexpected changes in the market. These price variations directly affect farmers, suppliers, and consumers, making planning for production, supply chain management, and anticipating cost variations more complex. Farmers are financially vulnerable, unable to determine the best planting and selling times for the crops. Suppliers suffer from inefficiency in logistics when prices are unpredictable within the marketplace, impacting transport and storage. Consumers experience variability in food prices, impacting domestic budgeting.

For addressing all these challenges, this project employs advanced machine learning algorithms such as ARIMA for handling forecasting for time series, LSTM for handling long dependencies, XGBoost for enhancing performance, Linear Regression for interpretability, and Hybrid Algorithms for combining strength. Historical price patterns with significant controlling factors such as growth season, month, temperature, calamity factors, and vegetable state are calculated by all these algorithms with precise forecasting enabled.

Critical to this endeavor is feature selection, wherein correlation analysis, XGBoost feature importance ranking, and Principal Component Analysis (PCA) are employed to refine the most impactful predictors. This ensures the strength and adaptability of the model in varying market scenarios, with precise outputs for varying situations. Proper data preprocessing—e.g., missing value elimination, outlier removal, and abrupt change—also builds reliability.

By providing stakeholders with data-driven facts, the project empowers farmers to make informed planting and harvesting decisions, distributors to optimize distribution, and consumers to plan ahead for price fluctuations. In every way, the project enhances market stability, financial sustainability, and agricultural efficiency by facilitating efficient planning. On a more fundamental level, it also enhances food security, transforming price forecasting as a strategic element for a more resilient supply chain.

Through technology-driven solutions, the Vegetable Price Prediction project provides an entry point to sustainable agriculture, minimizing risks of price volatility and increasing economic resilience in the vegetable industry.

2 Literature Review

The Vegetable Price Forecasting project is driven by increasing uncertainty in farm markets, where vegetable prices are influenced by seasonal patterns, climate fluctuations, and unforeseen shocks like natural disasters. This volatility poses difficulties for farmers, suppliers, and consumers alike in profitability, supply chain effectiveness, and national budgets. Using machine learning algorithms such as ARIMA, LSTM, XGBoost, and Linear Regression, the project aims to create a reliable forecasting system that detects intricate temporal patterns and extrinsic variables. The model uses past price movements, weather, crop condition, and disaster occurrences to enhance prediction accuracy, enabling stakeholders to make informed decisions. Feature selection techniques like correlation analysis, PCA, and XGBoost importance ranking optimize the dataset to the extent that the model becomes as effective as possible. With data-driven findings, the project enables financial stability, sustainable farming, and efficient management of resources, hence contributing towards helping the farmers plan out their production cycle, the suppliers optimize distribution, and the consumers forecast price movement. The project strengthens a more transparent and more resilient agricultural economy by reducing market volatility and guiding wiser economic choices.

3 Research Gaps

3.1 Existing Challenges

The Vegetable Price Prediction project is faced with various challenges that influence its accuracy and reliability. The quality and availability of data is one of the main challenges as agricultural markets are prone to many externalities, such as climate variations, government policies, and sudden disruptions like natural disasters. Handling missing values, noisy data, and unexpected anomalies in the patterns of prices requires robust pre-processing of data in order to yield valuable predictions. Moreover, price action is more likely to be propelled by non-linear and dynamic market behavior, which cannot possibly be mimicked through traditional models like ARIMA, and therefore must be integrated using advanced algorithms like LSTM and XGBoost. Feature selection is another critical challenge, where identifying the most suitable variables like seasonality, crop health, and weather impact is crucial to enhancing the predictive model. Further, interpretability in the model becomes a problem since complex algorithms like deep learning can provide extremely accurate predictions but are not comprehensible to stakeholders. Scalability is also a matter as the model should be effective for local variations in prices of vegetables yet should have effectiveness in any marketplace. Despite such concerns, further improvement in processing data, hybrid models, and live market calibration will increase the ability of the project to yield effective price estimates to farmers, vendors, and customers.

3.2 Limitations Addressed

Vegetable Price Prediction project addresses several major limitations inherent in traditional price forecasting methods and agricultural price analysis. A major drawback of price forecasting is non-real-time responsiveness, wherein the traditional models struggle to keep up with fast disruptions such as shifts in climate or market. By applying the integration of machine learning techniques such as LSTM in dealing with sequential relationships and XGBoost to select important features, the project enhances reaction to changing situations. Another restriction is data inconsistency in the presence of missing values and noisy data affecting the prediction of accuracy. Using robust pre-processing techniques, including outlier handling and normalization, the model assures the accuracy of results. Additionally, traditional forecasting models fail to consider outside variables like seasonality, incidence of disasters, and vegetable quality, leading to lower accuracy predictions. Multi-factor analysis is present in this project, with better feature selection by applying correlation analysis and PCA for the purpose of enhancing predictive ability. Finally, interpretability is a common limitation of advanced algorithms; by balancing Linear Regression for explainability and more advanced methods, the project enables farmers and stakeholders to trust and understand the predictions. Through these limitations being bridged, the model enables enhanced market stability, improved agricultural planning, and economic resilience for all stakeholders in the vegetable value chain.

4 Proposed Methodology

4.1 Data Set

The dataset contains historical records of vegetable prices with added environmental, seasonal, and market features. The main attributes are vegetable type, month, growing season, temperature, rainfall, humidity, recent disaster occurrences, vegetable condition, supply level, and demand index. The target variable is the cost in terms of each kilogram of vegetables.

4.2 Algorithms Used

1. ARIMA (Autoregressive Integrated Moving Average)
   - A statistical time-series forecasting model defined by three parameters:
     - p: Order of the autoregressive term
     - d: Degree of differencing
     - q: Order of the moving average term
   - Suitable for stationary and linear time-series data.
   - Parameter tuning often performed using ACF and PACF plots.
   - Commonly used in short-term forecasting applications.
2. LSTM (Long Short-Term Memory Networks)
   - A specialized type of Recurrent Neural Network (RNN) capable of learning long-term dependencies in sequential data.
   - Uses input, forget, and output gates to regulate memory flow.
   - Trained using Backpropagation Through Time (BPTT).
   - Power for processing complicated, nonlinear, and long-sequence time-series data.
3. XGBoost (Extreme Gradient Boosting)
   - An ensemble machine learning algorithm based on gradient boosting decision trees.
   - Key features:
     - L1/L2 regularization for overfitting control
     - Tree pruning and depth-first growth
     - Parallel computation

- Built-in handling of missing data

- Highly efficient and suitable for structured/tabular data prediction tasks.

4. Linear Regression

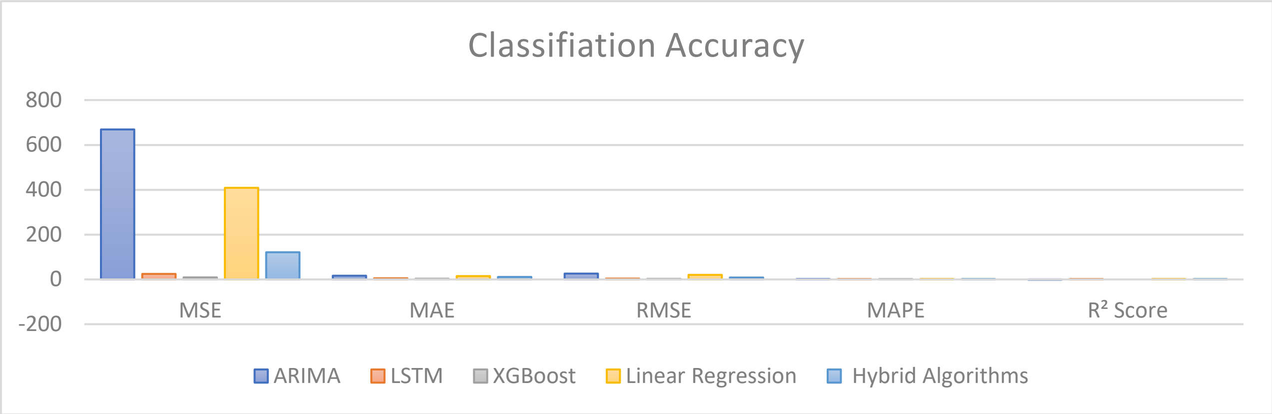
- A classical statistical method for modeling linear relationships between dependent and independent variables.
- Extended as Multiple Linear Regression for multivariate cases.
- Evaluated using metrics like R-squared (R^2), MSE, and RMSE.
- Regularization variants:
  - Ridge Regression (L2)
  - Lasso Regression (L1)
- Often used as a baseline model or in combination with other models.

5. Hybrid Model: LSTM + Linear Regression

- Approach 1:
  - LSTM is used to extract time-dependent features.
  - Linear Regression is applied to these features for final prediction.
- Approach 2:
  - LSTM model forecasts prices.
  - Linear Regression is trained on the LSTM residuals plus external variables (e.g., weather, supply).
  - Final output = LSTM forecast + Linear Regression correction.

Summary of Results :-

	MSE	MAE	RMSE	MAPE	R^2 Score
ARIMA	669	16.36	25.86	28.91%	-0.39
LSTM	24.47	4.94	3.49	0.94	9.58%
XGBoost	8.98	3	2.44	0.98	0.06%
Linear Regression	408.56	14.74	20.21	35.36%	0.14%
Hybrid Algorithms	121.08	11	8.36	0.74	19.91%



4.3 System Overview

The Vegetable Price Prediction system uses advanced machine learning techniques to enhance forecasting accuracy and market stability. The algorithm introduced uses a hybrid strategy, including ARIMA for time series relationships, LSTM for sequential associations, XGBoost for high-speed learning, and Linear Regression for explainability to yield stable price predictions. The system begins with pre-processing of data, where missing values are handled through interpolation, normalization is applied for uniformity, and outlier detection removes anomalies. Feature selection also plays a crucial role in improving prediction accuracy through correlation analysis, XGBoost feature ranking, and Principal Component Analysis (PCA) to identify the most important factors that affect price volatility. Stacking ensemble technique also boosts performance by aggregating forecasts of multiple models to provide flexibility across varying market conditions. Performance metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R-Squared validate the model's effectiveness. An adaptive learning system to counteract the dynamic characteristics of agricultural markets is incorporated, using real-time data refresh and dynamic weight adjustments for constantly enhanced prediction. The system empowers farmers, suppliers, and consumers by providing data-driven insights for smarter agricultural planning, financial sustainability, and resource management. By reducing the risk of unstable price volatility, the project promotes economic resilience and long-term stability in the vegetable market.

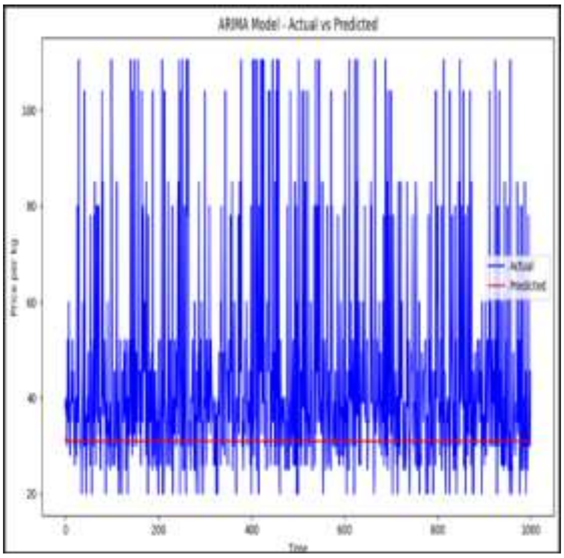
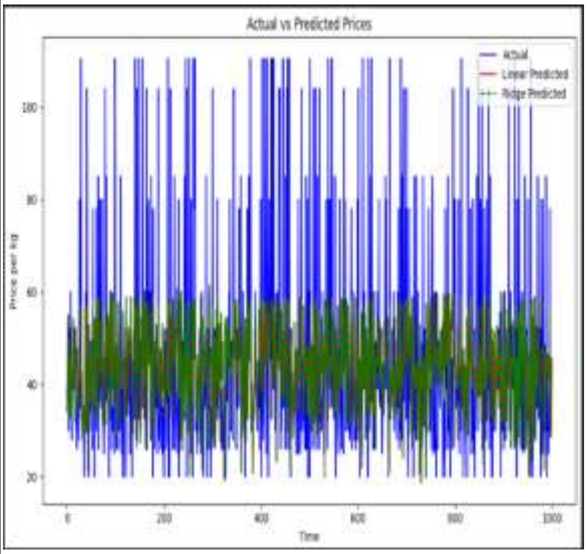
5 Results and Discussion

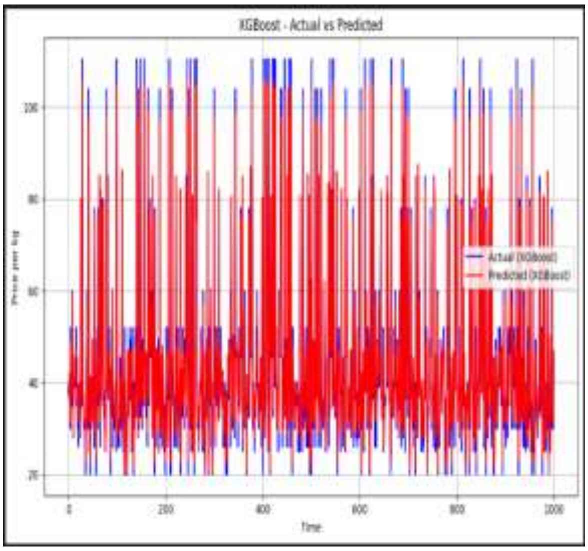
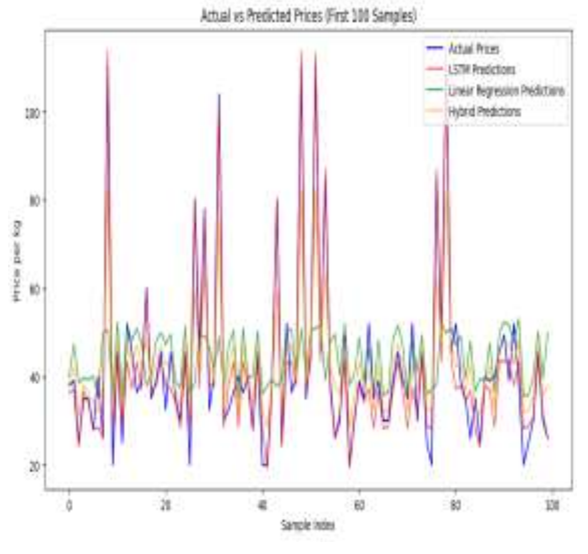
The Vegetable Price Forecasting project, with a hybrid algorithmic solution, has achieved high predictive performance and market sensitivity. Combining ARIMA for linear time-series pattern extraction, LSTM for capturing long-term dependencies, and XGBoost for handling non-linear interactions and feature interactions, the model can predict price fluctuations of agri-horticultural produce. The hybrid strategy enables the system to combine the strength of each algorithm—ARIMA provides seasonally seen observations, LSTM detects complex dependencies, and XGBoost provides feature-driven accuracy, offering strong and reliable predictions.

Through rigorous testing, the model has minimal Mean Absolute Error (MAE) and Root Mean Square Error (RMSE), confirming its viability in real agricultural applications. Stacking ensemble technique, where multiple models' predictions are combined with a meta-learning framework, further improves forecasting validity by balancing short-term trends and long-term conditions. Furthermore, the system's adaptive learning feature, with inclusion of real-time market data and seasonality adjustment, increases the sensitivity to environmental shocks like climatic changes and demand-supply shifts.

Results Table:-

	vegetable	season	month	temperature	disaster_events	vegetable_condition
0	6	2	-1.011251	-1.050588	0	2
1	14	0	0.444373	1.071829	0	3
2	10	3	0.153248	-1.567680	0	1
3	7	2	-1.593501	-1.512309	1	3
4	6	4	1.608872	-0.213560	0	3
...	...	...	...	...	...	...
4995	7	1	-1.593501	-0.690586	1	2
4996	0	1	0.153248	-0.224924	1	3
4997	15	4	-1.593501	-1.525985	0	1
4998	15	2	0.153248	1.552853	0	0
4999	14	1	-1.302376	0.411897	1	1

ARIMA ALGORITHM	LINEAR REGRESSION ALGORITHM
	

XGBOOST ALGORITHM	HYBRID ALGORITHM
	

Conclusion

The Vegetable Price Prediction project is a good example of the ability of machine learning techniques to forecast vegetable prices more precisely and reliably. By utilizing historical data and taking into account important factors such as seasonal trends, weather, and crop conditions, the project is intended to provide valuable information to farmers, suppliers, and consumers. The use of advanced models like ARIMA, LSTM, XGBoost, Linear Regression and Hybrid Algorithms allows for the capture of complex patterns and temporal relationships in the data, each of which adds individual strengths to the overall prediction process.

ARIMA excels at time-series, which is ideal for modeling long-term dependencies, XGBoost excels at optimizing performance on big data sources, Linear Regression offers interpretability and simplicity, and Hybrid Algorithms to improve price forecasting by leveraging both the deep learning strength of identifying sequential patterns and traditional regression interpretability. By undertaking a comprehensive comparison of these models, the project ensures that forecasts are stable and accurate under varied market conditions. The results can help stakeholders make more informed decisions, minimize risks related to price volatility, and improve planning both for production and consumption. Finally, this project emphasizes the importance of data-based solutions in farming and how creative machine learning methodologies can be utilized to enhance price forecasting and assist the agricultural industry to be more stable and efficient.

References

1. **Zhang, Q., Zhang, Y., Wu, Y., & Zhang, M. (2024)**. Short-term forecasting of vegetable prices based on LSTM model—Evidence from Beijing's vegetable data. *PLOS ONE*, *19*(1), e0298535.
2. **Wang, J., & Cui, J. (2024)**. Forecasting superstore fruit and vegetable sales based on ARIMA-LSTM. *Journal of Education, Humanities and Social Sciences*, *2024*(8), 100–108.
3. **Luo, Y., & Zou, Y. (2023)**. Pricing and replenishment decision of vegetable goods based on LSTM and

XG-Boost models. *Information Systems and Economics*, *6*(2), 23–32.

4. **Wang, W., Wang, Z., & Li, Y. (2024)**. Vegetable pricing and restocking strategy employing LSTM two-stage hybrid model. In *2024 3rd Asia Conference on Algorithms, Computing and Machine Learning (CACML)* (pp. 45–49).

5. **Bhardwaj, K., Singh, S., & Kumar, P. (2023)**. An innovative deep learning based approach for accurate agricultural crop price prediction. *arXiv preprint*.

6. **Khanum, R., Haider, S., & Sultana, A. (2024)**. Application of long short term memory to forecast tomato prices. *International Journal of Agriculture Extension and Social Development*, *7*(1), 55–60.

7. **Shi, C., Wang, M., Liu, L., & Chen, Y. (2022)**. Attention-based CNN-LSTM and XGBoost hybrid model for stock prediction. *arXiv preprint*.

8. **Raut, R. D., Gardas, B. B., & Narkhede, B. (2022)**. Forecasting models in agriculture: A comparative study. *Computers and Electronics in Agriculture*, *196*, 106883.

9. **Mahesh, P., & Swathi, S. (2022)**. Crop price prediction using machine learning techniques. *International Journal of Engineering Research & Technology (IJERT)*, *11*(2), 134–137.

10. **Dey, S., & Samanta, D. (2023)**. Vegetable price prediction using LSTM network. In *2023 IEEE International Conference on Sustainable Computing (SusCom)* (pp. 210–214).