



INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

“GPM: Applying A Hybrid Model To Predict Pricing For Garlic And Farmer’s Profitability”

Shashank J K^[1]

Parth^[2]

Authors:

Sagar M^[3]

Teja M^[4]

Dr. Manjunath K V^[5]

SOCSE, Presidency University, Bangalore, India

Guide:

Dr. Manjunath K V, Associate Professor, SOCSE, Presidency University

ABSTRACT

This study focuses on developing an automated garlic price prediction system in India using a hybrid model that integrates RF and LSTM networks. The system leverages historical price trends, and market conditions to enhance prediction accuracy. Individual models, including ARIMA, SVR, Prophet, XGBoost, were evaluated, with XGBoost demonstrating the highest standalone performance. However, due to the limitations of single models in capturing both non-linear relationships and sequential dependencies, a hybrid approach combining Random Forest Feature selection helps in choosing the most important data points, while LSTM (Long Short-Term Memory) is used to analyze patterns over time. Visualizations, including geospatial mapping and district-wise price distribution, provide insights into price variations across different regions. Future enhancements may include incorporating external socio-economic factors, real-time data integration, and advanced deep learning techniques to further improve forecasting accuracy. This research contributes to agricultural price intelligence systems, potentially benefiting farmers, policymakers, and market analysts in optimizing market strategies and mitigating price risks

Keywords:

Garlic Price Prediction, Machine Learning, LSTM, Mean Square Error, ARIMA , SVR, Prophet XGBoost, Hybrid Model(Random Forest + Long Short Term Memory)

1. Introduction

Agricultural commodity price forecasting plays a critical role in market stability, economic planning, and decision-making for farmers, traders, and policymakers. Accurate price prediction enables better financial planning, reduces risks associated with price volatility, and enhances supply chain efficiency. TWith the growing accessibility of extensive historical data on garlic prices and the need for precise price fluctuation predictions, there has been a transition from traditional statistical models like ARIMA to more advanced machine learning techniques [1, 10]. However, selecting the appropriate dataset for forecasting and optimizing machine learning models remain significant challenges. Key issues include parameter tuning, handling

nonlinearity, and avoiding the curse of dimensionality [5, 12].

Garlic prices in India are highly volatile due to external factors such as climate conditions, supply chain disruptions, and market demand fluctuations [17]. This unpredictability affects farmers, traders, and policymakers, influencing economic stability and agricultural planning. In recent years, variations in garlic price have had a considerable effect on India's agricultural economy, highlighting the need for precise predictive models to support stakeholders in making well-informed decisions.[16].

To overcome these challenges, this paper presents a hybrid machine learning model that integrates RF and LSTM networks for garlic price prediction. Random Forest is effective in feature selection and handling nonlinearity [3], while LSTM captures temporal dependencies in price trends [2, 11]. Hybrid models, which integrate statistical and deep learning techniques, have demonstrated superior predictive performance in agricultural price forecasting [13, 16]. By utilizing historical garlic price data and integrating external factors like weather patterns and market demand, the proposed model seeks to enhance prediction accuracy and offer valuable insights into India's garlic market [14, 18].

This system will not only aid farmers in planning their crop cycles more efficiently but will also help traders and policymakers in stabilizing market prices and preventing extreme fluctuations.

2. Related work and Literature Survey

Accurate garlic price forecasting is a critical research area driven by the need to enhance decision-making for farmers, traders, and policymakers. A range of statistical, ML, and DL models have been examined to forecast price fluctuations and reduce risks linked to market volatility. While conventional models like ARIMA and SVR have been commonly applied for time-series forecasting, recent progress in ML and DL has resulted in the development of more advanced predictive frameworks [10, 16]. However, each approach has its strengths and limitations, necessitating the exploration of hybrid models that integrate statistical techniques [5, 17].

2.1 Traditional Time-Series Models for Garlic Price Prediction

ARIMA model is among the commonly utilized approaches for price prediction. Studies have shown that ARIMA is effective for short-term forecasting due to its reliance on historical price trends [10]. However, it struggles with non-linear price fluctuations and external market shocks, which limit its predictive accuracy in volatile environments [7]. Additionally, research has demonstrated that ARIMA performs well in capturing seasonal variations. However, it does not account for external factors like weather conditions, transportation disruptions, and policy changes [12].

Support Vector Regression (SVR) has also been utilized for garlic price forecasting because of its capability to recognize non-linear patterns [19]. However, while SVR is effective in handling moderate-sized datasets, it demands precise parameter tuning and significant computational resources, which can limit its scalability for real-time applications [15].

Prophet is an additive regression model specifically developed for time-series forecasting, has also been explored for agricultural commodity price prediction. It is particularly useful for handling missing data and capturing seasonality [20]. However, Prophet's performance depends heavily on hyperparameter tuning and domain expertise to optimize forecasting accuracy.

2.2 Machine Learning Approaches for Garlic Price Forecasting

Recent advancements in ML have introduced more complicated models able of capturing complex, non-linear price fluctuations. models such as XGBoost and RF have gained popularity in agricultural price forecasting. Research indicates that XGBoost outperforms traditional statistical models by identifying intricate price

patterns and handling missing data more effectively [4, 12]. However, tree-based models require extensive hyperparameter tuning to prevent overfitting.

2.3 Deep Learning Approaches for Garlic Price Forecasting

Deep learning frameworks, particularly RNNs, have shown important potential in implementing long-term price trends. Among them, LSTM network are extensively used for their effectiveness in identifying long-term dependencies within sequential data[2, 11]. Research comparing LSTM with ARIMA and SVR has found that LSTM- based models outperform traditional techniques in capturing long- term trends and price fluctuations [18]. However, LSTM models require substantial training data and computational resources, making them challenging for small-scale implementation.

Convolutional Neural Networks (CNNs) have also been explored for agricultural price forecasting. Studies indicate that CNNs can effectively capture localized patterns in price fluctuations [20]. However, CNNs struggle with long-range dependencies, requiring hybrid approaches that integrate CNNs with recurrent architectures like LSTMs.

Model Type	Description	Hyperparameter Tuning
ARIMA	A statistical model used for time-series forecasting	p, d, q (parameters)
XG Boost	A gradient boosting algorithm that improves prediction accuracy by reducing errors iteratively.	Accuracy, RMSE, R^2 Score
SVR	A machine learning algorithm that captures complex patterns in price data.	Kernel Type, C (Regularization), Epsilon
Prophet	A regression model specifically designed to account for seasonality in time- series data.	Growth Type, Changepoint Prior Scale, Seasonality Mode
LSTM	A specialized type of recurrent neural network (RNN) built to identify and learn patterns in sequential data over long periods.	LSTM units control memory strength, while batch size decides how many samples the model learns from at once.

Table 2.1 Survey Models

3. Methodology

The methodology for this research integrates machine learning models, deep learning models, and hybrid models to provide a

comprehensive framework for real-time epidemic prediction and mapping. The steps outlined below describe the processes and

techniques used to collect data, develop models, and create visualizations that inform farmers and officials.

3.1 Data Collection

The first step in the methodology involves gathering the relevant data sources for modelling and analysis. The data collected includes:

Data Type	Description	Source
Garlic Price Data	Contains historical price data of garlic, .	https://agmarknet.gov.in/
Meteorological Data	Rainfall, temperature.	Meteorological Agencies, Kaggle

Table 3.1 Types of data

3.2. Model Development

3.2.1 Understanding Model Requirements

- Identified the need for a robust model to predict garlic prices accurately.
- Chose Random Forest (RF) for handling structured data and non-linearity.
- Chose LSTM for capturing long-term dependencies in sequential data.

3.2.2 Preprocessing

- Loaded historical garlic price data, ensuring completeness and consistency.
- Selected relevant features by eliminating non-informative variables.
- The data is divided into 80% for training and 20% for testing.
- Scaled numerical values using MinMaxScaler to standardize input data.
- Reshaped data for LSTM to fit the required 3D input format (samples, timesteps, features).

Data Cleaning	Removing duplicates, correcting errors, and handling missing values.	Pandas (Python)
Feature Engineering	Creating lagged features, aggregating data, and creating new variables	Python (Pandas, NumPy)
Normalization	Scaling features to ensure uniformity in range.	Min-Max Scaling

Table 3.2 Steps of Data Preprocessing

3.2.3 Implementing the Random Forest Model

- Built a Random Forest Regressor with 100 trees (n_estimators=100).
- Trained the model on the training dataset to capture feature relationships.
- Used the trained RF model to make initial price predictions.

3.2.4 Implementing the LSTM Model

- Designed an LSTM-based neural network with:
- LSTM Layer for learning time-dependent patterns.
- Dense Output Layer for final prediction.
- Dropout Layer to prevent overfitting.
- The model was configured using the Adam optimizer and MSE as the loss function.
- Used Early Stopping to halt training when validation loss stopped improving.

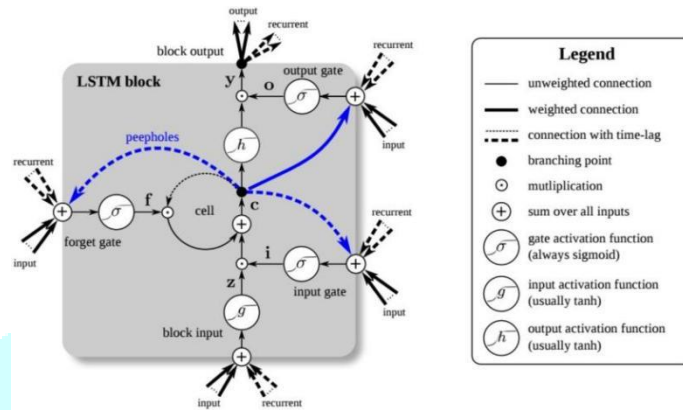


Figure 3.1 LSTM Model

3.2.5 Implementing the LSTM Model

- Designed an LSTM-based neural network with:
- LSTM Layer for learning time-dependent patterns.
- Dense Output Layer for final prediction.
- Dropout Layer to prevent overfitting.
- The model was configured using the Adam optimizer and MSE as the loss function.
- Used Early Stopping to halt training when validation loss stopped improving.

3.2.6 Combining RF and LSTM Predictions (Hybrid Model)

Applied weighted averaging to merge predictions from RF and LSTM:

$$\text{Hybrid Prediction} = \alpha \times \text{RF Prediction} + (1 - \alpha) \times \text{LSTM Prediction}$$

where $\alpha = 0.6$ (RF weight) and $(1 - \alpha) = 0.4$ (LSTM weight).

4. Results

4.1 Mean Absolute Error (MAE)

It measures how close predictions are to actual values by averaging their absolute differences. A lower MAE means a more accurate model.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

Where:

- N = Total no of observations
- y_i = The true value of the **i-th** observation
- $\hat{y}_i$ = Predicted value of the **i-th** observation
- $\sum$ = Summation symbol, summing over all observations

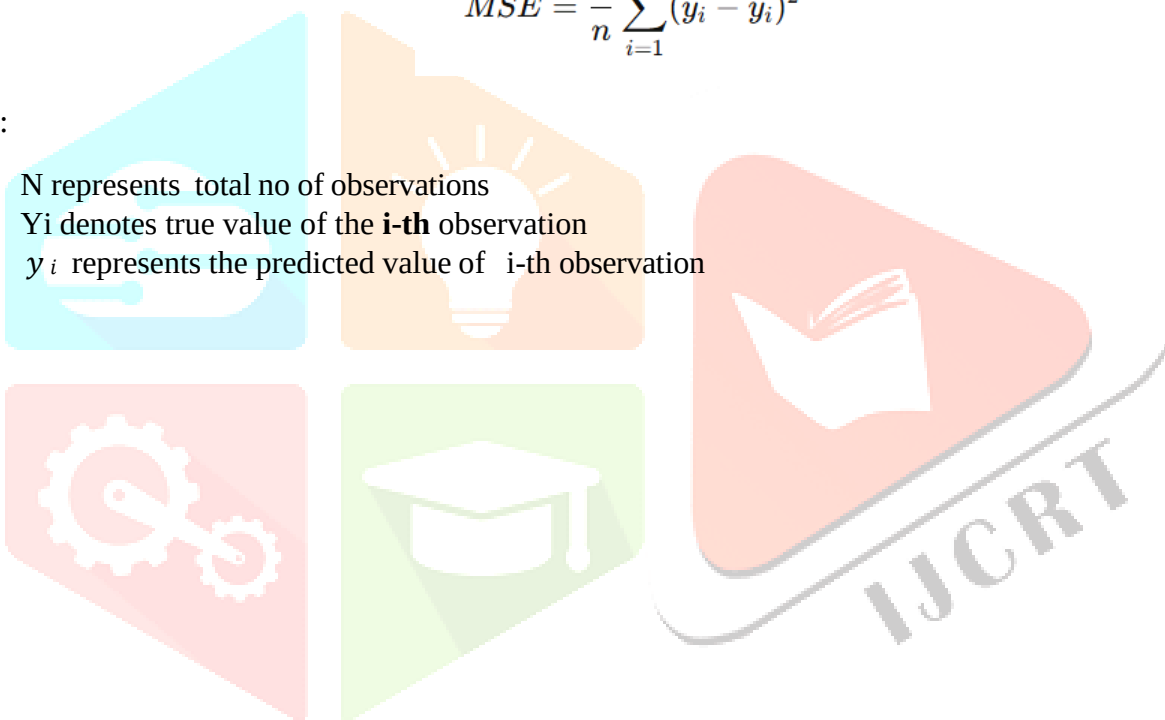
4.2 Mean Squared Error (MSE)

It evaluates accuracy of regression models like ARIMA and Random Forest where a lower MSE signifies improved performance.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Where:

- N represents total no of observations
- Y_i denotes true value of the **i-th** observation
- $\hat{y}_i$ represents the predicted value of **i-th** observation



Model	MAE	MSE
ARIMA	0.334	0.251
XG Boost	0.3127	0.220
SVR	0.3651	0.300
Prophet	0.3528	0.280
LSTM	0.3676	0.304
Hybrid Model	0.110	0.410

Table 4.1 Evaluation Metrics

4.3 Predictive Accuracy of Models

Accuracy is a common performance metric used in classification models, but in regression tasks (like price prediction), accuracy is usually derived from error metrics such as **MSE, MAE, or RMSE**.

$$\text{Accuracy} = \left(1 - \frac{\text{MSE}}{\text{Var}(Y)}\right) \times 100$$

Model	Accuracy Scores
ARIMA	66%
XG Boost	68.73%
SVR	63.49%
Prophet	64.72%
LSTM	63.24%
Hybrid Model	89%

Table 4.2 Accuracy Scores

The model accuracy comparison graph provides insights into the predictive models, this includes ARIMA, SVR, Prophet, XGBoost, LSTM, and a hybrid model combining RF and LSTM. The Hybrid model significantly outperforms the others, achieving an accuracy of **89%**, whereas the other models range between **63% to 69%** in accuracy.

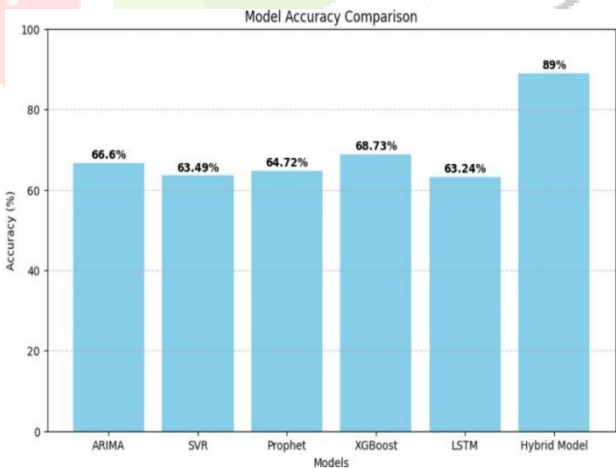


Figure 4.1 Accuracy score Bar Graph

ARIMA (66.6%) captures linear dependencies but struggles with sudden price fluctuations. **SVR (63.49%)** fails to capture long-term dependencies. **Prophet (64.72%)** lacks adaptability to sudden market changes. **XGBoost (68.73%)**struggles with sequential dependencies. **LSTM (63.24%)** specializes in longterm dependency

4.4 Visualization

Visualization plays a crucial role in analyzing and interpreting complex datasets by transforming raw data into graphical representations, making it easier to identify patterns, trends, and insights. In the given context of garlic price prediction, two types of visualizations have been used:

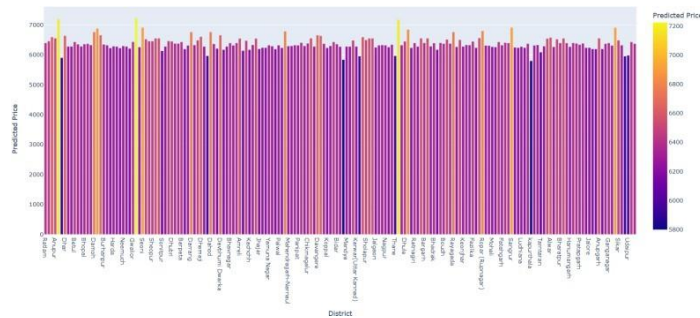


Figure 4.2 Price Prediction Bar Chart

4.4.1 Bar Chart for District-wise Price Prediction

- This bar chart represents the predicted garlic prices for different districts, with each bar representing a district and its height reflecting the predicted price.
 - The color gradient (ranging from purple to yellow) enhances readability by distinguishing lower and higher prices.
- This helps in quick comparisons across multiple regions, allowing stakeholders to pinpoint districts with the highest and lowest predicted prices



Figure 4.3 Map for Price Distribution

4.4.2 Geographical Map for Spatial Price Distribution

- This interactive map provides a spatial distribution of predicted prices across India, with each blue pin marking a district.
- This map helps traders, policymakers, and farmers make informed decisions regarding pricing trends, transportation, and market selection.

4. Conclusion

The automated garlic price prediction system developed in this project represents a significant advancement in agricultural price forecasting in India. By utilizing a hybrid Random Forest-LSTM model, the system effectively combines feature selection and sequential learning to improve price prediction accuracy. The results indicate that machine learning can be a powerful tool in helping farmers, traders, and policymakers make informed decisions, reducing financial risks associated with price volatility. Additionally, the system's ability to visualize predictions across different districts enhances its practical usability.

The importance of accurate price forecasting cannot be overstated, as it directly impacts farmers' income, market stability, and consumer affordability. Traditional forecasting methods, such as ARIMA and SVR, have demonstrated limitations in handling complex market dynamics, while deep learning-based approaches like LSTM offer better adaptability to changing trends. However, no single model is perfect, which is why this hybrid approach provides a balanced solution by combining the advantages of both machine learning and deep learning techniques.

This project lays the foundation for future developments in AI-driven agricultural forecasting systems. While the model has shown promising accuracy, its performance can be further improved by incorporating real-time data streams, external economic indicators, and advanced deep learning architectures. This research demonstrates that a data-driven approach to price prediction is not only feasible but also essential for building a more resilient and transparent agricultural market.

5.1 Challenges

1. **Data Limitations** – A major challenge lies in the availability and reliability of historical price data, as missing or inconsistent data points can affect the model's accuracy.
2. **External Influences** – Factors such as government regulations, global trade policies, and climate conditions are difficult to model, impacting the accuracy of predictions.
3. **Computational Complexity** – The hybrid approach requires significant computational resources, making it difficult to deploy in low-resource environments.

5.2 Future Directions

1. **Integration of Real-Time Data** – Enhancing the model with real-time market trends, weather patterns, and economic indicators will improve adaptability and forecasting precision.
2. **User-Centric Applications** – Developing mobile and web platforms will enable farmers and traders to access price forecasts conveniently, facilitating better decision-making.
3. **Explainability and Transparency** – Implementing explainable AI (XAI) techniques will enhance trust in the model by providing clear reasoning behind price predictions.

5. References

1. "Automated Agriculture Commodity Price Prediction System with Machine Learning Techniques" - Zhiyuan Chen^{*1}, Howe Seng Goh¹, Kai Ling Sin¹, Kelly Lim¹, Nicole Ka Hei Chung¹, Xin Yu Liew
2. "Support Vector Machines for Regression and Forecasting" – Vapnik, V.
3. "Prophet: Forecasting at Scale" – Taylor, S. J., & Letham, B.
4. "XGBoost: A Scalable Tree Boosting System" – Chen, T., & Guestrin, C.
5. "Long Short-Term Memory Networks for Time Series Prediction" – Hochreiter, S., & Schmidhuber, J.
6. "Hybrid Machine Learning Models for Price Prediction" – Zhang, G., Eddy Patuwo, B., & Hu, M. Y.
7. "Impact of Climate Factors on Agricultural Prices in India" – Kumar, R., & Singh, P.
8. "Comparative Study of Machine Learning Approaches for Market Forecasting" – Patel, J., Shah, S., Thakkar, P., & Kotecha, K.
9. "Random Forest Algorithm for Agricultural Price Prediction" – Breiman, L.
10. "Deep Learning in Agricultural Forecasting" – LeCun, Y., Bengio, Y., & Hinton, G.
11. "An Overview of Time Series Forecasting Techniques" – Hyndman, R. J., & Athanasopoulos, G.
12. "Integrating Machine Learning for Agricultural Market Intelligence" – Chakraborty, S., & Roy, D.
13. "Challenges in AI-Based Agricultural Forecasting" – Sharma, V., & Mehta, A.
14. "Big Data and AI in Agricultural Supply Chain Management" – Jha, S., & Agrawal, R.

15. "A Review of Hybrid Models in Predicting Commodity Prices" – Gupta, A., & Verma, S.
16. "Evaluating the Performance of Machine Learning Models in Price Forecasting" – Bose, K., & Sen, A.
17. "Leveraging LSTM for Market Trend Analysis" – Wang, Y., & Li, H.
18. "The Role of AI in Enhancing Agricultural Productivity" – Mishra, P., & Tiwari, R.
19. "Feature Engineering for Agricultural Price Prediction" – Das, A., & Mukherjee, S.
20. "Exploring the Influence of Macroeconomic Factors on Market Prices" – Banerjee, N., & Singh, A.
21. "Neural Networks for Time Series Forecasting: A Review" – Adhikari, R., & Agrawal, R. K.
22. "Machine Learning Applications in Agricultural Economics" – Fernandez, S., & Martin, J.
23. "The Use of LSTMs in Financial and Commodity Market Forecasting" – Huang, J., & Zhang, T.
24. "Bayesian Methods in Agricultural Price Prediction" – O'Hagan, A., & West, M.
25. "Explainable AI in Agricultural Forecasting Models" – Ribeiro, M. T., Singh, S., & Guestrin, C.
26. "A Comparative Analysis of Time Series Models for Price Prediction" – Box, G. E. P., & Jenkins, G. M.
27. "AI-Driven Decision Support Systems in Agriculture" – Bhattacharya, R., & Chattopadhyay, S.
28. "Understanding Price Volatility in Agricultural Markets" – Nelson, G. C., & Robertson, R.
29. "Predictive Analytics in Agri-Business: Challenges and Opportunities" – Kapoor, V., & Malhotra, R.
30. "The Role of Reinforcement Learning in Agricultural Price Forecasting" – Sutton, R. S., & Barto, A.