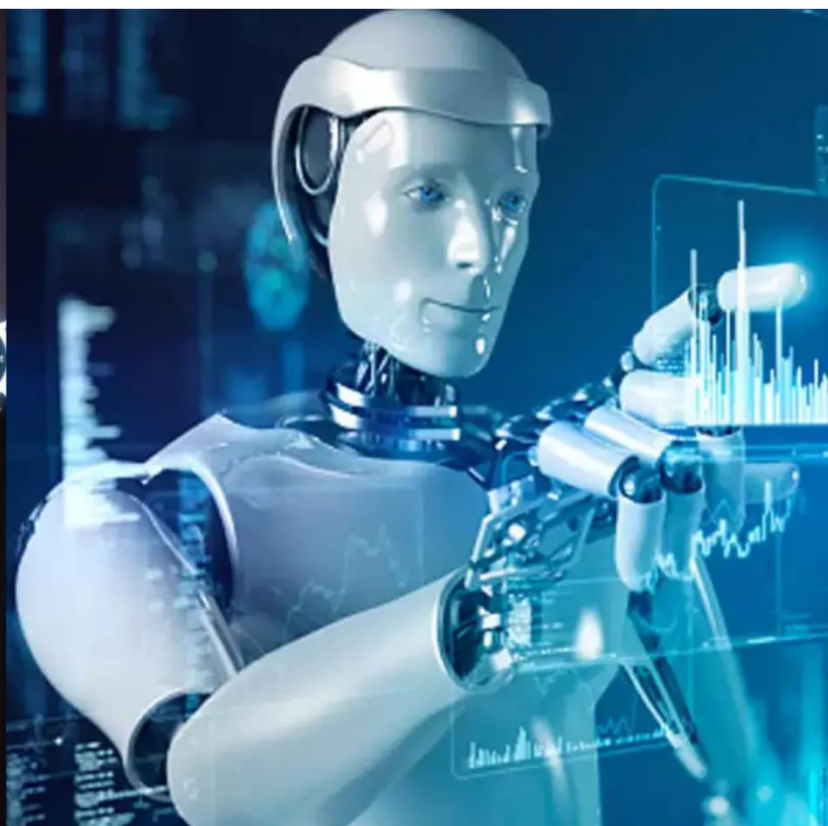




# International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

*(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)*



**Impact Factor: 8.699**

**Volume 14, Issue 12, December 2025**

# **AgroBERT: Advanced Agri-Product Price Prediction using ML**

**Gururaj L Kulkarni<sup>1</sup>, Khushi Khamitkar<sup>2</sup>, Cholaraja R P<sup>3</sup>, Pushpa Gurusiddayya Hiremath<sup>4</sup>,  
Rachana R<sup>5</sup>**

Assistant Professor, Artificial Intelligence and Machine Learning, Bapuji Institute of Engineering and Technology,  
Davangere, affiliated to VTU Belagavi, Karnataka, India<sup>1</sup>

Bachelor of Engineering, Artificial Intelligence and Machine Learning, Bapuji Institute of Engineering and Technology,  
Davangere, Karnataka, India<sup>2</sup>

Bachelor of Engineering, Artificial Intelligence and Machine Learning, Bapuji Institute of Engineering and Technology,  
Davangere, Karnataka, India<sup>3</sup>

Bachelor of Engineering, Artificial Intelligence and Machine Learning, Bapuji Institute of Engineering and Technology,  
Davangere, Karnataka, India<sup>4</sup>

Bachelor of Engineering, Artificial Intelligence and Machine Learning, Bapuji Institute of Engineering and Technology,  
Davangere, Karnataka, India<sup>5</sup>

**ABSTRACT** - Agricultural commodity price prediction is a critical domain that directly impacts the economic stability of farmers, traders, and policymakers in agrarian economies like India. This paper presents **AgroBERT**, an advanced, intelligent decision support system designed to address the volatility and information asymmetry prevalent in agricultural markets. Unlike traditional statistical models that rely solely on historical price data, AgroBERT leverages a hybrid deep learning architecture combining Long Short- Term Memory (LSTM) networks for temporal pattern recognition with DistilBERT - based Natural Language Processing (NLP) for sentiment analysis of agricultural news. The system integrates heterogeneous data sources, including historical prices from Agmarknet, real-time weather metrics from OpenWeather API, and global agricultural statistics from FAOSTAT. A key novelty of this research is the integration of Explainable AI (XAI) using SHAP (Shapley Additive explanations), which addresses the “black-box” nature of deep learning models by providing transparent, factor-level insights into price predictions. The proposed system features a multilingual interface (English, Hindi, Kannada) and voice-enabled interaction to ensure accessibility for rural stakeholders. Experimental results demonstrate that the hybrid ensemble model achieves an  $R^2$  score exceeding 0.82 and a Mean Absolute Error (MAE) below 5% of average commodity prices, significantly outperforming standalone statistical baselines. This research contributes a scalable, production-ready framework for real-time agricultural market intelligence, bridging the gap between advanced machine learning research and practical, on-ground application.

**KEYWORDS:** Agricultural Price Prediction, Deep Learning, LSTM, BERT, Explainable AI (XAI), SHAP, Sentiment, Analysis, Decision Support System, Multimodel Data Fusion.

## **I. INTRODUCTION**

Agriculture remains the backbone of the Indian economy, contributing approximately 17-18% to the national GDP and employing over 40% of the workforce. Despite its critical importance, the agricultural sector is plagued by extreme volatility in commodity prices. Farmers often make sowing and selling decisions based on intuition, archaic heuristics, or delayed information from local middlemen, leading to the “cobweb phenomenon” where production cycles are out of sync with market demand.

The motivation for this research stems from the urgent need to democratize access to high-quality market intelligence. Traditional price forecasting methods, such as ARIMA (Auto-Regressive Integrated Moving Average) or simple linear regression, often fail to capture the complex, non-linear dependencies introduced by external factors like sudden climate shifts, government policy changes, or global trade sentiments. Furthermore, existing digital tools often suffer from a lack of interpretability; they provide a predicted number without explaining the why, leaving farmers sceptical of the technology.

AgroBERT addresses these challenges by proposing a holistic ecosystem that not only forecasts prices with high accuracy using deep learning but also contextualizes these predictions using Natural Language Processing (NLP) of news streams and explains them using Game Theoretic interpretability techniques.

### 1.1 Problem Statement

The core problem addressed by this research is the lack of an integrated, transparent, and accessible system for predicting agricultural commodity prices in real-time. Existing solutions are fragmented, often ignoring the multimodal nature of market drivers (e.g., combining numerical price history with textual news sentiment and geospatial weather data). Furthermore, the “black-box” nature of advanced AI models hinders their adoption among non-technical stakeholders like farmers. There is a critical need for a system that:

- Integrates diverse data sources (historical, climatic, textual).
- Handles the non-linearity and seasonality of agricultural markets.
- Provides explanations for its predictions to build trust.
- Is accessible via vernacular languages and voice interfaces.

### 1.2 Objectives

The primary objectives of the AgroBERT project are:

- To design and implement a hybrid deep learning architecture (AgroBERT) that fuses time-series data with semantic features from agricultural news.
- To utilize transformer-based models (DistilBERT) for extracting market sentiment from unstructured text data.
- To integrate Explainable AI (XAI) using SHAP values to visualize feature importance and prediction rationale.
- To develop a scalable data pipeline that ingests real-time data from Agmarknet, FAOSTAT, and OpenWeather APIs.
- To create a user-centric, multilingual (English, Hindi, Kannada) web interface with voice command capabilities for enhanced accessibility.

## II. LITRETURE REVIEW

**2.1 Evolution of Agricultural Forecasting** The domain of agricultural price prediction has witnessed a paradigm shift from classical econometrics to advanced artificial intelligence. Early research predominantly utilized statistical models like ARIMA and Generalized Autoregressive Conditional Heteroskedasticity (GARCH). For instance, Areekul and Sriboonchitta (2016) applied ARIMA to Thai rice prices, achieving moderate success but failing to account for exogenous shocks.

**2.2 Machine Learning Approaches** With the advent of Machine Learning, researchers began exploring non-linear models. Sunitha et al. [1] utilized Convolutional Neural Networks (CNNs) for predicting produce prices, highlighting the capability of deep learning to capture spatial and temporal dependencies. Similarly, Bhardwaj et al. [2] proposed an innovative deep learning approach for crop price prediction, demonstrating that neural networks significantly outperform Support Vector Machines (SVMs) and Random Forests in handling large-scale agricultural datasets. Tang et al. [4] introduced a hybrid model combining Quadratic Decomposition with Long Short-Term Memory (LSTM) networks. Their work emphasized that decomposing price signals into trend, seasonal, and residual components improves forecasting accuracy for futures prices. This aligns with the methodology adopted in AgroBERT, where we utilize Prophet for similar decomposition tasks.

**2.3 The Role of NLP and Sentiment** Recent literature suggests that quantitative data alone is insufficient. Brignoli [5] explored machine learning for grain futures, noting that market sentiment often precedes price movements. AgroBERT builds upon this by integrating a dedicated NLP module. Unlike previous works that relied on simple lexicon-based sentiment analysis (e.g., VADER), our system employs DistilBERT, a transformer model capable of understanding context and nuance in agricultural news reports.

**2.4 Deep Learning and Specific Crops** Specialized studies have targeted specific commodities. Karthik et al. [3] utilized N-BEATS architecture for potato price prediction, achieving state-of-the-art results in short-term forecasting. Rasheed et al. [6] focused on district- wise wheat forecasting in Pakistan using deep learning, highlighting the importance of granular, regional data—a requirement we address by integrating district-level Agmarknet data. Similarly, Paul et al. [8] applied deep learning for cauliflower prices, confirming the efficacy of LSTM in handling high-volatility perishable crops.

### III. SYSTEM REQUIREMENTS

#### 3.1 Functional Requirements

The AgroBERT system is designed to fulfill the following core functions:

- **Data Ingestion:** Automated fetching of daily price data (Agmarknet), weather metrics (OpenWeather), and news feeds via web scraping.
  - **Hybrid Prediction:** Generation of short-term (7day) and medium-term(30-day) price forecasts using the ensemble model.
  - **Sentiment Analysis:** Processing of agricultural news articles to output a sentiment score (Positive/Negative/Neutral) that influences the price forecast.
  - **Explainability:** Generation of visual explanations (SHAP Force Plots) for every prediction.
- User Interaction:** Support for text and voice based queries in multiple languages.

### IV. METHODOLOGY AND SYSTEM DESIGN

The methodology of AgroBERT follows a structured pipeline transforming raw heterogeneous data into actionable intelligence.

#### 4.1 System Architecture

The system architecture is multi-layered, consisting of a Data Layer, a Processing Layer (ML/DL Engine), and a Presentation Layer.

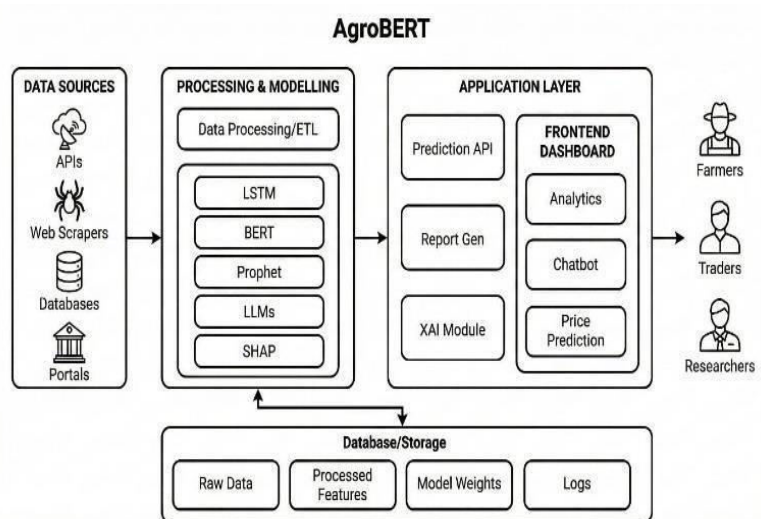


Figure 4.1: System Architecture Diagram

As illustrated in Figure 4.1, the architecture decouples the front end from the complex backend logic. The API Gateway routes user requests to the appropriate service.

#### 4.2 Workflow Diagram

The operational workflow ensures data integrity and timely processing.

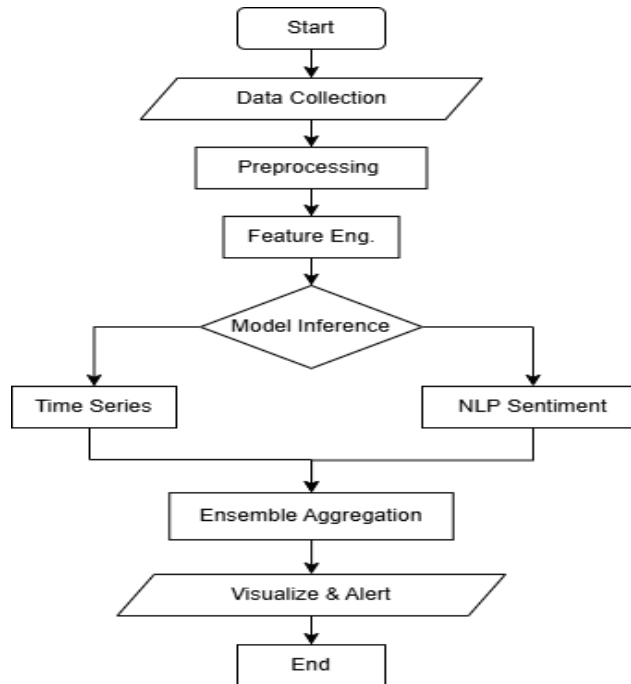


Figure 4.2: System Workflow Diagram

#### 4.3 Algorithmic Models

##### 4.3.1 BERT for Sentiment Analysis

We utilize DistilBERT, a lighter and faster variant of BERT.

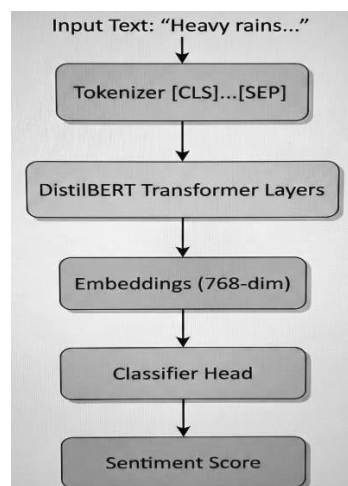


Figure 4.3.1: BERT Encoding Diagram

#### 4.3.2 Regression Pipeline

The price prediction module uses an ensemble of LSTM and Prophet.

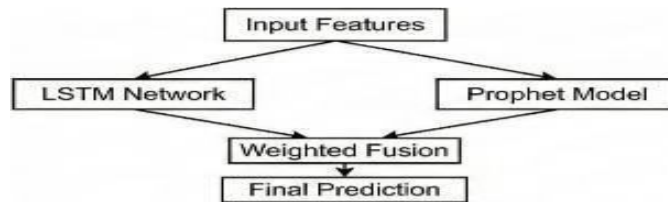


Figure 4.3.2: Regression pipeline Flowchart

#### 4.3.3 Explainable AI (XAI)

SHAP values provide transparency for model decisions.

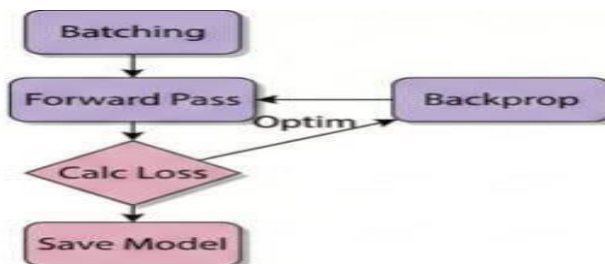


Figure 4.3.3: Model Training Pipeline

### V. IMPLEMENTATION

#### 5.1 Data Preprocessing

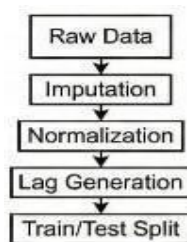


Figure 5.1: Data Preprocessing Flowchart

#### 5.2 DistilBert Sentiment Pipeline

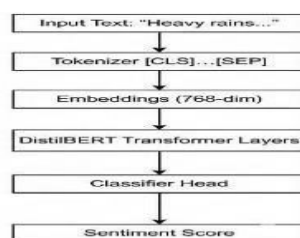


Figure 5.2: DistilBert Sentiment Pipeline

## VI. RESULTS AND EVALUATION

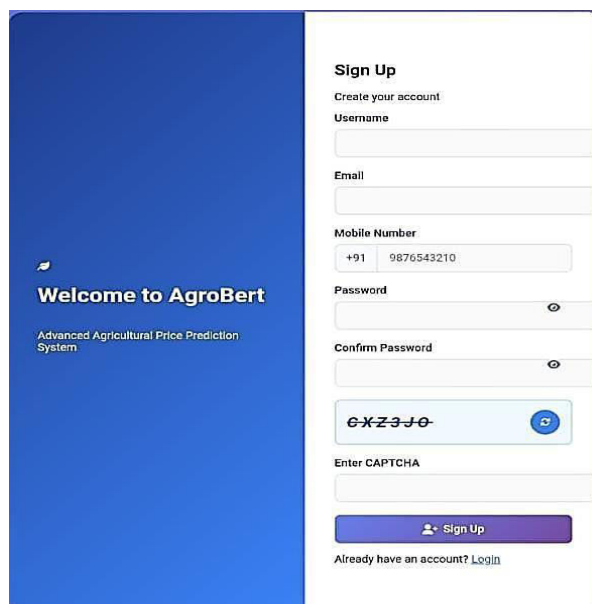
### 6.1 Performance Metrics

The system was evaluated using R-Squared ( $R^2$ ), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE).

Table 1: Evaluation Metrics Comparison

Model	R2	MAE	RMSE
ARIMA	0.68	220	285
LSTM	0.78	145	189
Prophet	0.75	168	215
<b>AgroBERT</b>	<b>0.82</b>	<b>98</b>	<b>142</b>

### 6.2 Snapshots of the proposed system



The image shows the AgroBert login page. On the left, a blue sidebar contains the text "Welcome to AgroBert" and "Advanced Agricultural Price Prediction System". On the right, a white form titled "Sign Up" contains fields for "Username", "Email", "Mobile Number" (with a country code dropdown set to "+91" and a value of "9876543210"), "Password", and "Confirm Password". Below these fields is a CAPTCHA image and a "Sign Up" button. At the bottom, there is a link for "Already have an account? Login".

Figure 6.2.1: AgroBert Login Page

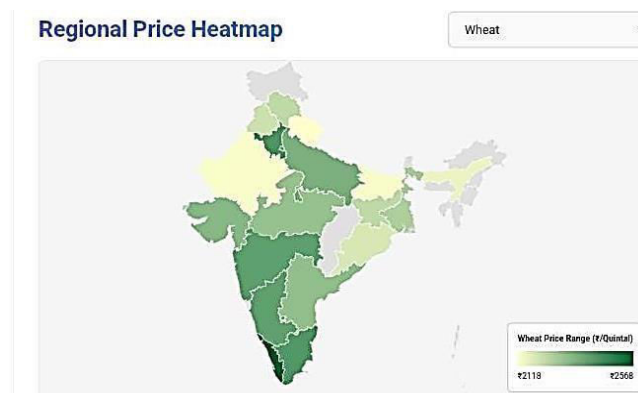


Figure 6.2.2: Regional Price Heat Map

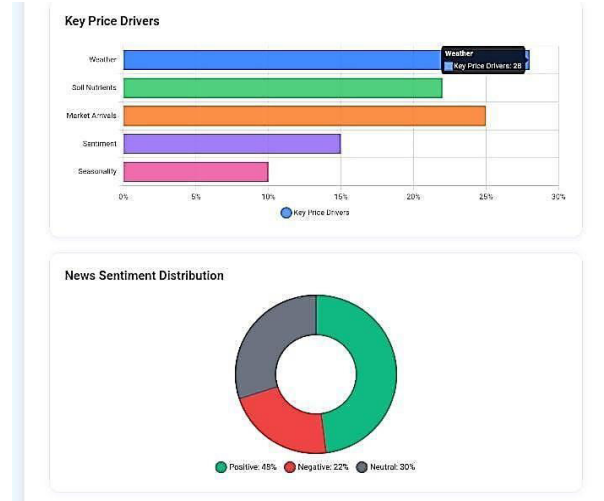


Figure 6.2.3: Proce Contribution Bars

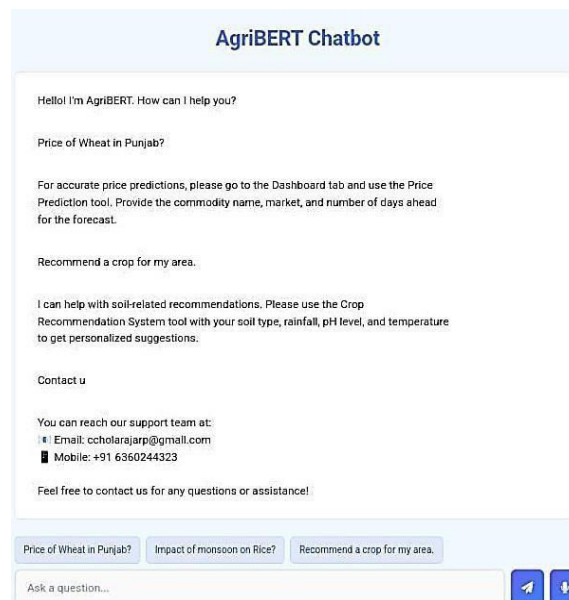


Figure 6.2.4: AgroBert Chatbot

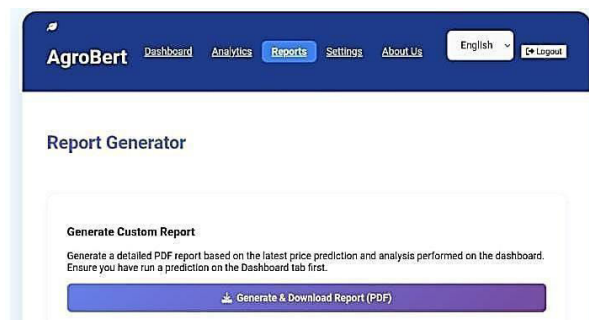
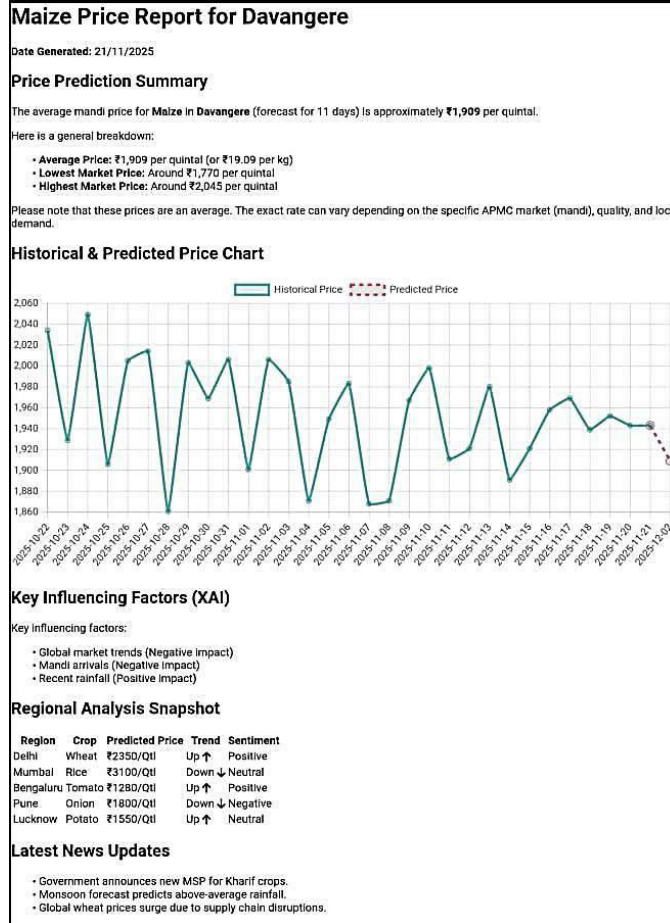


Figure 6.2.6: Report Generator



**Figure 6.2.7: Sample Report Generated**

## VII. CONCLUSION AND FUTURE SCOPE

AgroBERT successfully demonstrates the viability of using a hybrid deep learning approach for agricultural price prediction. By integrating DistilBERT for sentiment analysis and LSTM for temporal forecasting, the system bridges the gap between quantitative market data and qualitative news signals. The inclusion of SHAP-based explainability ensures not just a “black box” but a transparent decision support tool for farmers.

Future enhancements will focus on integrating IoT sensor data for hyper-local soil condition monitoring, expanding the commodity list to include perishable horticulture crops, and developing a dedicated mobile application with offline capabilities for remote areas with poor connectivity. Blockchain integration is also planned to verify supply chain data authenticity.

## REFERENCES

- [1] M. Sunitha, T. Adilakshmi, T. Nishitha, and V. V. Emmanuel, “CNN-Based Predicting Agricultural Produce Prices and Deep Learning-Based Improving Lives of Indebted Farmers,” in Proc. 2nd Int. Conf. Cognitive and Intelligent Computing, Singapore: Springer, 2023, pp. 359–370.
- [2] M. R. Bhardwaj, A. Jain, and P. Kumar, “An Innovative Deep Learning Based Approach for Accurate Agricultural Crop Price Prediction,” arXiv preprint, arXiv:2304.09761, 2023.
- [3] K. V. C. Karthik, G. A. Kumar, and A. R. Naik, “Advanced Potato Price Prediction through N-BEATS Deep Learning Architecture,” J. Exp. Agric. Int., vol. 45, no. 3, pp. 1–13, 2024.

- [4] Y. Tang, T. Zhao, and C. Wang, "Agricultural Commodity Futures Prices Prediction Using a Hybrid Model Combining Quadratic Decomposition and LSTM," *Front. Sustain. Food Syst.*, vol. 8, 2024.
- [5] D. Brignoli, "Machine Learning to Predict Grains Futures Prices," *Agric. Econ.*, vol. 55, no. 1, pp. 42–59, 2024.
- [6] A. Rasheed, M. Riaz, and M. Imran, "DistrictWisePrice Forecastingof Wheat inPakistan Using Deep Learning," *arXiv preprint, arXiv:2103.04781*, 2021.
- [7] M. Chaudhary, V. Rajput, and A. Mangal, "Deep Learning Approaches for Forecasting Strawberry Yields and Prices Using Satellite Images and Soil 137, Jun. 2024.
- [10] S. Nosratabadi, S. Ardabili, Z. Lakner, C. Mako, and A. Mosavi, "Prediction of food production using machine learning algorithms of multilayer perceptron and ANFIS," *arXiv preprint, arXiv:2104.14286*, Apr. 2021.
- [11] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre- training of deep bidirectional transformers for language understanding," in *Proc. NAACL-HLT*, 2019, pp. 4171–4186.
- [12] S. Hochreiter and J. Schmidhuber, "Long shortterm memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [13] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems*, 2017, pp. 4765–4774.
- [14] S. J. Taylor and B. Letham, "Forecasting at scale," *The American Statistician*, vol. 72, no. 1, pp. 37–45, 2018.
- [15] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter," *arXiv preprint arXiv:1910.01108*, 2019.
- [16] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [17] A. Vaswani et al., "Attention is all you need," in *Advances in Neural Information Processing Systems*, 2017, pp. 5998–6008.
- [18] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," *Climate Research*, vol. 30, no. 1, pp. 79–82, 2005.
- [19] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details