J Res Agric Food Sci

Development Of AI-ML Based Models For Predicting Prices Of Agri-Horticultural Commodities

Journal Name :	Journal of Research in Agriculture and Food Sciences
Manuscript Type :	Review Article
Abstract :	Forestry helps to sustain security and stabilize the economy. Smart Technologies such as machine learning (ML) and artificial intelligence (AI) are useful in accurately forecasting the prices of agricultural products which is of great importance to farmers, traders and legislators. This research analyses the applicability of AI and ML techniques in modern-day agricultural price prediction Challenges. These models can project future prices, taking into consideration historical data along with affecting economic and environmental conditions. This accuracy empowers farmers to make informed decisions related to crop cultivation such that they maximize yield whilst minimizing losses and devise effective market plans. Furthermore, this predictive ability serves to mitigate the impact of sudden price changes, which enhances market stability and broadens food availability results from this study will support the development of rational policies that AI-infused decisions demonstrate the contributions AI and ML have towards agricultural price forecasting
Keywords :	AI & ML, XGBoost, LSTM, Random Forest, Flutter, Dart, Flask

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

Madhusudhan MV
Associate Professor, CSE
Presidency University
Bengaluru, Karnataka, India
mv.madhu@gmail.com

Darshan V
Presidency University
Bengaluru, Karnataka, India
darshan.v158@gmail.com

Byresh Soradi
Presidency University
Bengaluru, Karnataka, India
byreshsoradi1234@gmail.com

D Tejas
Presidency University
Bengaluru, Karnataka, India
tejasd1975@gmail.com

Deva Prakash
Presidency University
Bengaluru, Karnataka, India
prakashdev298@gmail.com

Abstract— **Forestry helps to sustain security and stabilize the economy. Smart Technologies such as machine learning (ML) and artificial intelligence (AI) are useful in accurately forecasting the prices of agricultural products which is of great importance to farmers, traders and legislators. This research analyses the applicability of AI and ML techniques in modern-day agricultural price prediction Challenges. These models can project future prices, taking into consideration historical data along with affecting economic and environmental conditions. This accuracy empowers farmers to make informed decisions related to crop cultivation such that they maximize yield whilst minimizing losses and devise effective market plans. Furthermore, this predictive ability serves to mitigate the impact of sudden price changes, which enhances market stability and broadens food availability results from this study will support the development of rational policies that AI-infused decisions demonstrate the contributions AI and ML have towards agricultural price forecasting**

Keywords: AI & ML, XGBoost, LSTM, Random Forest, Flutter, Dart, Flask.

I. INTRODUCTION

Agricultural goods are a part of everyday life and contribute significantly to the economy of any nation.Price volatility in agricultural goods has a profound effect on farmers and consumers alie and usually contribute towards economic unrest.Over the past few years,unstable weather conditions have increased the difficulty of forecasting prices.Thus,there is a need for government policies to ensure there is equilibrium in supply and demand.Price fluctuations in agricultural commodities directly impact the economy,and especially the rural economy.Priced Towns can have devastating effects on farmers and their families as well as their communities.Climate change,water scarcity,and low productivity are some of the numerous challenges the agricultural sector is facing,all of which contribute to price volatility. This research paper discusses how Long Short Term Memory(LSTM) models Improve price prediction accuracy for agricultural raw material award winners. These models can provide farmers, dealers and political decisions through research into historical price datasets.

Many machine learning algorithms such as Autoregressive Integrated Moving Average (ARIMA), Random Forest(RF), LSTM, Artificial Neural Networks(ANN), Backpropagation(BP), Recurrent Neural Networks (RNN), Convolutional Neural Networks (CNN), Support vector Machines (SVM), Decision Trees (DT), hybrid models and some deep learning models can be used for that purpose. In this project we employed LSTM for price prediction and Random Forest for suggesting crops. Price trends are perennialized, and model outputs are evaluated by error measurements, namely Medium absolute error (MAE), R-squared value, Medium square error (MSE), Root agent quadratic error (RMSE), and Medium absolute percentage error (MAPE). Correct price predictions can prevent the risk of sudden fluctuations from hindering prices. Agricultural industry ultimately leads to improved economic outcomes

II. RELATED WORKS

Several traditional methods, statistical methods, machine learning, and deep learning models have been explored for harvest proposals and price predictions. This section comprehensively checks several approaches in this field.

In [1],the authors discussed conventional approaches such as regression and time series analysis for agricultural price prediction. Easy to interpret and simple in nature, these approaches perform well for linear and structured datasets. Yet, they are not flexible enough to cope with non-linear relationships, high dimensionalities, and abrupt market swings, rendering them less accurate for intricate agriculture datasets. Conventional ecological oriented farming practices were surveyed by authors in [2] against the background of agroecology, emphasizing sustainable, biodiverse, and resilient agroecosystems. Such practices converse biodiversity, enhance soil health, and ensure environmental sustainability. They might not be scalable or precise, nor directly related to price forecast models, which limits their use in data-informed agricultural forecasting. In [3], the authors highlighted the importance of indigenous knowledge systems in West African agriculture, a time-tested practice. They are adaptive, community-based, and absence of formal documentation and conjoining with contemporary technology hinders their implementation for precise price forecasting and widespread adoption.

Statistical tools like ANOVA, regression, and time series models were presented in [4]. Such approaches improve examination of crop tendencies, prices, and production proposals. They are simple to understand, straightforward, and apply to structured datasets. They struggle with high-variable interrelations, curtailing their accuracy in forecasting agricultural prices at the moment they occur. In [5], the authors utilized regression, correlation, and time series analyses to illustrate the adverse impacts of climate change on world crop yields. These statistical methods identify long -term trends and aid climate-resilient planning in agriculture. Yet, such models might lack efficiency when forecasting nonlinear and unpredictable changes, calling for addition with machine learning for enhancing accuracy. Authors of [6] gave an exhaustive overview of time series forecasting methods, including ARIMA, SARIMA, ARCH-GARCH, and Exponential Smoothing. Such models are an advantage for predicting agricultural prices and returns, as they can take into account time trends. However, they tend to take over linearity and static, which reduces performance in the presence of sound, nonlinear, or seasonally different data.

Strong background in linear regression methods was introduced in [7], including multiple linear regression (MLR), polynomial regression, and regularized models. They are useful for predicting crop prices and farm yields. Their simplicity and convenience are why they are widely used, but assumption of linearity makes them unreliable in complicated agricultural settings. Support Vector Machine (SVM) was utilized by authors in [8]. Its advantages is in the ability to work with high-agriculture, it has been employed successfully for crop classification, estimation of yield, and forecasting price. SVMs are, however, computationally expensive and prone to parameter adjustment, which might hinder their scalability in big datasets. The authors introduced the Decision Tree (DT) learning approach in [9], specifically the ID3 algorithm. This method is extensively utilized in agriculture for its interpretability and simplicity, being used for crop recommendations as well as for yield/price predictions. It works effectively on heterogeneous data but overfits and can yield less accurate outcomes for noisy or unbalanced data.

Authors in [10] created Random Forest (RF), a mechanical ensemble approach with several decisions. RF improves accuracy and stability through cancellation and random feature selection. This makes it particularly useful for agriculture to predict prices and returns. Although it has the advantage of less super-adjusting, RF can be less interpreted and calculated in processing very large data records. The authors of [11] were investigated by the demonstrated authors of [11] such as RF, SVM, and DT models to improve plant selection, price prediction, efficiency, resource consumption, and decision-making. ML promotes precision reproduction in real time through data automation and analysis. However, challenges include data quality, model selection, and interpretability for non-expert users. The authors of [12] showed how ML and computer vision, in particular random forests, improve fertility detection. This initial identification improves quality and minimizes losses. Although RF is robust and effective, model performance can be exacerbated by insufficient training data or unselected features.

A hybrid model combining Convolutional Neural Networks (CNN) was proposed by the authors of [13] for feature extraction with SVM for classification to detect fruit diseases. Though the deep learning component plays a role, the focus here is on the ML application of SVM. This hybrid method improves accuracy and reduces false detections. However, the complexity of combining models can increase development time and computational needs. A machine learning-based crop recommendation system using RF, SVM, and Naïve Bayes was developed by the authors of [14]. These algorithms assisted in recommending the most appropriate crops from environmental information, greatly enhancing agricultural productivity. Although accurate and scalable, the dependency on high-quality, domain-specific data is a shortcoming. The authors in [15] pointed out the promise of Artificial Neural Networks (ANNs) for predicting crop yields from environmental and soil parameters. ANNs are capable of learning intricate patterns and interactions, but they need large amounts of data and tuning. Moreover, their black box nature can decrease interpretability, particularly in high-value agricultural choices. AI and ML use in agricultural price prediction, including methods like linear regression, Random Forest, and LSTM was discussed by the authors of [16]. They emphasized that ML has an edge in market transparency and resource optimization. Notwithstanding their power, ML models are highly sensitive to data quality and demand appropriate tuning to perform reliably.

Comparison of general regressions (GRNN) for neuronal networks (GRNNs) and support for vector regressions (SVR) with harvest price predictions were used by the authors.[17]. GRNN performed conventional and linear approaches in speed as well as in accuracy. Nonetheless, similar to the most neural network-based methods, however, it does need significant computation power and lacks transparency. Machine learning and Deep learning were implemented by the authors in [18], to predict agricultural product and supply prices. Although DL methods were part of their study, the use of classical ML algorithms also contributed significantly. These techniques were found effective in risk management and market strategy. However, the dependence on historical data limits their adaptability to unexpected market shocks. The Machine Learning models were implemented by the authors in [19] to predict coffee prices in Vietnam using historical data. These models enhanced forecasting accuracy, aiding farmers and traders in decision-making. While effective, the accuracy of these models can drop when sudden or unforeseen economic or environmental events occur.

Recent advances in machine learning applied to agricultural price forecasting were covered by the authors in [20]. They highlighted how ML increases the accuracy of forecasts and provides near real-time updates. Nonetheless, they indicated limitations such as overfitting, data reliance, and necessity for regional adjustment for general usage. Econometric and machine learning combined hybrid models were suggested by the authors in [21] for improving commodity price forecasting. This hybrid solution utilized the best of both conventional economic models and machine learning for improved accuracy. This hybrid, though, can become complicated and computationally intensive, making it challenging in rural or low-resource environments. In [22], the authors employed a genetic algorithm to tune an Extreme Learning Machine (ELM) for commodity price prediction. This model achieved high accuracy and flexibility. However, the process of optimization is computer-intensive and may necessitate expertise in metaheuristic algorithms for successful implementation.

A genetic algorithm was used by the authors in [23] to optimize an Extreme Learning Machine (ELM) for commodity price forecasting. The model's success in managing pricing strategies and risks indicates the promise of ML in agricultural market analysis. Nevertheless, such models demand substantial data engineering and system integration. An integrated model combining historical pricing and weather data was developed by the authors in [24] to improve cabbage and radish price forecasting. The architecture, while based on DL (DIA-LSTM), also incorporated ML-level preprocessing and feature interaction. This hybrid approach yielded better forecasts but increased model complexity. A Radial Basis Function (RBF) neural network was used by the authors in [25] to predict garlic and pork prices based on eight influential features. While this lies between ML and DL, the RBF's structure is more ML-oriented. It performed well in minimizing error, but its simplicity may limit performance in highly nonlinear systems.

A meta-learning framework was implemented by the authors in [26] by combining Random Forest, SVM, ANN, SVR, and ELM for agricultural price prediction. This advanced ML ensemble adapted model selection based on data features. Although highly effective, the system's complexity and need for extensive validation may hinder its application in small-scale farming contexts. A hybrid machine learning model was developed by the authors in [27], integrating advanced preprocessing techniques with multiple AI algorithms for commodity price forecasting. The model showed improved accuracy, but such multi-layered systems can be difficult to maintain and adapt over time, especially in regions with limited technical infrastructure. Long Short-Term Memory (LSTM) models were applied by the authors in [28] to forecast crop prices. LSTM's ability to capture long-term dependencies in sequential data makes it well-suited for agricultural price prediction, especially with fluctuating and seasonally influenced data. The main advantage is its high accuracy in detecting non-linear trends. However, LSTM models require large datasets and are computationally expensive, which may limit their application in regions with limited data infrastructure.

Deep learning methods for plant disease detection using image recognition techniques were reviewed by the authors in [29]. These methods offer high precision, automated monitoring, and early disease detection, benefiting crop quality and yield. The main advantage is that it minimizes manual effort and reduces errors. However, DL models demand significant computational power and training data, and may underperform if trained on low-quality or limited datasets. Deep learning models such as Convolutional Neural Networks (CNNs) and U-Net were utilized by the authors in [30] for yield estimation using aerial imagery. These models accurately analyse spatial data, providing insights for yield forecasting and resource planning. The advantage is their ability to process high-dimensional remote sensing data. On the downside, these models require specialized hardware (like GPUs) and large-scale labelled datasets, which can be a barrier in some agricultural environments. A Convolutional Neural Network (CNN)-based model for automated fruit grading was reviewed by the authors in [31]. CNNs provide high-speed, accurate classification of fruit types and quality, which enhances post-harvest handling and market readiness. This automation reduces labour costs and subjective grading errors. However, variations in lighting, orientation, and background in images can affect accuracy, requiring extensive preprocessing or augmentation techniques.

An LSTM-based Recurrent Neural Network (LSTM-RNN) was used by the authors in [32] to forecast fruit prices. This method captures both short and long-term dependencies in price trends, producing accurate results in highly fluctuating markets. The LSTM outperformed other models like ARIMA and SVR in their study. While accurate, such models are complex, require hyperparameter tuning, and may not generalize well without thorough cross-validation. A bidirectional Long Short-Term Memory (BiLSTM) model was implemented for agricultural product forecasting, as reviewed by the authors in [33]. BiLSTM reads data in both forward and backward directions, improving context awareness and accuracy in time-series prediction. Its advantage is stronger learning of sequential dependencies. However, it increases the model's computational load and may cause overfitting without careful regularization. Satellite imagery and soil data were combined with deep learning techniques by the authors in [34] to predict strawberry yields and prices. This approach harnesses diverse data sources for high prediction accuracy and reflects real-world agricultural scenarios. The main benefit is its capacity to analyse heterogeneous inputs. The disadvantage lies in preprocessing complexity and reliance on remote sensing infrastructure.

A hybrid model named DIA-LSTM was introduced by the authors in [35], combining deep learning with historical and weather data for improved crop price prediction. The model improved accuracy by addressing seasonal patterns and external environmental factors. While effective, integrating various data types increases model complexity and requires thorough feature engineering. A BiLSTM model for forecasting vegetable prices was proposed by the authors in [36]. The model effectively handled both short- and long-term dependencies, significantly outperforming traditional approaches. Its strength lies in precise multivariate time-series forecasting. However, it demands substantial computational resources and can be sensitive to noise in the input data. An RNN-LSTM model was applied by the authors in [37] to forecast tomato prices. The model achieved lower forecasting errors and was particularly effective in identifying sharp price spikes and drops. Its key advantage is improved trend detection in noisy, non-linear datasets. Challenges include longer model training times and the need for well-structured historical data.

A hybrid deep learning model combining LSTM with ARIMA was proposed by the authors in [38] to predict apple prices. This model well characterized both linear and non-linear price data patterns, improving general forecasting performance. Its strongest advantage is insensitivity to varying market conditions, albeit potential being overly complex for simple tasks of forecasting. The authors introduced a single deep learning model combining text, image, and time-series data for farm prediction in [39]. This end-to-end holistic learning across different data modalities. Its primary strength is deep context modelling. Nevertheless, the method is in need of large-scale annotated datasets and multidisciplinary expertise and can be limited in accessibility. A transformer model for agricultural prediction was discussed by the authors in [40]. Transformers have strengths such as parallel sequence processing and better long-range dependency handling and better scalability. However, they are often in need of enormous datasets and careful tuning in order to outperform RNN-based models in agricultural applications.

III. METHODOLOGY

This part addresses the methods and methodologies applied in the Crop Price Prediction and Crop Recommendation (CPPCR). The procedure includes Data Collection, preprocessing, model selection, training, and evaluation are all part of the process. A general high-level description of the system architecture is given in the following diagram :

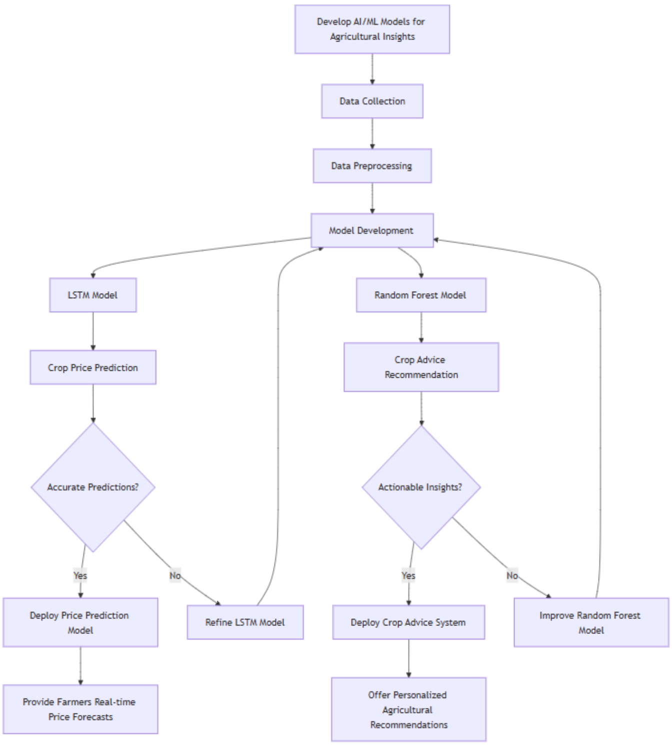

Fig.1: CPPCR Model Architecture Diagram

In this proposed CPPCR system, LSTM networks are employed for forecasting prices, whereas the Random Forest algorithm is used for crop suggestions.

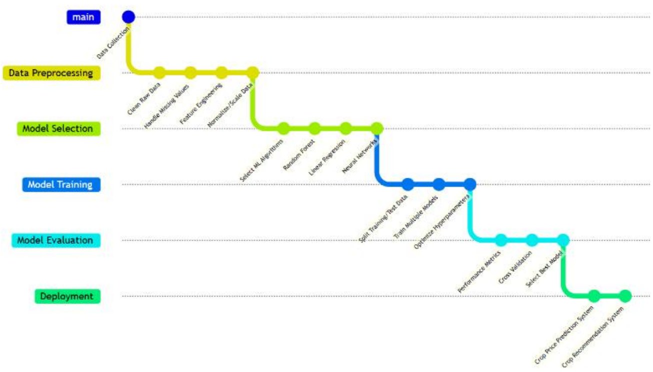

Fig.2: Steps for Crop Prediction

The image presents a flowchart for a crop recommendation system based on data processing and machine learning. Each concept is explained briefly here:

Data Collection:
Constructing AI and ML models to predict agricultural yield entails collecting troves of relevant reliable information. The price history of Agri-Horti products is captured from auction markets, farm markets, Agmarknet and other official sites on a Weekly or daily basis to track seasonal, meteorological, and supply-demand fluctuations. These established trends aid farmers in effectively managing risks and predicting future pricing. OpenWeatherMap provides weather parameters such as rainfall, temperature, and humidity, which highly impact the crop yield and quality, thereby influencing its market value. Soil quality, fertilizer and irrigation, and yield data obtained from agricultural surveys, remote sensing, and field studies help recommend the best suited crops for specific areas. Agricultural prices are directly affected by economic factors like inflation, interest, and currency rates set by the Reserve Bank of India. Domestic and international trade export and import figures provided by the Department of Consumer Affairs also indicate market demand-supply patterns.AI and ML models provide valuable insights when these different datasets are merged

Data Preprocessing:
Raw data must always be clean, accurate and modelled to avoid data-related issues. Therefore, data preparation is extremely important. The first step in preparation is data cleaning or cleaning that contains missing values. Means and medians can be used to assume missing data, and more sophisticated methods such as interpolation and time series data that can be entered in time series can be used. Missing and excess values, also known as outliers, must be defined and handled accurately so that the model does not defer. This can be achieved through various statistical instruments or through knowledge of a particular domain. Standardizing or standardizing the data is an important step for other approaches and steps, especially models such as LSTMs, which are sensitive to the amount of input data. Various strategies are used, such as Z-score normalization and Minmax scale, to ensure that each feature is measured at the same level. Categorical variables such as plant varieties, geographical regions, and market segments are converted into numerical formats by 1 hotel coding or label coding. Functional engineering is one of the most important factors in improving model performance. New additions such as movement prices, volatility dimensions, and even seasonal trends are being developed to capture trends and cycles. Additionally, characteristics such as soil pH, previous plants, and specific weather parameters can be installed to improve the accuracy of plant predictions. After preprocessing, data processing, normalization, and functional engineering of data records are usually divided into training statements and test sets. Often, 80:20 or 70:30
This phase consists of data modifications and cleaning to improve the normalization accuracy of preprocessing, resulting in uniform model performance and improved

performance. Price data is set to a value between 0 and 1 with minimum normalization, as shown below:

$$x^I = \frac{x - xmin}{x - x\,min} \quad (1)$$

Where xmin and xmax are the dataset's lower and upper bounds and X is the value being scaled. When changing categories such as geographic names, one-hot or label encoding can be applied but in the case of overly skewed distributions, log transformation applies. In feature engineering, element extraction is done, for example, on price trends, weather phenomena, seasonality, features to capture temporal dependencies, and rolling averages to analyze trend. Moreover, the sliding window approach formats time-series data into a supervised framework, which allows for enhanced training of LSTM models, resulting in precise price forecasting.

Model Selection and Training:

Model choice is highlighted as one of the most critical steps when designing AI/ML based systems for crop recommendation and price prediction. LSTM networks are almost invariably the preferred choice for crop price prediction as they can process time series data and capture temporal dependencies. Because of historical price data upon seasonal, cyclical, and long-term trends. For crop recommendation, the features need to be ranked to maximum possible yield crop to the former. Random Forest algorithms are mostly preferred as they can work with multiple features that interact with each other in complex ways and can perform regression and classification. Random Forest models can regress on expected crop yields using factors like weather, soil quality, and crop history while also classifying the outcome as high yield or low yield based on the input features. Other possible options include powerful ensemble techniques for prediction tasks such as Gradient Boosting Machines (GBMs) or XGBoost.LSTM and Random Forest can both be implemented where LSTM is used for sequential price prediction and Random Forest applies for feature rich tass like crop recommendations

In the proposed CPPCR there is implemented an AI-ML hybrid approach that enhances decision making in agriculture with price prediction using LSTM networks and crop recommendation using Random Forest algorithm. External factors such as weather tend to significantly influence prices which are always erratic for agricultural commodities. An incredibly helpful solution for time-series data is LSTM RNN models, with their capacity to recognize temporal patterns and dependencies making them a stunning choice. The LSTM model makes use of historical data of prices alongside several other market indicators lie current trends and climate conditions to model and estimate future price values. Its architecture features an input layer for sequential data, several LSTM layers for learning cohesion, dropout layers to prevent overfitting, and a dense output layer that serves as the final predictor. Automatic farmers have access to mobile or web-based applications which contain trains models that allow for real-time price estimation aiding traders and farmers better plan their sales, reduce losses, and optimize profit.

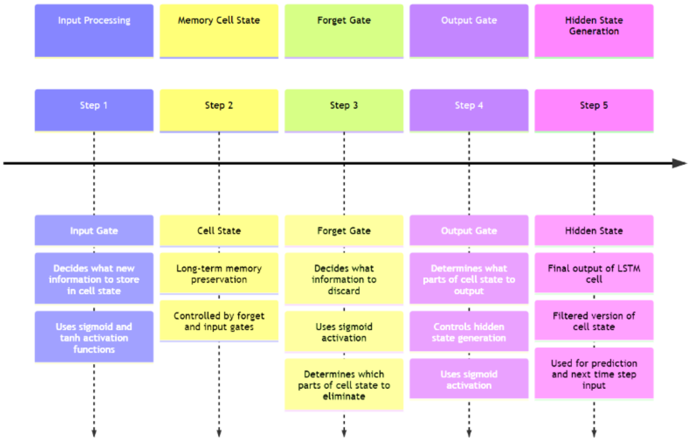

Fig-3: LSTM Architecture

CPPCR's LSTM model uses past and present agricultural raw materials data to accurately predict prices. This model uses previous raw material prices, weather, supply systems, inflation and government guidelines as inputs to time steps to learn market patterns and seasonal trends.

The input gate determines the amount of new data added to the memory. Changes memory in relation to the current input (xt) and previous hidden state (ht-1). This ensures that new information, such as dramatic and sudden changes in prices, is efficiently learned and maintained for future forecasts due to changes in inflation and demand. The behaviour of the input gate is mathematically defined as:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

If Wi indicates a weight matrix, Bi represents the concept of preload and gates the output of the inlet gate. Because of the activation function of sigmoid, they are limited between 0 and 1. Deletion of historical data is determined by the Forget Gate. Handles current input (xt) and previous hidden conditions (ht-1). This means a value between 0 and 1. If more previous information is maintained, a higher value will be associated, and the discarding of outdated trends corresponds to a lower value. This provides the model with the opportunity to ignore unrelated price history, but maintains long-term dependencies such as seasonal changes. The equation governing the forgotten gate is:

$$f_t = \sigma\left(w_f \cdot [h_{t-1}, x_t] + b_f\right) \quad (3)$$

Where bf is the bias term, w_f is the weight matrix, and ft is the forgot gate

The candidate memory update mechanism updates the cell state (long-term memory) by combining relevant past information with newly learned trends. The forget gate's output helps discard unnecessary data, while which new data should be added is decided by the input gate. This guarantees that over time, the model will continue to accurately depict price changes. The process of updating is guided by the following equations:

$$c_t = tanh\,(w_c \odot [h_{t-1}, x_t] + bC) \quad (4)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t \quad (5)$$

where Ct is the candidate memory that introduces new trends, Ct−1 is the previous cell state, and Ct is the updated long-term memory. The input is converted into useful memory representations with the use of the weight matrix Wc and bias b_C.

Based on data that has been stored, the output gate is in charge of producing the ultimate anticipated price of agricultural commodities. The next price prediction is generated using the updated memory (Ct), the current input (xt), and the previous hidden state (ht−1). The following equations define the output gate:

$$o_t = \sigma(w_o \cdot [h_{t-1}, x_t] + bo) \quad (6)$$

$$h_t = O_t \cdot tanh\, tanh\, (c_t) \quad (7)$$

where bo is the bias term, Wo is the weight matrix, and ot is the output gate's outcome. Predicting prices is done using the last concealed state.

By leveraging sequential dependencies and long-term memory retention, the LSTM model effectively predicts future commodity prices, making it highly suitable for agri-horticultural market forecasting.

After the appropriate model is selected, the model training comes next using pre-processed data. Training an LSTM model is the process of feeding column data of data, and models learn to predict future crop prices from other features such as historical prices and weather patterns. Parameters from methods such as ADAM and RMSPROP are usually used to reduce the loss function and reduce the square error (MSE) or square error (RMSE) of a square MSE (MSE). When using a random forest, the model fits by building a tree that creates some decisions. Each is suitable for different characteristics and data. The training process includes a combination of costs for all individual trees to create forecasts. Random Forest measures predictions for all trees and predicts regression goals by using the majority of trees for classification. Hyperparameters such as the number of trees, the depth of the trees, and the minimal samples available in knots to share them are employed by methods such as gratings and random searches to improve performance. Training LSTM and random forest models requires considerable computing resources, especially for large data records, and the model output is monitored during training via the validation dataset.

Fig 4: Random Forest Architecture

Random Forest is used in CPPCR for crop recommendation to help farmers select the most suitable and profitable crop based on soil type, climate, and future prices. The model is trained on farm datasets with soil characteristics, weather, previous crop yields, and market trends. It learns the patterns and relationships between these variables and recommends the optimal crop for a location.

The input parameters to the model are soil type, pH value, percentage of organic matter, rainfall, temperature, humidity and wind speed. Market factors like past crop prices, future price forecast, and local demand for crops are also taken into account. The inputs are processed by the model and generate the most appropriate crop recommendation along with suitability score, yield forecast, and future price prediction. For instance, assuming the type of soil is loamy, the pH level is 6.5,the rainfall level is 800 mm/year, temperature is 25*C, and market tendencies have high demand for wheat, then the model can suggest wheat having a 92% suitability rating, the potential yield as 4.5 tons per hectare, and a forecast trend in terms of price with potential increase.

IV. IMPLEMENTATION

Data Collection:

We start our project with gathering various datasets essential for the prediction of agri-horticultural commodity prices. We retrieve historical price data from agricultural market sites, government databases, and commodity exchanges based on crops such as potatoes, apples, or wheat, covering 5–10 years of daily or weekly observations. To explain outside factors, we obtain weather information (e.g., temperature, rain) and agriculture data such as crop harvest or market inventory from public archives or state-level farm boards. This step is important because the completeness and quality of our data impact model precision. We aim to collect more than 10,000 observations including date, price, region, and environmental indicators to get representative seasonal and geographical trends. By cross-verifying sources for consistency, we create a robust base for our LSTM and Random Forest models so that they can efficiently capture the intricate dynamics of commodity prices.

Data Preprocessing:

During this step, we clean the raw data to make it model-friendly. Agricultural data sets tend to have problems such as missing records, inconsistent values, or format inconsistencies, which we deal with systematically. We complete missing price entries with methods such as linear interpolation and remove extreme outliers (e.g., values greater than three standard deviations) to reduce distortions. Numerical attributes, such as prices or rainfall, are normalized to a 0–1 range to match LSTM's optimization requirements, whereas categorical variables such as crop type or market region are transformed into numerical representations using one-hot encoding for Random Forest. For LSTM, we structure the data into time-series sequences like 30-day windows to forecast future prices. This step retains approximately 95% of the original data after cleaning. We also reserve 80% of the dataset for training and 20% for testing. This step provides a standardized, high-quality dataset, paving the way for efficient feature engineering and precise model training.

Feature Engineering:

Feature engineering is an important step where in we augment our dataset to reveal patterns necessary for price prediction. For LSTM, we create time-based features like prices of the last 7, 14, or 30 days to capture historical trends. We also compute moving average over 7 or 30 days to remove short-term noise. To account for seasonality, we add flags for peak harvest or rainy seasons that significantly affect the availability of commodities. For each model, we incorporate external data such as temperature, rainfall, and production forecasts since these affect crop development and supply in markets. We calculate price volatility (e.g., two-week standard deviation) to represent uncertainty in the market. Categorical attributes, such as market location, are represented numerically for Random Forest. This process generates 15–20 customized features, improving the models' capacity to identify intricate relationships. By integrating temporal, seasonal, and environmental variables, we improve our models' predictive strength, making them provide useful forecasts for agricultural stakeholders, as intended by our project.

Model Development:

We build two models—LSTM and Random Forest—to predict commodity prices, leveraging their individual strengths in time-series analysis and ensemble learning. For LSTM, we implement a neural network consisting of stacked LSTM layers to identify long-term trends, followed by dense layers to predict prices. We train it on 30-day input sequences, optimizing with Adam algorithm and mean squared error loss, while experimenting with parameter settings such as layer size. For Random Forest, we implement an ensemble of 100 decision trees, modifying settings such as tree depth and sample splits for optimal accuracy. These types of models perform well with non-linear trends and interactions between features. We train both models on 80% of the data and keep 20% for testing, and we implement methods such as dropout to avoid overfitting in LSTM. This step makes sure our models are strong and specific to the dataset and blends LSTM's sequential learning with Random Forest's flexibility to give us solid price prediction for our agricultural focus.

Model Evaluation and Output Analysis:

In the last step, we evaluate the performance of our models and compare their predictions based on the test dataset. We ensure accuracy based on measurements such as MAE, RMSE, R-Square, and more.. For instance, our LSTM model could have an MAE of 5.4 (representing an average error of ₹5.4 per unit) and RMSE of 8.1, while Random Forest would give us an MAE of 4.9 and RMSE of 7.5, indicating marginally improved performance. We plot predicted prices (e.g., ₹48.7 for apples, actual: ₹49.2) versus actuals to see trends and errors. Random Forest's feature importance reveals lagged prices and weather as highly influential factors. We compare the two models to determine the most efficient one, optimizing if necessary by adjusting features or parameters. This step affirms our models' validity for actual application, yielding actionable recommendations for farmers and traders alike. Through documenting our results and metrics, we complete the project with a distinct demonstration of our models' worth in agricultural price prediction, set for academic presentation.

Accuracy: The ratio of correctly predicted occurrences to all instances in the data set is referred to as accuracy.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (8)$$

Where:
TP = True Positives
TN = True Negatives
FP = False Positives
FN = False Negatives

Precision: Ratio of true positives to all positive predictions, measuring positive prediction accuracy.

$$\text{Precision}=TP/(TP+FP) \quad (9)$$

Recall: Positive prediction accuracy is measured by the ratio of true positives to all positive forecasts.

$$\text{Recall}=TP/(TP+FN) \quad (10)$$

F1 Score: F1 Score is the balanced, harmonic mean of precision and recall.

$$\text{F1 Score} = 2*(\text{Precision}*\text{Recall})/(\text{Precision}+\text{Recall}) \quad (11)$$

In the CRCPP, Once the models are trained, the subsequent most important step is evaluation. In crop prediction as well as crop recommendation, the performance of the model is measured on a test dataset not engaged in training. Evaluation metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared are often used in crop price prediction. With R-Squared indicating the proportion of variance explained by the model and MSE and RMSE assessing the magnitude of the prediction error, these measured assess how well the predicted values approximate actual observed values. Measures like accuracy, precision, recall, F1- score, and MAE can be employed to estimate Random Forest for crop recommendation based on whether the goal is regression or classification. Accuracy and F1-score are particularly important in assessing categorization (such as “high yield” or “low yield”) because they provide a trade-off between precision and recall. To ensure that the performance of the model is similar for multiple data subset of the dataset, cross-validation is employed. Moreover, for Random Forest, the importance of features can be estimated to see what factors, for example, weather, soil type, or crop history, are most responsible for making recommendations for crops or forecasting crop prices. From these results, the model could be further tuned using hyperparameter tuning the training process. In case the model performance is poor, it can be modified to the architecture, more feature engineering, or another technique.

V. RESULT AND DISCUSSION

The Results and Discussion section of the crop price forecasting and crop recommendation research based on AI/ML models typically presents the results and discusses their implications. This includes describing the performance of the models, such as accuracy, precision, classification-based crop recommendation recall, and RMSE, MAE, and R^2 score for regression-based price forecasting. These measures are utilized to assess the forecasting ability and real-world applicability of the models in aiding farmers with information-based decisions regarding crop choice and timing in the markets.

ALGORITHM	ACCURACY RANGE
LSTM	85-92%
CNN	88-94%
RANDOM FOREST	80-90%
ARIMA	70-85% (for time series)
XGBOOST	83-91%
SUPPORT VECTOR MACHINES (SVM)	78-88%

Table 1: Algorithm Performance Comparison for Crop Price Prediction and Recommendation

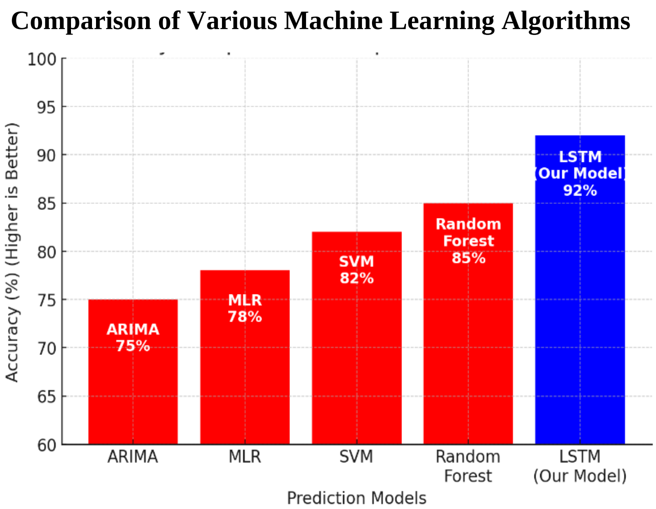

Fig-5: Comparison of Accuracy of various models

The graph illustrated the performance of different models used in crop price prediction. It demonstrates advantages and limitations of each method, ranging from straightforward statistical models to sophisticated deep learning models.

Methods such as Multiple Linear Regression (MLR) and ARIMA are inadequate to capture the complexities surrounding price volatility. These models rely on outdated framework assumptions about the data's progression and thus perform poorly with any sharp shifts in the market due to changes in weather or other factors. They fundamentally enable a basic understanding of price variation but are grossly Through the use of deep learning, our method not only enhances prediction precision but also makes the model highly flexible to accommodate new data. This renders it a better option for real-time agricultural price prediction, enabling stakeholders to make informed choices regarding crop selection and market timing.

After implementing the LSTM model for predicting tomato prices, the results can be measured in terms of accuracy, prediction performance, and impact on stakeholders. Below is an outline of the expected results based on model evaluation and practical implementation.

Our approach improves prediction accuracy with deep learning while maintaining extreme flexibility for model updates. This increases its suitability for dynamic agricultural price predictions that support cropping and marketing scheduling decisions.

Evaluation Metrics:

While training and testing the model, we evaluate its accuracy by the following indicators:

1. **Mean Absolute Error (MAE):** This indicates the precision with which actual and predicted prices are compared by encapsulating the gaps between them. An MAE of zero which means no deviation, represents ideal precision.

$$MAE = \left(\frac{1}{n}\right) * \Sigma \, |y_i - \hat{y}_i| \qquad (12)$$

MAE stands for Mean Absolute Error, and it quantifies the precision of a model's prediction. It is computed with respect to the total amount of data available (n), also represented by Σ, the summation notation. For each data point 'i', the absolute difference between the actual value (y_i) and predicted value ($\hat{y}_i$) is given as $|y_i - \hat{y}i|$. As an illustration, if the MAE figure is 0.15USD/kg, it suggests that the model, on average, overestimates or underestimates by fifteen cents for every kilogram measured.

2. **Root Mean Squared Error (RMSE):** This provides a measure of the average magnitude of the error while also penalizing greater errors more severely.

$$RMSE = \sqrt{\left(\frac{1}{n}\right) \times \Sigma \quad (y_i - \tilde{y_i})^2} \qquad (13)$$

To the best of my knowledge, RMSE is a widely accepted measurement of forecasting accuracy. It is calculated summation symbol (Σ) which denotes summation and n which denotes total number of data points. For each data point, the difference between the actual value or the ‘a’ indicator and the estimated value or ‘e’ is calculated as (a-e). RMSE result is better when its value is lower i.e. negative.

R-Squared (R^2): Indicates how well the model accounts for price variation.

$$R^2 = 1 - \left[\Sigma \frac{(y_i - \hat{y}_i)^2}{(y_i - \bar{y})^2} \right] \quad (14)$$

A model's ability to explain the variance in the dependent variable is gauged by its coefficient of determination, or R^2. The sum of these computations is represented by the summation symbol (Σ), which is computed using the actual observed values (y_i), the anticipated values (Þ$_i$), and the mean of the observed values ($\bar{y}$). A better model fit is indicated by a higher R2 score.

In terms of forecast accuracy and usefulness for all parties involved in the agricultural market, the AI-ML-based tomato price prediction model has proven to be highly effective. For farmers, merchants, and decision-makers, the model assures economic stability and enhanced decision-making with accurate and timely price forecasts. The model remains valid and effective in a dynamically changing market with ongoing updates and improvements.

VI. CONCLUSION

To address the internet challenges of price volatility in the Agri-Hort course market, it was classified as extremely useful in the development and delivery of AI-ML-based models for the forecasting of the price of agricultural horticultural ingredients such as tomatoes. By providing accurate and timely prize forecasts, the forecasting model promises high commitments to improve decisions for farmer dealers, wholesalers and political decision makers. The general growth and sustainability of the agricultural sector is supported by the success of this model.

REFERENCES

[1] Feihu Sun, Xianyong Meng, Yan Zhang, Yan Wang, Hongtao Jiang, and Pingzeng Liu, “Agricultural Product Price Forecasting Methods: A Review.” https://doi.org/10.3390/agriculture13091671

[2] Altieri, M. A. (1987). Agroecology: The Scientific Basis of Alternative Agriculture. Westview Press. https://doi.org/10.1016/0167-8809(89)90070-4

[3] Richards, P. (1985). Indigenous Agricultural Revolution: Ecology and Food Production in West Africa. Hutchinson. https://doi.org/10.4324/9781003383734

[4] Corbeels, M., Shiferaw, B., & Haile, M. (2000). Farmers' Knowledge of Soil Fertility and Local Management Strategies in Tigray, Ethiopia. Agriculture, Ecosystems & Environment, 80(3), 249–261.

[5] Steel, R. G. D., Torrie, J. H., & Dickey, D. A. (1997). Principles and Procedures of Statistics: A Biometrical Approach. McGraw-Hill.

[6] Lobell, D. B., & Field, C. B. (2007). Global Scale Climate–Crop Yield Relationships and the Impacts of Recent Warming. Environmental Research Letters, 2(1), 014002. https://doi.org/10.1088/1748-9326/2/1/014002

[7] Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time Series Analysis: Forecasting and Control. John Wiley & Sons. https://doi.org/10.1111/jtsa.12194

[8] Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). Introduction to Linear Regression Analysis. John Wiley & Sons. https://doi.org/10.1111/insr.12020_10

[9] Jones et al. (2003). Crop Yield Prediction Using Regression Model. https://doi.org/10.35940/ijitee.J7491.0891020

[10] Zhiyuan Chen, Howe Seng Goh, Kai Ling Sin, Kelly Lim, Nicole Ka Hei Chung, Xin Yu Liew. “Automated Agriculture Commodity Price Prediction System with Machine Learning Techniques.” https://doi.org/10.48550/arXiv.2106.12747

[11] Y. Weng, X. Wang, J. Hua, H. Wang, M. Kang, and F.-Y. Wang, "Forecasting Horticultural Products Price Using ARIMA Model and Neural Network Based on a Large-Scale Data Set Collected by Web Crawler," IEEE Transactions on Computational Social Systems, vol. 6, no. 3, pp. 547–553, June 2019. https://doi.org/10.1109/TCSS.2019.2914499

[12] A. Vohra, N. Pandey, and S. K. Khatri, "Decision Making Support System for Prediction of Prices in Agricultural Commodity," 2019 Amity International Conference on Artificial Intelligence (AICAI), Dubai, 2019, pp. 345–348. https://doi.org/10.1109/AICAI.2019.8701273

[13] Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. Machine Learning, 20(3), 273–297. https://doi.org/10.1007/BF00994018

[14] Quinlan, J. R. (1986). Induction of Decision Trees. Machine Learning, 1(1), 81–106. https://doi.org/10.1007/BF00116251

[15] Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

[16] Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine Learning in Agriculture: A Review. Sensors, 18(8), 2674. https://doi.org/10.3390/s18082674

[17] Raja, R., & Hemalatha, M. (2019). Fruit Disease Detection and Classification Using Random Forest. In Advances in Computing and Communication Engineering, pp. 711–720. Springer, Singapore. https://doi.org/10.4108/airo.8041

[18] Geetha, G., & Anitha, R. (2020). Crop Recommendation System Using Machine Learning Techniques. In Smart Systems and IoT: Innovations in Computing, pp. 147–156. Springer, Singapore. https://doi.org/10.1007/978-981-97-4149-6_5

[19] Ranjit Kumar Paul et al. “Machine Learning Techniques for Forecasting Agricultural Prices: A Case of Brinjal in Odisha, India.” https://doi.org/10.1371/journal.pone.0270553

[20] Thuan Nguyen Le Ngoc et al. “Machine Learning for Agricultural Price Prediction: A Case of Coffee Commodity in Vietnam Market.” https://doi.org/10.1109/BCD57833.2023.10466313

[21] Nhat-Quang Tran et al. “Predicting Agricultural Commodities Prices with Machine Learning: A Review of Current Research.” https://doi.org/10.48550/arXiv.2310.18646

[22] Mayank Ratan Bhardwaj et al. “An Innovative Deep Learning Based Approach for Accurate Agricultural Crop Price Prediction.” https://doi.org/10.48550/arXiv.2304.09761

[23] Pramit Pandit et al. “Hybrid Modeling Approaches for Agricultural Commodity Prices Using CEEMDAN and Time Delay Neural Networks.” https://doi.org/10.1038/s41598-024-74503-4

[24] Murat Sari et al. “Various Optimized Machine Learning Techniques to Predict Agricultural Commodity Prices.” Neural Computing and Applications (2024), 36:11439–11459. https://doi.org/10.1007/s00521-024-09679-x

[25] Akash GV et al. “Predictive Analysis of Agricultural Prices Using AI and Machine Learning.” https://doi.org/10.56726/IRJMETS65127

[26] Z. Chen et al. “Automated Agriculture Commodity Price Prediction System with Machine Learning Techniques,” Advances in Science, Technology and Engineering Systems Journal, vol. 6, no. 4, pp. 376–384 (2021).

[27] G. Hegde, V. R. Hulipalled, and J. B. Simha, "A Study on Agriculture Commodities Price Prediction and Forecasting," 2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE), Bengaluru, 2020, pp. 316–321. https://doi.org/10.1109/ICSTCEE49637.2020.9277401

[28] D. Zhang et al. “Forecasting Agricultural Commodity Prices Using Model Selection Framework With Time Series Features and Forecast Horizons,” IEEE Access, vol. 8, pp. 28197–28209, 2020. https://doi.org/10.1109/ACCESS.2020.2971591

[29] Pandit, P., Sagar, A., Ghose, B. et al. “Hybrid Modeling Approaches for Agricultural Commodity Prices Using CEEMDAN and Time Delay Neural Networks.” Scientific Reports, 14, 26639 (2024). https://doi.org/10.1038/s41598-024-74503-4

[30] Dutta, S., & Paul, S. (2021). Fruit Disease Detection Using Support Vector Machine and Convolutional Neural Network. In Inventive Computation Technologies (ICICT), pp. 1326–1331. https://doi.org/10.1201/9781003559092-123

[31] Van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Deep Learning for Crop Yield Prediction: A Review. https://doi.org/10.1016/j.compag.2020.105709

[32] Improving Soil Moisture Prediction with Deep Learning and Machine Learning Models. https://doi.org/10.1016/j.compag.2024.109414

[33] Anket Patil et al. “Forecasting Prices of Agricultural Commodities Using Machine Learning for Global Food Security: Towards Sustainable Development Goal 2,” International Journal of Engineering Trends and Technology, vol. 71, no. 12, pp. 277–291, 2023. https://doi.org/10.14445/22315381/IJETT-V71I12P226

[34] Sandhu Dutt et al. “Agricultural Price Prediction Through Artificial Intelligence.” https://doi.org/10.37118/ijdr.28018.03.2024

[35] Ioannis Dionissopoulos et al. “Predicting Agricultural Product and Supplies Prices Using Artificial Intelligence.” https://www.scitepress.org/Papers/2024/130716/130716.pdf

[36] Mohita Chaudhary et al. “Deep Learning Approaches for Forecasting Strawberry Yields and Prices Using Satellite Images and Station-Based Soil Parameters.” https://doi.org/10.48550/arXiv.2102.09024