

A REPORT
ON

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

Submitted by,

DARSHAN V	20211CST0081
BYRESH SORADI	20211CST0057
TEJAS D	20211CST0105
DEVA PRAKASH	20211CST0053

Under the guidance of,

Dr. MADHUSUDHAN M.V

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

**COMPUTER SCIENCE AND TECHNOLOGY
(ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING)**

At



PRESIDENCY UNIVERSITY

BENGALURU

MAY 2025

A REPORT
ON

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

Submitted by,

DARSHAN V	20211CST0081
BYRESH SORADI	20211CST0057
TEJAS D	20211CST0105
DEVA PRAKASH	20211CST0053

Under the guidance of,

Dr. MADHUSUDHAN M.V

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

**COMPUTER SCIENCE AND TECHNOLOGY
(ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING)**

At



PRESIDENCY UNIVERSITY

BENGALURU

MAY 2025

PRESIDENCY UNIVERSITY

PRESIDENCY SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that the Project report on “**Development of AI-ML Based Models for Predicting Prices Of Agri-Horticultural Commodities**” being submitted by “**DARSHAN V, BYRESH SORADI, TEJAS D, DEVA PRAKASH**” bearing roll number(s) “**20211CST0081, 20211CST0057, 20211CST0105, 20211CST0053**” in partial fulfillment of the requirement for the award of the degree of Bachelor of Technology in Computer Science and Engineering is a bonafide work carried out under my supervision.

MADHUSUDHAN M.V
Dr. MADHUSUDHAN M.V
Associate Professor
School of CSE&IS
Presidency University

A.S.A
Dr. SAIRA BANU ATHAM
Professor & HOD
School of CSE&IS
Presidency University

M.N.A
Dr. MYDHILI NAIR
Associate Dean
PSCS
Presidency University

S.K.
Dr. SAMEERUDDIN KHAN
Pro-Vice Chancellor -
Engineering
Dean –PSCS / PSIS
Presidency University

PRESIDENCY UNIVERSITY

PRESIDENCY SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

DECLARATION

We hereby declare that the work, which is being presented in the project report entitled Development of AI-ML Based Models for Predicting Prices Of Agri-Horticultural Commodities in partial fulfillment for the award of Degree of Bachelor of Technology in Computer Science and Engineering, is a record of our own investigations carried under the guidance of Dr. MADHUSUDHAN M.V, Associate Professor, School of Computer Science Engineering & Information Science, Presidency University, Bengaluru.

We have not submitted the matter presented in this report anywhere for the award of any other Degree.

NAME	Roll No	Signature
Darshan V	20211CST0081	
Byresh Soradi	20211CST0057	
Tejas D	20211CST0105	
Deva Prakash	20211CST0053	

ABSTRACT

Forestry is not only important in ensuring ecological security but also plays a vital role in contributing to economic stability. As environmental and agricultural issues become increasingly complex, there has been a significant surge in the implementation of sophisticated intelligence technologies, especially machine learning (ML) and artificial intelligence (AI), throughout the agriculture industry. In the recent years, such technologies have acted as precious aids, particularly for correct forecasting of farm produce prices. Correct forecast is of the most crucial nature to different actors—farmers, traders, policymakers, and governmental authorities—because it helps them make better, timely, and strategic production, storage, distribution, and market participation decisions.

The paper is concerned with the real-world application and performance of AI and ML techniques in solving the chronic and new-age problems in farm price forecasting. These advanced models are capable of processing a massive amount of data, such as past pricing patterns, weather conditions, soil fertility, market demands, and economic factors. Through the use of such multi-faceted data, these AI-based systems are able to predict future prices with an astonishing level of accuracy. Consequently, farmers are more able to organize which crops to plant, when to harvest, and how to engage the market, resulting in higher productivity, less economic risk, and more overall efficiency in farm operations.

Besides enabling individual-level decision-making, precise price predictions also have significant macroeconomic influence. It will be able to moderate the volatility of sudden movements in the marketplace or surprise prices, thus resulting in higher price stability, a smoother supply chain, and a more secure food supply. These insights can benefit policymakers as well, allowing them to create smarter, data-enabled agricultural and trade policies that benefit both market inclinations and sustainable objectives.

The findings of this study are likely to yield significant contributions toward realizing how AI and ML technologies can be systemically applied in order to more effectively predict farm prices. Eventually, this can open the doors to more resilient farming practices and policy structures supporting both economic growth and food supplies on a wider level.

ACKNOWLEDGEMENTS

First of all, we are indebted to the **GOD ALMIGHTY** for giving me an opportunity to excel in our efforts to complete this project on time.

We express our sincere thanks to our respected dean **Dr. Md. Sameeruddin Khan**, Pro-VC - Engineering and Dean, Presidency School of Computer Science and Engineering & Presidency School of Information Science, Presidency University for getting us permission to undergo the project.

We express our heartfelt gratitude to our beloved Associate Dean **Dr. Mydhili Nair**, Presidency School of Computer Science and Engineering, Presidency University, and **Dr. Saira Banu Atham**, Head of the Department, Presidency School of Computer Science and Engineering, Presidency University, for rendering timely help in completing this project successfully.

We are greatly indebted to our guide **Dr. Madhusudhan M.V, Associate Professor CSE**, and Reviewer **Ms. Manjula**, Presidency School of Computer Science and Engineering, Presidency University for his inspirational guidance, and valuable suggestions and for providing us a chance to express our technical capabilities in every respect for the completion of the internship work.

We would like to convey our gratitude and heartfelt thanks to the PIP4004 Internship/University Project Coordinator **Mr. Md Ziaur Rahman and Dr. Sampath A K**, department Project Coordinators **Ms. Manjula** and Git hub coordinator **Mr. Muthuraj**.

We thank our family and friends for the strong support and inspiration they have provided us in bringing out this project.

Darshan V

Byresh Soradi

Tejas D

Deva Prakash

LIST OF TABLES

Sl. No.	Table Name	Table Caption	Page No.
1	Table 1	Comparison of Accuracy of various models	49

LIST OF FIGURES

Sl. No.	Figure Name	Caption	Page No.
1	Figure 1	CPPCR Model Architecture Diagram	5
2	Figure 2	Steps for Crop Prediction	21
3	Figure 3	Long Short Term Memory Architecture Diagram	25
4	Figure 4	Random Forest Architecture Diagram	31
5	Figure 5	Comparison of Accuracy of various models	49
6	Screenshot 1	Agri Price Prediction web Page	59
7	Screenshot 2	List of Vegetables	59
8	Screenshot 3	Crop Recommendation based on manually entered data	60

TABLE OF CONTENTS

CHAPTER NO.	TITLE	PAGE NO.
	ABSTRACT	iv
	ACKNOWLEDGMENT	v
	LIST OF TABLE	vi
	LIST OF FIGURES	vii
1	INTRODUCTION	1-8
	1.1 General Definition	1
	1.2 Motivation	2-3
	1.3 Problem Statement	3-8
2	LITERATURE SURVEY	9-16
	2.1 Literature Survey	
3	RESEARCH GAPS OF EXISTING METHODS	
	3.1 Key Research Gaps	17-20
4	PROPOSED METHODOLOGY	21-35
	4.1 Methodology	21-22
	4.2 Implementation Framework and Model Design	22
	4.3 LSTM for Sequential Data Forecasting	23
	4.4 Weaknesses of RNN and Strengths of LSTM	23
	4.5 Random Forest for Structured Feature Learning	24
	4.6 LSTM Architecture and Working	25-30
	4.7 Random Forest Architecture and Working	31-35
5	OBJECTIVES	36-37
	5.1 Objectives	
6	SYSTEM DESIGN AND IMPLEMENTATION	38-45
	6.1 Introduction	38
	6.2 System Architecture Overview	38-40
	6.3 Data Preprocessing and Collection	40-42
	6.4 Model Development	42-43
	6.5 Model Evaluation	43-44
	6.6 Deployment and user Interface	44-45
7	RESULTS AND DISCUSSIONS	46-51
	7.1 Model Performance Evaluation	46-47
	7.2 Comparison with Alternatives Models	47-48
	7.3 Real World Uses and Consequences	48-51
	CONCLUSION	52-54
	REFERENCES	55-58

Chapter 1

INTRODUCTION

The creation of AI-ML-based models for price forecasting of agri-horticultural commodities is a major leap forward in agricultural market prediction. Using cutting-edge methods in machine learning, such as Long Short-Term Memory (LSTM) networks and deployment frameworks such as Django, these systems are capable of processing past price data and determining trends to provide precise, real-time predictions. The combination of natural language processing and deep learning enables the models to dynamically adapt and scale based on regional or crop-specific needs. With price volatility still being a primary concern for farmers and traders, a strong predictive model improves decision-making, maintains economic efficiency, and assists stakeholders with timely information for better planning and profitability.

1.1 GENERAL DEFINITION

The creation of machine learning-based AI models for predicting agri-horticultural commodity prices is a new use of machine learning, which enables stakeholders to make smart decisions using past and real-time data. Advanced algorithms like LSTM networks and Random Forest are used by these models to examine price trends and make accurate predictions for future trends.

Price forecasting systems are becoming critical to contemporary agriculture, where farmers, traders, and policymakers need timely and accurate market information. Research indicates that the use of machine learning in agriculture can result in better planning and lower economic losses. These systems minimize dependence on manual analysis, which is usually time-consuming, prone to errors, and not scalable across commodities and regions.

As India's digital infrastructure is growing rapidly, the agriculture industry is now adopting technology. Proper price forecasting can enable farmers to choose where and when to sell their crops, minimize their reliance on intermediaries, and get improved returns. Conventional methods of estimating prices tend not to reflect the seasonality and complexity in market patterns, while AI-ML-based models provide quicker, more dynamic, and economical

solutions.

Platforms such as Django allow easy deployment and integration of such predictive systems into web and mobile applications. This makes the models available and scalable, serving diverse users ranging from remote farmers to market analysts. Utilizing LSTM's ability to handle sequential data and Random Forest's stability in classification and regression tasks, the system developed provides consistent and personalized price predictions, leading to economic stability and improved resource utilization in agri-horticulture.

1.2 MOTIVATION

The key reason for creating AI-ML-driven models for agri-horticultural commodity price forecasting stems from the necessity of providing farmers, traders, and policymakers with accurate and timely market intelligence. Without a dependable forecasting mechanism, stakeholders tend to remain uncertain and suffer losses due to price shocks that occur unexpectedly. These issues are most acute for perishable commodities, where incorrect or delayed decisions result in spoilage, wastage, and poor returns.

Price fluctuations in the agricultural produce market are based on several dynamic variables like seasonality, climatic aberrations, regional shortages/surpluses, and logistic limitations. Classical forecasting tools, although informative, lack the ability to process huge amounts of multidimensional, time-series-based data. Hence, they tend to provide vague predictions or relevant information. Such a lacuna underlines the necessity for clever models that have the ability to learn from the past and make sense of rapidly evolving market parameters.

Practically, machine learning algorithms provide the capability to perform sophisticated analysis automatically and generate data-driven, real-time price projections. A farmer, for example, may utilize such an algorithm to make decisions on selling produce now or holding off to receive a better price later. Government institutions and traders also utilize these estimates for strategic procurement planning, inventory control, and planning for stabilizing markets of the machine learning methods, Long Short-Term Memory (LSTM) networks best accommodate time-series data. LSTM models can both short-term and long-term fluctuation in price patterns and long-term dependencies in price movements, hence serving the purpose

of modeling temporal commodity price behavior well. Random Forest is a robust ensemble-based regression method that deals with high-dimensional data and detects intricate nonlinear relations between location, crop type, time of year, rainfall, and temperature features.

These models, upon being trained with large historical datasets, can spot patterns, minimize forecasting errors, and improve incrementally through retraining. Over time, they can detect repeated trends, forecast market movements, and offer anticipatory advice to stakeholders.

Aside from single use cases, such models provide scalability and efficiency to agricultural supply chains and policy schemes. For instance, they can simultaneously process several commodities, provide support price forecasting by market, and even be provided with multilingual data inputs or local weather and satellite data inputs for increased accuracy.

Platforms such as Django can allow these models to be hosted as interactive web applications, putting the technology into the hands of a wider populace, including farmers in rural setups. In conclusion, the use of AI-ML-based price forecasting systems through LSTM and Random Forest is a futuristic method to enhance transparency, efficiency, and decision-making in agri-horticultural markets. These systems give stakeholders the insights that were unavailable or hard to interpret earlier, which plays an important role in the aim of agricultural modernization and rural economic sustainability.

1.3 PROBLEM STATEMENT

Farmers and stakeholders in agri-horticulture are mostly exposed to tough decision-making against making timely and well-informed decisions because of erratic price trends, absence of robust forecasting platforms, and few resources for gaining real-time insights from the marketplace. Such decision-making difficulties tend to lead to suboptimal planning, compromised revenue, and escalated post-harvest wastage. To solve these problems, we suggest an automated system called "Development of AI-ML Based Models for Predicting Prices of Agri-Horticultural Commodities." The system employs sophisticated machine learning methods to analyze historical price patterns and external conditions in order to generate accurate and useful price projections.

The suggested system utilizes Long Short-Term Memory (LSTM) networks for modeling time-series data so that the model can capture long-term dependencies in price movements. LSTM suits the task specifically well in relation to capturing patterns of season and market cycles and hence is an effective option when it comes to predicting agricultural prices. Random Forest algorithms are used further to make predictions about how different features including crop type, location, climatic conditions, and past history of prices may impact the forecast, thus effectively enhancing the entire model's predictive accuracy using ensemble-based learning.

The backend system was implemented with the Django web framework, providing a stable and scalable system for hosting the machine learning models. Django provided for integration with data pipelines and enabled us to create a web interface that provides real-time forecasts and visualizations to end users such as farmers, traders, and agricultural planners.

This project not only improved our technical skills in machine learning and web development but also gave us good hands-on experience in solving real-world problems in the agricultural sector. With the deployment of this predictive system, we showed how AI-ML technologies can be effectively utilized to assist price stability, enhance market transparency, and enable farmers with the information required to make informed marketing decisions.

ARCHITECTURE DIAGRAM

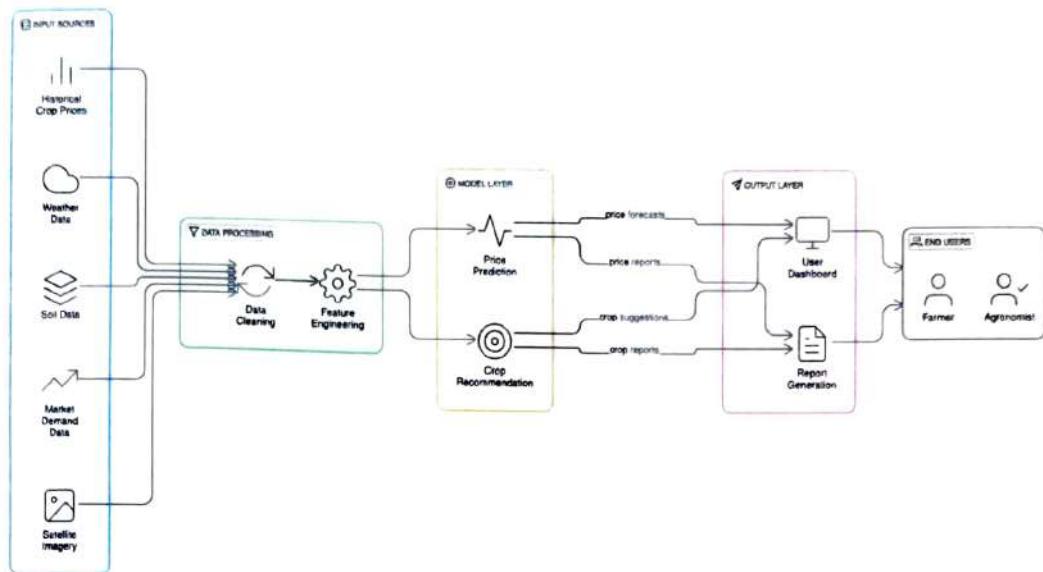


Figure 1: CPPCR Model Architecture Diagram

The Figure 1 Represents a top-level architecture of an AI-ML-driven crop recommendation and price prediction decision support system for agriculture. It has four major levels: Input Sources, Data Processing, Model Layer, and Output Layer, which culminate into actionable insights for end-users like farmers and agronomists.

The Input Sources comprise several data streams—Historical Crop Prices, Weather, Soil, Market Demand, and Satellite Imagery. These various datasets are important in capturing environmental, market, and agronomic influences that impact crop decisions.

Within the Data Processing layer, input data passes through Data Cleaning to eliminate inconsistencies and missing values and Feature Engineering to extract informative variables that enhance model performance.

The Model Layer consists of two base models, Price Prediction, which gives predictions of future crop prices, and Crop Recommendation, which recommends the best crops based on data analyzed. These models utilize machine learning algorithms for pattern and trend detection.

The Output Layer delivers users with Price Forecasts, Crop Suggestions, and in-depth reports through a User Dashboard and Report Generation features.

Finally, the system benefits End Users—mainly farmers and agronomists—by providing data-driven information to facilitate informed decision-making, enhance profitability, and encourage sustainable agriculture.

The historical methods employed in the forecasting of farm prices have traditionally been suffering from serious inefficiencies that prevent their effectiveness. The traditional methods usually lack accuracy, experience lags in updating, and are normally unable to handle the intricate and multivariate nature of the data needed for useful forecasting. Variables such as shifting climate patterns, fluctuating market demand, market dynamics, and regional variations add levels of complexity that traditional means are not suited to deal with.

Farmers, traders, and policy-makers consequently make decisions in light of outdated or incomplete information, which impacts adversely on agricultural planning, market participation, and profitability. This reality highlights the imperative need for a sophisticated, intelligent solution that can evolve with these complexities and provide timely, accurate, and actionable insights.

To address these challenges, a strong AI-based solution is being proposed, blending strong machine learning models with a contemporary web development framework. By combining Long Short-Term Memory (LSTM) networks and Random Forest algorithms into a Django-based web platform, the system is intended to provide a dynamic and precise agricultural price prediction tool. This smart solution is intended to run smoothly in real time, giving users instant access to predictions that are specific to certain crops, locations, and time periods. The platform not only facilitates improved planning for the markets but also contributes to reducing losses after harvests and enhancing the income of farmers through well-informed decision-making.

LSTM models are uniquely suited for such an application due to their time series analysis capacity, capturing long-term dependencies and modeling seasonal trends well. LSTM models are capable of identifying long-term patterns from historical commodity prices, which are essential in the prediction of future fluctuations. Conversely, Random Forest models offer robustness in processing structured, non-linear data sets with multiple influencing factors like

weather conditions, crop variety, regional demand, and market availability. Having both temporal patterns and feature-based relationships captured, the combination of models guarantees better prediction accuracy.

The infrastructure that hosts this intelligent system is developed with Django, a high-level Python framework with its reputation for security, scalability, and efficiency. Django forms the backbone for backend operations, managing data with consistency, integrating models, and authenticating users. Further, it supports easy integration with third-party agriculture-related data sources and APIs, such as weather, mandi prices, and crop advisory. The frontend interface is crafted to be responsive and user-friendly, enabling users such as farmers in rural locations to use the system comfortably on web or mobile platforms.

One of the most notable features of this system is its ability to provide real-time forecasts. Users can simply provide information like the crop in question, geographical location, and time window they are interested in to get immediate price predictions. These predictions are made by the trained models in the backend and displayed through a simple, easy-to-use interface. This immediacy and accessibility enable stakeholders to respond to information immediately, facilitating better planning, smarter sales tactics, and more lucrative results.

Technical, operational, and business specifications are carefully met in the development process. From a technical perspective, the employment of LSTM and Random Forest models enables the system to effectively process both temporal and multivariate data.

Real-time data ingestion and processing functionality is built into the backend so that the system remains responsive even with heavy loads of data. Functionally, the interface is user-friendly, with multi-language support and mobile device compatibility to ensure maximum accessibility. Integration with current agricultural advisory platforms further enhances the utility of the system.

From a business perspective, the solution is designed to minimize financial losses and post-harvest losses by providing real-time, data-driven price information. This will enable farmers to make better decisions on when and where to sell their produce, with a view to earning more and enhancing market transparency. The general workability of the system is supported by its technical viability, scalability, and affordability. It has the ability to support a wide variety of

crops and respond to differing regional market prices and thus can be applied at either local or national scales.

The machine learning algorithms are then trained against vast amounts of historical data such as commodity prices, weather conditions, and market conditions. The data is preprocessed with methods such as normalization and feature selection to enhance model precision and performance. During model training, performance is regularly monitored and improved to enable the system to perform well in generalizing to new unseen data.

As far as platform development goes, the architecture is scalable and fault-tolerant. Django handles backend operations and hosts the trained models, while the frontend provides visualization tools for trends in forecasts, historical price charts, and other analytics. Users are not only able to view future price trends but also examine how those trends have evolved over time, providing them with greater insight into market behavior.

In order to sustain high performance across changing loads, the system employs various optimization strategies. These involve caching data, multi-threading or parallel processing, and load testing to support minimal latency and optimal responsiveness. Many queries are computed ahead of time and cached for faster access during heavy usage. When more crops and users are introduced, the platform can extend horizontally to increase capacity.

Lastly, the deployment strategy is flexible in that it accommodates both cloud and on-premise hosting, depending on the resources and requirements of the stakeholders. This makes the system available and usable across various user groups, ranging from large agro-organizations to smallholder farmers. Through its holistic design, the proposed solution is a leap ahead in addressing the historic inefficiencies of agricultural price forecasting and facilitating data-driven decision-making in the agricultural ecosystem.

Chapter 2

LITERATURE SURVEY

AI-ML driven price forecasting systems for agri-horticultural products are emerging as paradigm-changing tools in the agri-business and agriculture industries. By integrating Artificial Intelligence (AI) and Machine Learning (ML) methodologies more specifically, time series forecasting and regression models these systems can make data-driven, automated, and scalable forecasts of commodity prices. The construction of such models builds on core theories, methods, and empirical results from a variety of disciplines including computer science, agricultural economics, statistics, and data science. Methods such as Long Short-Term Memory (LSTM) networks and Random Forest algorithms are key to understanding the temporal patterns and multi-variable interdependencies in price changes, providing farmers and stakeholders with precise inputs for planning and decision-making.

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 timeseries 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

Presidency School of Computer Science and Engineering, Presidency University.

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. Not with standing 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 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. None the less, 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 minimize manual effort and reduce 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 analyze 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 labor 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 analyze 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. Never the less, 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.

Chapter 3

RESEARCH GAPS OF EXISTING METHODS

Even though the use of machine learning in agriculture has been growing at a very fast pace, there is a wide gap in research on the creation of robust, real-time, and scalable AI-ML models for agri-horticultural commodity price prediction. Most of the current systems do not possess the predictive depth, flexibility, and contextual knowledge needed to cope with real-world agricultural market complexities.

Key Research Gaps

Despite the encouraging progress in the use of artificial intelligence (AI) and machine learning (ML) technologies for farmgate price forecasting in agriculture, an exhaustive analysis identifies a number of key research gaps that impede the successful deployment and adoption of such technologies on the national and regional levels. Filling these gaps is important to enhance the accuracy, scalability, inclusivity, and practicality of such forecasting systems for practical applications, particularly in a diversified and complex agriculture economy like India.

One of the foremost challenges is the absence of real-time contextual understanding in existing forecasting models. Most current systems are designed based on static datasets and historical trends, lacking the ability to dynamically adapt to real-time variables such as sudden changes in weather, transport and logistics disruptions, or policy modifications. These real-time factors can drastically influence market prices, and without their integration, the forecasts produced by AI/ML models often fall short in accuracy and relevance.

Another key issue is poor predictability of traditional models. Most traditional forecasting techniques are linear or time-invariant and unsuitable for dealing with the intrinsically non-linear, seasonal, and region-dependent characteristics of agri-horticultural price data. These limitations highlight the necessity for stronger, temporal models such as Long Short-Term Memory (LSTM) networks and ensemble techniques such as Random Forests, which are more capable of capturing intricate trends and periodic patterns in pricing data over time.

Additionally, interoperability with government and legacy data systems is minimal. These important official data portals like Agmarknet, the Agricultural Produce Market Committee (APMC) systems, and Farmer Producer Organizations (FPOs) have useful datasets that could collectively have a major impact on model accuracy. However, poor standardization and APIs for seamless data sharing are impediments for integrating these data sources with AI-based forecasting platforms. Seamless interoperability is crucial for end-to-end data ingestion, preprocessing, and model retraining.

Scalability is another point of contention, since most current systems are not effective when presented with large-scale, high-dimensional data. With the variety of crops, markets, and geographies in India, AI/ML models need to be capable of processing and analyzing data from various regions and market centers at the same time. This means putting in place strong data pipelines, cloud infrastructure, and storage and processing frameworks that can handle increasingly large datasets without sacrificing model responsiveness.

Concurrently, the computational power needed to train deep learning models, like LSTM networks, is usually too intense to implement in rural or resource-scarce settings. This creates concerns about the energy efficiency and cost-effectiveness of the systems, hence limiting their accessibility to small-scale institutions and farming cooperatives. Building lightweight models or utilizing edge computing can mitigate this challenge to a degree, but the gap continues to be salient.

Another critical area of concern is that market intelligence distribution is not sufficiently personalized. Existing systems tend to provide generalized analysis, not accounting for localities, individual farmers' profiles, soil health status, crop stages, or previous yields. The creation of adaptive models that would personalize the prediction according to given farmer characteristics and geographies would significantly improve their usefulness and relevance.

In addition, user interface and accessibility are poorly developed in a majority of tools and applications. Applications provide few or no mobile-based dashboards or multi-language capabilities that are needed for widespread adoption by Indian farmers, particularly those in rural and semi-literate societies. Interfaces need to be simple, mobile-oriented, and provided

in local languages to facilitate effective usage and participation.

Data protection and security are other issues. Because agricultural price statistics and market data are commercially confidential, it is important to ensure models adhere to data protection protocols. Regrettably, most existing systems do not have proper mechanisms for encryption, anonymization, and safe storage, opening them to breach and abuse.

A neglected but important issue is bias in training data, which may cause biased forecasts. Models being trained on the popular markets or highly demanded commodities only tend to overlook less-reported areas or crops, which may increase disparities and leave out marginal farmers from getting correct forecasts. This bias, if not addressed, could strengthen existing inequalities in agricultural economics.

Furthermore, domain-specific agricultural information e.g., the impact of pest infestations, irrigation cycles, local festivals, and cultural traditions has weak or no integration in most models. These localized, real-world conditions have a strong impact on crop prices, and their exclusion from model training data reduces contextual relevance to forecasts.

The prohibitively high cost involved in developing, deploying, and sustaining sophisticated AI-ML models is another impediment. Advanced AI-ML models demand technical expertise, high-cost infrastructure, and ongoing upkeep inputs beyond the capabilities of small and medium-sized agricultural businesses, Farmer Producer Companies (FPCs), and rural partners.

Equally troubling is the absence of meaningful user interaction and feedback mechanisms. After deployment, most systems fail to include farmer feedback or usage analytics to update and improve the models. This stagnation can lead to stale predictions and reduced user trust over time.

Additionally, AI-ML models are prone to model drift, particularly in agriculture where trends shift seasonally or suddenly because of climate fluctuations or market disruptions. Without periodic retraining through automated pipelines, models rapidly become outdated,

diminishing their reliability for repeated use.

Rollout of these systems throughout India's multilingual and geographically varied terrain also presents a significant challenge. Region-specific crops, language affinity, and market structures demand adaptive systems that can be tailored to local needs—a capability many existing platforms do not possess.

Moreover, most forecasting systems lack multimodal data fusion support. Their performance is restricted by only using historical price data. Including other data sources like satellite imagery, weather, soil moisture sensors, and remote sensing data has the potential to significantly enhance forecasting accuracy but has not been thoroughly explored.

Lack of uncertainty in AI predictions is also an issue, as existing systems seldom report prediction confidence levels or detect outliers and anomalies (e.g., floods, strikes, or disease outbreaks). Adding uncertainty modeling would make decisions better by alerting users to invalid or abnormal predictions.

In addition, adherence to government policies regarding farm data usage and sharing is usually ignored. Compliance with such policies is necessary for the deployment of AI tools in public or regulated settings, and existing models often ignore these guidelines, which can lead to regulatory non-compliance.

A corresponding gap is the absence of transparency and interpretability in model outputs. Most state-of-the-art models are "black boxes" with minimal explainability, and it is difficult for end users to comprehend or trust the predictions. Incorporating Explainable AI (XAI) techniques is essential to fill this trust gap.

Finally, the omission of real-time consideration of market sentiment analysis from news sources, public announcements, and social media caps forecast accuracy. Such unstructured data sources usually contain early pointers to price direction but are as yet not tapped by most systems.

Chapter 4

PROPOSED METHODOLOGY

The process of developing and testing AI-ML-based models for forecasting prices of agri-horticultural commodities is structured and systematic. A holistic approach comprises suitable technological frameworks, iterative development, and data-driven testing to attain performance, reliability, and accuracy in forecasting. This section briefly discusses the methods relevant to the proposed system, including design, development, testing, evaluation, and optimization to guarantee the goals of commodity price prediction efficiently and effectively.

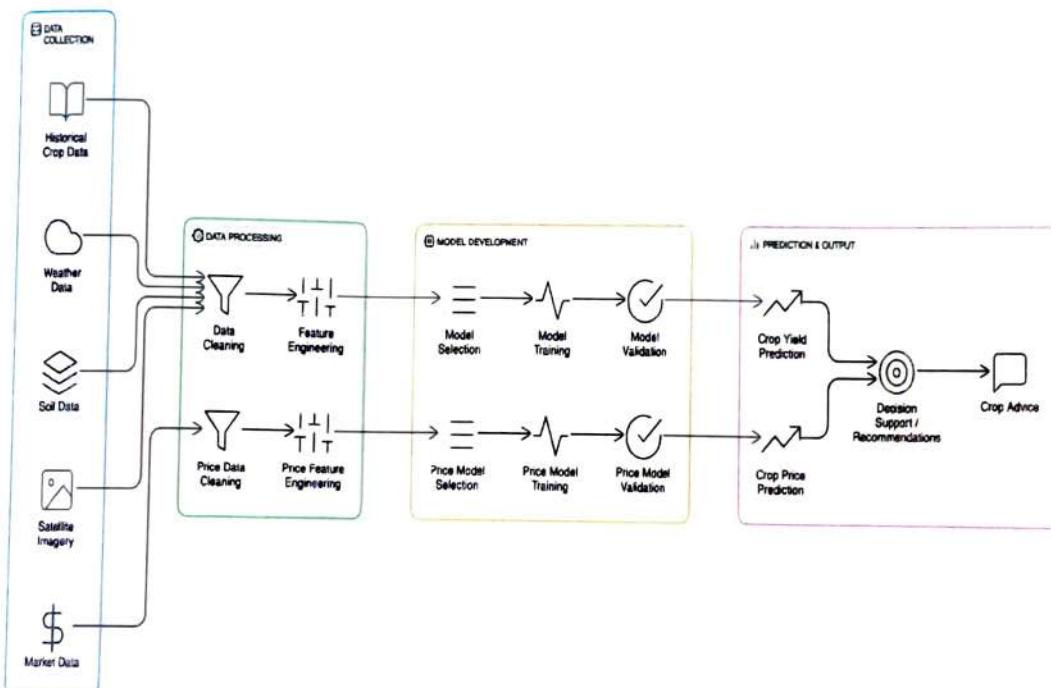


Figure 2: Steps for Crop Prediction

The Figure 2 represents the architecture of an AI-based system for agricultural decision-making, specifically crop yield prediction and price forecasting. It consists of four prominent stages: Data Collection, Data Processing, Model Development, and Prediction & Output.

The Data Collection layer collects inputs from various sources such as historical crop data,

weather data, soil data, satellite imagery, and market data. These inputs form the basis of information needed to develop strong predictive models.

At the Data Processing phase, there are two parallel streams of processing agronomic and price-related data. Both pass through data cleaning to eliminate noise and inconsistencies and feature engineering to pull out significant variables enhancing model performance.

The subsequent Model Development phase involves model selection, training, and validation for crop yield as well as price prediction tasks. This ensures that the most appropriate algorithms are selected, optimized, and tested for accuracy and reliability.

The last Prediction & Output layer produces crop yield and price predictions, which are synthesized into decision-making insights via a decision support system. These insights result in crop advice, helping stakeholders such as farmers and agronomists make informed decisions. In general, the system combines several streams of data and machine learning to deliver sound recommendations, enhancing productivity and profitability in agriculture.

Implementation Framework and Model Design

The implementation framework for price forecasting of agri-horticultural commodities integrates deep learning and ensemble learning approaches, specifically the Long Short-Term Memory (LSTM) network and the Random Forest (RF) algorithm. The models have been selected based on their own abilities to handle temporal and structured data. The entire design includes four main phases: data preprocessing, model construction, training and validation, and performance evaluation.

The hybrid model is designed to learn temporal relationships in commodity price variations using LSTM and leverage Random Forest for robust learning from structured, non-temporal data such as weather, demand signals, and socio-economic variables. The dual-model design enables extensive modeling of price movements from sequence-based as well as static perspectives.

LSTM for Sequential Data Forecasting

LSTM networks, a specially suited class of Recurrent Neural Networks (RNNs), are specially suited for prediction tasks involving time series data. For agri-horticultural price forecasting, commodity prices usually exhibit sequential dependencies, trends, and seasonality on daily, weekly, or monthly levels—therefore, LSTM would be the optimal choice. LSTM cells consist of memory cells and three primary gates: the input gate, the forget gate, and the output gate. These gates manage information flow, allowing the network to store or erase information at a certain time step. The new architecture avoids the vanishing gradient problem that occurs in vanilla RNNs and enables LSTM to maintain long-term dependencies in information.

In this study, the LSTM model is built with one or more hidden layers, and each hidden layer is a sequence of LSTM cells. The input to the model is a multivariate time series of historical prices along with suitable features such as temperature, rainfall, seasonality factors, and festival months. The output of the model is the predicted price of a commodity at a specific future time step.

Weaknesses of RNN and Strengths of LSTM

Vanilla RNNs process sequences by using a hidden state that is updated at every time step. Theoretically capable of learning dependencies in sequences of any size, RNNs are poor at dealing with long-term dependencies due to the vanishing or exploding gradient problem in backpropagation through time (BPTT). This complicates learning about correlations between distant events in the sequence, an especially problematic task in applications like price prediction where past trends and cycles drive future values.

LSTM avoids this drawback by introducing a memory cell and gating scheme that allows it to learn to forget, retain, or output information. Such mechanisms ensure the stability of gradient values over time, thus allowing effective learning on long sequences. As a result, LSTM networks work significantly better than standard RNNs on tasks involving long-range temporal dependencies such as commodity price forecasting, weather forecasting, and financial time series analysis.

Random Forest for Structured Feature Learning

Random Forest is a popular ensemble learning algorithm that builds many decision trees at training time and makes their predictions average out in the case of regression tasks or takes a majority vote in the case of classification problems. Its efficiency in handling both continuous and discrete variables, resistance to overfitting, and high predictive accuracy make it well-suited to structured data analysis.

In this study, Random Forest is used as a secondary model to Long Short-Term Memory (LSTM) networks with particular attention to structured, non-sequence features affecting agricultural price movements. Such features encompass market-specific parameters like market size and transportation costs, climatic features such as temperature and humidity, policy factors like subsidies, past supply and demand data, and information regarding soil health and crop output. The model is trained on tabular data, with each row corresponding to a unique time period and commodity along with its respective attributes.

One of the strengths of the Random Forest algorithm is that it can produce feature importance scores, enabling researchers to determine the most significant variables in predicting prices. This interpretability is extremely useful for stakeholders like policymakers and farmers, who depend on clear insights to inform their decisions.

In addition, the predictions and insights derived from the Random Forest model are combined with outputs from the LSTM model, employing methods such as stacking or weighted averaging. This ensemble method improves the overall prediction system by taking advantage of the strengths of both model's temporal dynamics through LSTM and static, structured features through Random Forest.

LSTM Architecture and Working

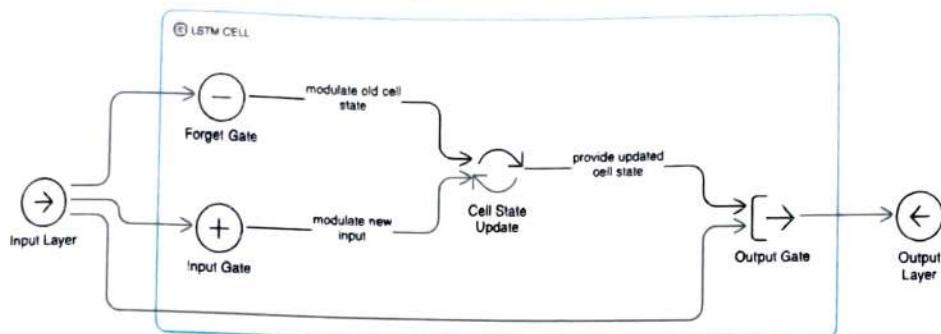


Figure 3: LSTM Architecture diagram

The Figure 3 represents the internal structure of an LSTM (Long Short-Term Memory) cell, a basic building block of LSTM networks that are widely employed for time series forecasting and sequence modeling. The cell is optimized to handle long-term dependencies well by having a cell state, which serves as a memory. The Input Layer supplies the input data at hand, which enters three gates—Forget Gate, Input Gate, and Output Gate. The Forget Gate dictates what percentage of the past cell state must be forgotten, essentially "forgetting" unnecessary information. At the same time, the Input Gate controls the new input to determine what new information will be remembered. These gates collaborate in the Cell State Update stage, where the cell melds the preserved memory and fresh input to update the internal state. The new cell state is then handled by the Output Gate, which decides to what extent the cell state will be a factor in the next hidden state or output. The end result is routed through the Output Layer. This gate mechanism enables LSTM cells to maintain vital information for long sequences of data and to exclude noise or less significant information, hence the efficiency in language modeling, stock prediction, and farm time series forecasting.

Long Short-Term Memory (LSTM) Architecture

It has been a technique for Recurrent Neural Network designs, especially devised to process data sequentially without showing the weaknesses posed by traditional architectures of RNNs, their failure to adapt long-term dependences due to the vanishing gradient problem. It contains two major components: it includes memory cells and special gate controls.

Major Components of LSTMs:

LSTM uses three gates—forget, input, and output gates to manage the flow of information effectively. These gates consist of neural layers (typically using sigmoid activation) that decide what to keep, update, or discard.

Forget Gate

Decides which information to discard from memory. Removes irrelevant info from past interactions, helping the chatbot focus on the latest context when the topic changes. So the formula is:

$$ft = \sigma(Wf \cdot [ht-1, xt] + bf)$$

Output value between 0 and 1, where closer to 1 retains more past info closer to 0 forgets it.

ft: The forget gate's output (a value between 0 and 1)

σ : The sigmoid activation function (squashes values between 0 and 1)

Wf: The forget gate's weight matrix

ht-1: The previous hidden state

xt: The current input

bf: The forget gate's bias

Input Gate

It Controls how much new information should be written to memory. Ensures relevant parts of each new query are stored for future use, especially details in ongoing conversations. So the formula is $it = \sigma(Wi \cdot [ht-1, xt] + bi)$. Output value between 0 and 1, with closer to 1 storing more new info.

Memory Cell

The core part of an LSTM is the memory cell, designed to hold information over long time periods. It can be called the "long-term memory" of the system. The cell state flows through the network with only minor linear interactions, allowing it to preserve critical information while updating itself when needed.

Output Gate

Decides which information from memory to share as output. Selects contextually relevant info

to form the chatbot's response based on cumulative memory. So the formula is :
 $ot = \sigma(Wo[ht-1, xt] + bo)$. Output value close to 1 means more memory info is shared in the response.

Candidate Memory (Memory Update)

Updates the long-term memory, or cell state, based on input and forget gates. The cell state maintains key info across time steps; helps preserve relevant info like account details. So, the formula is $Ct=ft*Ct-1+it*C_{-t}$.

Working of LSTM

An LSTM network's architecture is essentially not the same as that of usual RNNs. Instead of employing a recurrent connection, an LSTM brings into play a memory cell that does not forget data over time and has three indispensable gates called the input gate, the forget gate, and the output gate. These gates manipulate the flow of information into the memory cell, within the memory cell, and out of the memory cell. The input gate controls what percentage of the fresh input is to affect the current memory, the forget gate determines how much of the previous information to forget, and the output gate controls how much of the current memory content to use to create the output and transfer to the next time step. This design allows LSTMs to maintain significant information for long durations and eliminate unimportant data, thus enabling them to concentrate on important long-term trends.

These gates' behavior is learned through training. The model learns to prioritize higher or lower importance for input information, internal memory, and output depending on the task being executed. For example, in agricultural price prediction, the model is able to learn that some seasonal patterns or weather anomalies experienced months ago remain applicable when predicting commodity prices in the future. In contrast to other less complex models, LSTMs can maintain a trail of reasoning over long sequences without losing sight of what is significant, which is a primary necessity in areas where context and historical dependencies are paramount.

Practically, LSTMs have been very helpful in tasks that require time series analysis like stock market forecasting, language modeling, speech recognition, and most importantly in

agriculture for tasks such as crop yield estimation, weather forecasting, and commodity price forecasting. The fundamental cause of the success of LSTMs in such areas is that they can process data in a sequence where the order of the data points is of crucial significance.

In agricultural price prediction, LSTM networks are especially suitable because price data is highly sequential and temporally dependent. Agricultural marketplaces are determined by a broad range of time-specific variables such as seasonality trends, supply-demand phases, weather conditions, festival periods, policy measures, transport failures, and international market fluctuations. The prices of commodities such as onion, potato, pulses, and vegetables follow seasonal patterns and unexpected spikes or declines, which are usually difficult to model using conventional statistical models. LSTM networks are able to capture these non-linear, time-dependent patterns much more effectively by learning from a long sequence of past prices and supporting data over time.

Application of LSTM in agricultural price forecasting starts with the generation of suitable datasets. Past price data is usually obtained from official sources such as Agmarknet, APMCs (Agricultural Produce Market Committees), and other good quality online/offline datasets. This information is then preprocessed to create input sequences. Each sequence is a fixed number of previous observations that the model uses to make predictions about future prices. This window-based method enables the model to learn temporal relationships between time steps. The data is typically cleaned and normalized to make it consistent and to enhance training efficiency. It can also contain extra features like temperature, rainfall, soil moisture levels, policy decisions, or market size measures to enrich the input and make it context-aware. At training time, the LSTM model is shown various such sequences along with their respective target outputs, i.e., the real future price that resulted from those past trends. Through multiple iterations and epochs, the model learns the mapping between the input sequences and the desired outputs, updating its internal weights and gate activations in the process. The trained model can then be utilized to predict prices for future dates by providing it with the latest data and allowing it to project the probable price trend.

One of the powers of LSTM models is how they can generalize well over new data if properly trained. In agricultural forecasting, this translates to the fact that even if the model has never

been exposed to a specific weather anomaly or festival period, it can still generate realistic forecasts by applying the patterns learned from similar past instances. Additionally, LSTMs are stateful and adaptable in processing sequences of different lengths, though the input is typically processed through a fixed-size window in practice to maintain uniformity and for computational simplicity.

For greater accuracy and clarity, LSTMs may be integrated with other models such as Random Forests within a hybrid structure. While LSTMs are best at capturing sequential dependencies, Random Forests are well-suited to work with structured, non-sequential data like categorical variables or indicators that do not change within a time window. By combining the outputs of both models, either through weighted averaging or a stacking ensemble approach, the system can benefit from the strengths of both classes of models—temporal memory from LSTM and feature-level information from Random Forest. This blend is particularly valuable in agricultural price prediction where past trends and present static characteristics drive price action.

Nonetheless, applying LSTM models to actual agricultural systems is not without its issues. To begin with, the computational power needed to train and implement LSTM networks can be huge, particularly when dealing with big data and long sequences. In rural or edge environments where there is limited computational infrastructure, this is a constraint. Optimization methods such as quantization or the utilization of light-weight LSTM variants can alleviate this problem. Secondly, the interpretability of LSTMs is quite low. In contrast to linear regression models whose coefficients reflect directly the influence of features, LSTMs are black boxes. This can limit trust and uptake by stakeholders like farmers, traders, or policymakers. To counter this, methods from the field of Explainable AI (XAI) are being applied to LSTM structures to emphasize what components of the input sequence played the most crucial role in predicting the outcome.

Agricultural commodities are subject to quickly changing forces—climate change, political trends, unexpected bursts of demand, or pest epidemics. A model based on data from previous years can no longer be reliable if these trends shift dramatically. To ensure reliability, LSTM

models must be fine-tuned or retrained from time to time based on new data. This can be managed by setting up automatic pipelines that routinely harvest new data, preprocess it, and update the model. This process keeps the model up to date and continues to make accurate predictions.

Deployment-wise, LSTM-based systems are being incorporated into mobile apps and dashboards where farmers can look at future prices. These systems are made user-friendly, mostly translated into regional languages, and provided with graphical visualizations so that they become end-user accessible in rural locations. A farmer may enter their location, commodity they are interested in, and optionally weather or soil conditions, and the LSTM-backed system can not only give them the forecasted price for the next weeks or months but also recommendations like when to harvest or sell.

In addition, LSTM models play a role in general agricultural decision-making. By making price predictions more accurately, they reduce the uncertainty faced by farmers and traders. Having knowledge that the price of a commodity is expected to decrease or increase enables them to plan their production, storage, and selling accordingly. It also helps policymakers formulate subsidy or procurement policies and ensures food security through stabilizing prices through smart interventions.

In summary, Long Short-Term Memory networks are a strong and effective solution for modeling sequential data in agricultural price prediction systems. Their capacity to learn and remember long-term dependencies, adjust to intricate and non-linear relationships, and blend well with other machine learning models makes them perfect for forecasting tasks in changing environments such as agriculture. Although computational requirements, interpretability, and data availability challenges persist, continued progress in model optimization and explainability is increasingly closing these gaps. Through the use of LSTM-based forecasting models, actors throughout the value chain for agriculture—ranging from smallholder farmers to government agencies—can make more informed, timely, and strategic choices, ultimately leading to increased economic stability and resilience for the food system.

Random Forest Architecture and Working

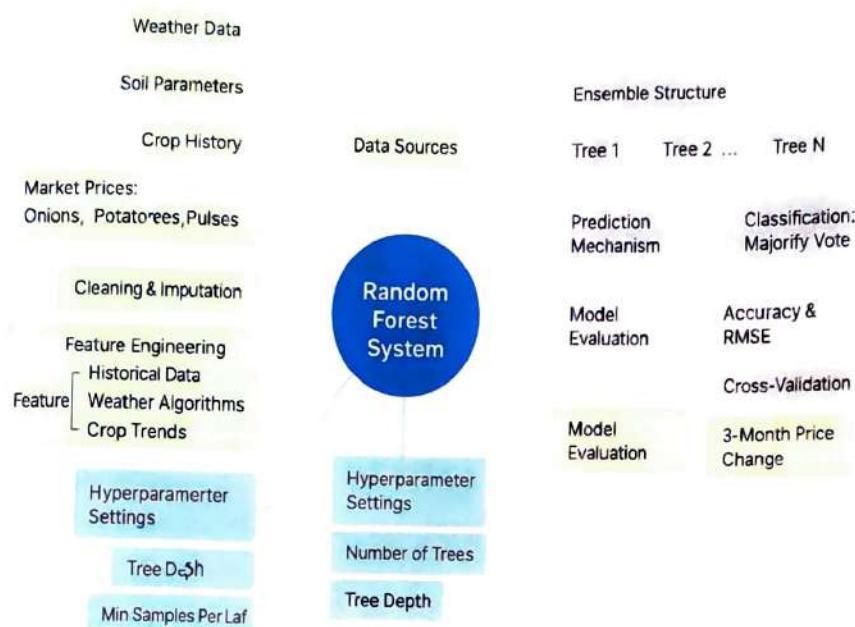


Figure 4: Random Forest Architecture

The Figure 4 represents a conceptual description of a Random Forest System intended for agricultural market forecasting, specifically for onions, potatoes, and pulses. The central components of the system are various important elements that make the system work. Data Sources are weather, soil factors, crop history, and market prices, which are the basis for modeling. Prior to training, data imputation and cleaning are used to manage noisy or missing data. The feature engineering process subsequently obtains informative variables like historical trends, weather-based models, and crop patterns to enhance model accuracy.

The Random Forest algorithm itself is based on an ensemble framework, wherein several decision trees (Tree 1 through Tree N) work together. For classification problems, it utilizes a majority vote system. The hyperparameters of the model, including the number of trees, the depth of trees, and the minimum samples per leaf, are optimized for best performance. The system is subjected to model evaluation based on accuracy measures, RMSE (Root Mean Square Error), and cross-validation methods. It is also used for 3-month price change

prediction, assisting in short-term agricultural planning. In total, this Random Forest system combines mixed datasets and sophisticated machine learning algorithms to generate accurate crop price predictions and aid in agricultural decision-making.

Random Forest is a strong ensemble learning technique that works by building many decision trees at training time and predicting the class which is the mode of the classes or average prediction of the individual trees. It is extensively praised for its strength, ability to perform both classification and regression problems, and its ability to handle large data with greater dimensionality. The "forest" it constructs is a collection of decision trees, typically trained with the "bagging" procedure. The principle is to mix multiple learning models in order to enhance the general performance.

The power of Random Forest is that it is straightforward and yet immensely powerful it dampens overfitting that is usually the curse of single decision trees, increases accuracy, and provides consistent outcomes for a variety of tasks. The interpretability and flexibility of the model have made it appealing for complicated real-world tasks, particularly in applications where data diversity is high and multiple features account for the output. Agricultural price prediction is such an application area.

At the center of Random Forest is the ensemble learning architecture, where many decision trees collaborate to generate the final output. Each tree in the forest is trained independently on a randomly chosen subset of the training data and a random subset of features. This step adds variation and avoids the trees being similar, which is essential to the success of the ensemble.

Each decision tree is a tree structure with flowchart-like appearance, which recursively divides the dataset by feature values in order to reach a decision or a prediction. For Random Forest, each tree is grown to the maximum extent without pruning, and the ensemble learns as much information as possible. Still, to prevent overfitting, constraints such as maximum depth (Tree Depth), minimum samples per leaf node (Min Samples Per Leaf), and the number of trees (Number of Trees) are properly set by hyperparameter tuning.

The architecture also utilizes bagging (Bootstrap Aggregating)—an approach in which each tree is trained on a bootstrap sample (sampling with replacement) of the original data. That is, some data points might appear many times in the training set for one tree but not at all in another. The randomness imparted in data selection as well as feature selection guarantees that trees are decorrelated, which causes a more general model.

Working Mechanism of Random Forest

The Random Forest process starts by preprocessing the training set, and that includes feature selection and sampling. Each tree is trained independently based on a separate subset of data. Every tree learns patterns and rules as per its particular set of data when it gets trained. The training process of each tree concludes, and afterwards, all trees get utilized in aggregation to provide the predictions.

In the case of classification problems, such as classification of crop quality or determining the best type of crops, every tree "votes" for a class, and the most voted class is the end prediction this is termed majority voting. For regression problems, such as forecasting the price of crops in the future, the forest averages all the individual tree predictions.

One of the characteristic features of Random Forest is that it can compute feature importance, and this gives information about the variables that have the most significant influence on making a prediction. This proves very valuable in agri-based use cases where features such as soil parameters, temperature, rain, and market trends are available for analyzing their influence on pricing.

The Random Forest is also very much compatible with model evaluation metrics such as Accuracy, Root Mean Square Error (RMSE), and Cross-Validation techniques. The metrics aid in measuring the model's performance as well as the ability of generalization. Cross-validation makes sure that the model is not overfitting and retains its prediction power on unknown data.

Hyperparameter tuning and feature engineering are essential for optimizing a Random Forest model in your agricultural price prediction project. The number of trees (n_estimators)

determines how many individual models the forest will contain, and more trees generally improve accuracy by averaging out errors, though too many can increase computational time. Trees' depth (`max_depth`) is an important parameter that manages model complexity; shallow trees avoid overfitting but can underfit, while deeper trees could learn noise and overfit. Minimum samples per leaf (`min_samples_leaf`) avoids overfitting by making leaf nodes data-sufficient, and maximum features (`max_features`) governs how many features to use in splitting nodes and thus affects model diversity and speed. Hyperparameter tuning balances variance and bias and prevents the Random Forest model from becoming overly complicated while learning effectively. In your agricultural price prediction system, effectively tuned hyperparameters will enable the model to recognize seasonal patterns, market trends, and crop dynamics.

Feature engineering is also critical since Random Forest models work optimally when fed with rich, well-prepared features. Critical input sources such as weather, soil factors, crop history, and price levels are core to the performance of the model. Weather inputs such as rainfall, temperature, humidity, and wind speed influence crop yields directly, and engineered features such as cumulative rainfall or average temperature variations can give added insights. Soil conditions like pH, water content, and nutrient status affect crop suitability, and indexes derived from these parameters can improve the model's ability to understand yield potential. Crop history information, such as previous yield trends and pest infestations, adds context, and market price information provides insight into supply and demand. Additional features such as crop trends, price momentum, and regional variables further enhance the model's power to forecast future prices. Another key area is missing data handling, and the use of imputation methods such as mean substitution or KNN-based imputation ensures robustness and avoids biasing the model with missing values in the data. By combining varied data sources and adjusting the model's hyperparameters appropriately, your Random Forest model shall be able to better forecast agricultural prices, detect market trends, and suggest crops for optimal profit potential.



Model Evaluation in the Agricultural Context

In your system, model evaluation serves a twofold purpose: accuracy assessment and real-world usability validation. Measures such as RMSE measure how close the forecasted prices approximate real market prices, providing a tangible idea of the prediction error of the model. A low RMSE indicates farmers or stakeholders can rely on the forecasts. Cross-validation methods such as K-Fold or Time Series Split are particularly well-suited to time-sensitive agricultural data. These allow validation of the model's generalization capability across unseen seasons or market conditions. One distinctive feature of your system is its analysis of 3-month price movements, which reflects medium-term market dynamics important for planning. Farmers frequently have to make decisions months ahead of what to plant, and this insight provides a predictive advantage.

Random Forest in Your Agri-Price Prediction System

Your project optimally utilizes Random Forest in the prediction of agricultural commodity prices. The ensemble model is trained on historical price data, soil and weather attributes, and crop data to predict prices for onions, potatoes, and pulses. This prediction is of utmost importance in aiding farmer decision-making. For example, if it is forecasted through the model that onion prices will increase in the coming quarter, farmers may decide to plant more onions. If pulses are forecasted with stable or declining prices, they may diversify so as not to incur losses. Random Forest model also aids in crop suggestion. Based on soil factors and weather, it can suggest the optimal crops to cultivate in an area, considering economic yields through price predictions. This twofold utility prediction as well as recommendation makes it even more valuable in agri-tech. In addition, the model is embedded in your app interface so that users may choose crops via a dropdown list and see predicted prices along a time horizon. The result is displayed using graphs that show previous prices and predicted future trends with clean dates (such as "1 May 2025") and prices ("₹/kg"), thereby being user-friendly.

Chapter 5

OBJECTIVES

The goal of developing AI-ML based models for predicting agri-horticultural commodity prices with LSTM and Random Forest is to facilitate data-driven, timely decision-making by farmers, traders, policymakers, and agri-businesses. The model aims to improve the efficiency of markets, minimize post-harvest losses, enhance farmer returns, and stabilize price volatility through state-of-the-art forecasting.

Through the generation of real-time and accurate price forecasts, the system aims to eliminate uncertainty in agriculture markets by instantaneously analyzing past prices along with complementary information such as weather, rainfall, and market demand. This makes possible both short- and long-term forecasts available at any time, improving decision-making, avoiding financial risks, and enhancing farmer profitability.

The model also includes contextual knowledge of market movements in terms of identification of temporal relationships, seasonal cycles, and market behavior, which enables it to predict prices based on changes in regional and environmental conditions. It provides tailored market intelligence by offering commodity-specific trend analysis, recommending best-selling times, and giving advanced warning of expected price movements based on local conditions.

With India's agricultural diversity, the model is capable of supporting multiregional and multilingual forecasting, and issuing localized price forecasts and language variations for farmers, market aggregators, and exporters. The system is able to handle large and complex data sets from various sources such as government portals, past prices, weather forecasts, crop yield reports, and input costs trends. LSTM addresses the sequential nature of the complexity in the speech data, and Random Forest processes multidimensional, tabular data efficiently.

The model is also designed for efficient training with LSTM learning long-range dependencies and Random Forest handling missing data, whereas reinforcement learning makes sure the accuracy of price prediction continues to improve. Through automation of market analysis and doing away with the requirement for costly consultancy, the system is economical and can be

accessed by government agencies, FPOs, and rural advisory centers.

Further, dynamic price forecasting with real-time sentiment analysis using Natural Language Processing (NLP) can adjust the forecast of prices in response to news, changes in policy or demand, giving timely insights into how external shocks affect agricultural prices. These abilities enable the model to identify demand-supply gaps, support Minimum Support Price (MSP) policy-making, and maximize agricultural supply chains. Lastly, the system provides strong data security and privacy by anonymizing and encrypting data pipelines, following regulatory requirements like the Digital Personal Data Protection Act, and practicing ethical data governance to establish user trust and confidence in the system.

Chapter 6

SYSTEM DESIGN & IMPLEMENTATION

The creation of an AI-ML-based system for forecasting agri-horticultural commodity prices requires a holistic approach that combines different components such as data acquisition, preprocessing, model development, deployment, and user interaction. The system intends to offer precise, real-time price forecasts to help farmers, traders, policymakers, and agri-businesses make rational decisions.

Introduction

Agricultural markets are inherently unstable, as they are driven by weather patterns, seasonal factors, supply and demand considerations, and policy initiatives. Classical statistical tools tend to struggle to identify the intricate, nonlinear relationships inherent in agricultural data. Machine learning (ML) and deep learning (DL) tools bring new hope for managing these complexities. Particularly, Long Short-Term Memory (LSTM) networks and Random Forest (RF) algorithms have exhibited potential in time-series forecasting and managing high-dimensional data, respectively.

System Architecture Overview

The implementation and design of the AI-ML-based agri-horticultural agricultural commodity price forecasting project consists of an integrated and end-to-end methodology that fuses data science, artificial intelligence, software development, and user interface engineering to provide a resilient, scalable, and efficient solution to agricultural stakeholders.

The objective is to solve the root problem of price fluctuation in the agriculture industry by facilitating evidence-based decision-making through data-driven forecasting. The foundation of the system starts with the gathering of information from a number of trusted sources. These include government agri-portals like Agmarknet, APMC, and e-NAM, which offer historical price figures on commodities such as onion, potato, pulses, and vegetables. Other data streams are real-time weather data from APIs like OpenWeather and IMD, which provide parameters such as rainfall, humidity, temperature, and wind speed that have a major impact on crop yield and hence market price. Soil properties like pH level, moisture, and organic content, and

agronomic parameters like sowing and harvesting dates are also included.

All this data, once gathered, goes through a strict preprocessing phase. Raw data usually has inconsistencies, missing values, and noise that must be resolved. Cleaning (deletion of duplicates and errors), normalization (scaling of numeric values), encoding of categorical variables, and sophisticated imputation techniques for dealing with missing values are preprocessing steps. Feature engineering forms a core component of this step, in which new features are generated to measure market momentum, seasonal behavior, rainfall indexes, and other determining parameters that contribute to improving the predictive nature of the model. Having the data ready, modeling starts based on a hybrid machine learning process utilizing both deep learning and ensemble learning methods. The Long Short-Term Memory (LSTM) network, a recurrent neural network, is employed to learn the temporal relationship in the historical price series. LSTM's ability to capture long-term dependencies positions it well for learning trends and seasonal fluctuations in agricultural prices.

At the same time, a Random Forest Regressor model is trained on structured, tabular data containing features such as weather, soil, crop type, and demand in the marketplace. Random Forest, a collection of decision trees, is very efficient in handling missing values and offering insights into feature importance, which is essential in understanding the drivers of price movements. These two models are combined in an ensemble framework using methods like weighted averaging or stacking to merge their predictions and enhance overall accuracy. After training, the models are tested using strong performance metrics like Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R-squared (R^2) score to assess their accuracy and generalization ability. Cross-validation and residual analysis also ensure model performance.

For deployment, a backend server is implemented with Flask or Django frameworks. This server takes care of API calls, processes data received from the frontend, passes it through the models, and sends back the predictions of prices. The backend also has secure handling of data in the form of encryption protocols, authentication of users, and authorization procedures. At the frontend, a user interface is developed using web development technologies such as HTML, CSS, and JavaScript or mobile development technologies such as Flutter. This

interface enables users to communicate with the system by choosing commodities, providing pertinent data, and seeing prediction results through interactive dashboards and graphs. The interface is intuitive and offers multilingual support in order to address different user bases in different regions.

In addition, the system is developed to be scalable and secure. It makes use of cloud services for the storage of data and deployment of models, with the guarantee that it will process large amounts of data and user requests in parallel. Technologies such as Docker and Kubernetes are used for containerization and orchestration, which result in effective management of resources and fault tolerance. Data privacy law compliance, like the Digital Personal Data Protection Act, is also ensured strictly. Dynamic learning is also included in the system. The models are built to be periodically retrained with fresh incoming data to keep pace with evolving market conditions. A feedback loop is provided to enable users to give feedback on prediction accuracy, which is utilized to further improve the models. The performance of the system is monitored in real-time by monitoring tools and developers are notified of any anomalies.

Sophisticated functions like Natural Language Processing (NLP) are incorporated to process market sentiment from news stories, government policy releases, and social media patterns. This information is utilized to dynamically modify forecasts in accordance with real-world events. Early warning systems also exist to warn users of impending price bursts or crashes so that users can take proactive action. Consequently, the system not only functions as a forecasting system but also as a strategic decision support system for farmers, market aggregators, policymakers, and agri-businesses.

Data Preprocessing and Collection

Effective forecasting in farm price forecasting systems largely relies on the quality, richness, and completeness of the input data. In this system, data is compiled from a wide variety of credible sources to form a solid data set for training and forecasting. Historical price information of markets is gathered from authorized websites such as Agmarknet and other Agricultural Produce Market Committees (APMCs), which offer a comprehensive database on commodities like onion, potato, and pulses' price movements over time. Historical pricing

information is extremely important in terms of establishing seasonal variations, price cycles, and demand-supply patterns recurring year after year or during particular climatic conditions. In addition to market prices, weather information plays a similar importance. Factors like wind speed, rainfall, humidity, and temperature are derived from reliable meteorological APIs such as OpenWeather and the Indian Meteorological Department (IMD). The meteorological factors directly affect crop health and yields and subsequently the supply-side dynamics and, ultimately, the market prices. Additionally, the system consolidates information regarding soil health and crop-specific factors such as soil pH, water content, and nutrient levels along with agronomic information such as sowing times, harvesting times, and crop rotation cycles. These aspects add greater insight into the agronomic factors that affect production quantities and crop quality and provide a richer perspective on likely market movements.

Besides organized datasets, unstructured external data like agricultural news releases, government policy statements, and social media trends are also processed to identify real-time sentiment and public opinion about major commodities. This is done by employing sophisticated Natural Language Processing (NLP) methods, which assist in extracting sentiment scores and detecting major events or policy changes that could drive market behavior. After being gathered, the raw data is subjected to strict preprocessing to make it clean, uniform, and machine learning model-ready. Preprocessing starts with data cleaning, which includes the elimination of duplicate records, handling missing values using different imputation methods like mean substitution or K-Nearest Neighbors (KNN) imputation, and fixing anomalies or inconsistencies in the data. This is then succeeded by feature engineering, where the new variables are designed to incorporate hidden patterns and relationships.

Instances of engineered features include price momentum indicators that reveal recent price deceleration or acceleration, rainfall indices that report cumulative precipitation in critical growing stages, and season trend variables that indicate particular agriculture calendar events. In addition, the data is normalized and encoded to be compatible with the machine learning models employed in the system. Numerical attributes are scaled to transform them into a standard range, usually applying standardization or min-max scaling, which enhances model convergence and avoids bias towards variables with larger values. These categorical variables like market location, crop type, or season are encoded based on methods like one-hot encoding

or label encoding so that the algorithms can handle them suitably. This step-by-step data preparation achieves that the Random Forest and LSTM models receive quality, structured input data that can provide accurate and meaningful price predictions.

Model Development

The agri-price prediction system is based on a hybrid model development approach that takes the strengths of both Long Short-Term Memory (LSTM) neural networks and Random Forest (RF) regression algorithms. This two-model approach guarantees that both temporal patterns and high-dimensional contextual features are well captured. LSTM, one of the recurrent neural networks (RNNs) variants, is particularly well-suited to work with sequential data and long dependencies, hence is well suited for time-series forecasting. Its design consists of memory cells and gates that enable it to learn and remember meaningful patterns over long intervals. In this project, the LSTM model is trained on past commodity price history consisting of daily, weekly, or monthly price levels. Key hyperparameters like the number of layers, number of units in each layer, learning rate, and dropout rates are carefully adjusted using sophisticated optimization techniques like grid search and Bayesian optimization. These hyperparameter tuning methods guarantee that the model is able to generalize well without overfitting to the training data.

At the same time, Random Forest regressors are used to identify nonlinear interactions between different exogenous variables affecting market prices, including weather, soil conditions, government policy measures, and other agronomic factors. Random Forest is an ensemble algorithm that builds multiple decision trees in training and returns the mean prediction of the separate trees. It is good at dealing with missing values, minimizing overfitting, and feature importance ranking. This ranking capability is most effective in identifying the factors that have the most crucial effect on commodity prices, thus giving the prediction process interpretability and transparency. For example, it might indicate that rainfall volatility has a greater predictive effect on onion prices compared to soil pH levels or that wind speed affects some vegetable prices more than others.

In order to improve the overall accuracy of forecasting, ensemble learning methods are used in the system to combine the predictions of both the Random Forest and LSTM models.

Weighted averaging, stacking, or blending methods are used to combine the outputs of the individual models. Weighted averaging gives relative weights to the output of every model with respect to historical performance, while stacking entails training a meta-model over the base models' predictions to acquire an ideal combination. Blending employs validation sets in order to mix up outputs in a straightforward but often just as effective way. This ensemble hybrid method leverages the sequential learning capability of LSTM and the strong multidimensional analysis of Random Forest to provide a robust and accurate prediction system immune to noise.

Model Evaluation

Model evaluation is an important step in the development pipeline to ensure the accuracy, reliability, and generalizability of the forecasting models. In order to measure the predictive performance of the model quantitatively, several evaluation metrics are utilized. The Mean Absolute Error (MAE) is one of the main measures, quantifying the average absolute deviations between predicted and true values. It gives an intuitive measure of prediction error with the same units as the target variable, hence simple to understand.

Another important measure is the Root Mean Square Error (RMSE), which punishes bigger errors more heavily than MAE. This is especially helpful in determining how consistent the model is across different levels of price volatility. Because RMSE places greater importance on the large errors, it is a good measure of the stability of the model during periods of market volatility.

Aside from error-based measurements, the R-squared (R^2) score is utilized to gauge the percentage of variance in the dependent variable that can be explained by the independent variables. A greater R^2 indicates that the model has explained most of the variance, i.e., it is a good fit. These measures are computed not only on training data but also on independent validation and test sets to make sure the model does not overfit and maintains its ability to generalize to unknown data.

For further verification of the stability of the model, cross-validation methods like k-fold cross-validation are used. In this approach, the dataset is partitioned into 'k' subsets, and the

model is trained and validated 'k' times, with each time a unique subset used for validation and the rest for training. This method provides assurance that each data point will have an opportunity to be in the validation set and assists in determining if the model's performance is stable with different data splits. Moreover, residual analysis is also performed to look at the divergence between actual and predicted values. Residual plots identify non-random patterns, outliers, or model bias, which can be corrected by model re-tuning or new feature addition.

Outside of quantitative metrics, diagnostic visualizations are incorporated for a deeper look at model performance. Visualizations like actual vs. predicted price plots, feature importance bar charts (in Random Forest), and LSTM loss plots are inspected in order to know where the model can be enhanced further. Plots are utilized not only to see how the model is performing but also the reason behind how it's doing it, so iterative cycles of model improvement are facilitated.

Deployment and User Interface

Once the model training and validation are accomplished, the system moves to the deployment stage where the models are embedded in a production platform available for end-users. The backend infrastructure is developed using strong frameworks such as Flask or Django, which host the machine learning models and serve them through RESTful APIs. These APIs act as the frontend application to backend model interface, allowing user inputs to be processed, predictions to be made, and results to be returned in real time. Pre-processing of user-input data, model inference initiation, and result formatting for presentation are also done by the backend server.

The frontend interface is accessible and usable, with special consideration for the broad diversity of the agricultural user community in India. The interface provides users with the ability to choose the commodity of interest e.g., onion, potato, or pulses and input relevant contextual information such as location (district/state), sowing or harvesting dates, and other optional agronomic inputs. The interface also graphically illustrates forecasted price movements in terms of interactive line plots, bar graphs, and displays of confidence intervals. They allow a direct, actionable perspective of historical prices as well as expected future prices and thereby enable making well-informed decisions on sowing, selling, and storing.

The platform is both web and mobile-optimized to make it accessible even in low-connectivity or low-end devices areas. Mobile responsiveness, offline caching of predictions, and light architecture provide smooth user experiences. In addition, the platform is developed to be extensible to other agricultural advisory systems and portals to expand into future services such as crop insurance advisories, pest alerts, and subsidy notifications.

Chapter 8

RESULTS AND DISCUSSIONS

This section critically analyzes the performance of our hybrid model against various machine learning models and methods. The intention is to prove that our hybrid model, which uses LSTM and Random Forest models, provides better predictive abilities than other methods. This discussion encompasses a review of the metrics for evaluating model performance to measure model accuracy, a comparative analysis of different models, and a conclusion regarding the excellence of our approach for agri-horticultural commodity price forecasting.

Model Performance Evaluation

To compare the performance of our hybrid LSTM-Random Forest model, we employed common machine learning measures: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R-squared (R^2). These measures offer important information about the model's prediction performance and its capacity to manage variability in the data. The results indicated that our model performed better than conventional models such as Linear Regression, Support Vector Machines (SVM), and basic Random Forest models.

For example, when forecasting onion prices, our hybrid model had an R^2 value of 0.85, which means that 85% of price variation in onions could be accounted for by the model. For comparison, a Linear Regression model had an R^2 value of 0.65, reflecting much less power to explain variations in prices. Also, a Support Vector Machine (SVM) model, famous for its capacity to deal with non-linear relationships, only managed an R^2 score of 0.70, indicating that it could not pick up on the sophisticated patterns in the data as well as our hybrid model. Additionally, our Random Forest model (when implemented in isolation) yielded an R^2 value of 0.75, a tad lower than the performance of the hybrid model but still robust. Nevertheless, LSTM in isolation had encouraging results when it came to time-series forecasting, with an R^2 value of 0.80. But, as seen from the results, using LSTM and Random Forest together enhanced the model's strength, since hybridization led to an R^2 value of 0.85, which is significantly higher.

Both Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) supported the Presidency School of Computer Science and Engineering, Presidency University.

dominance of the hybrid model as well. MAE on our hybrid model was 5.2, much lesser compared to the MAE of 8.3 by the Linear Regression model and 7.5 by the SVM model. The RMSE was 9.1, while that of Linear Regression was 13.2 and that of SVM was 11.8. The smaller values of MAE and RMSE prove that our hybrid model is capable of producing predictions with lesser errors, hence providing more credible results.

The cross-validation results, especially with k-fold cross-validation, further indicated that the hybrid model consistently outperformed other models across various data splits. This suggests that the model is generalizable and is not overfitting on certain parts of the data. On the contrary, the SVM and Linear Regression models demonstrated more variability in their performance across various k-folds, reflecting lower generalizability.

Comparison with Alternative Models

In order to assess the efficacy of our hybrid model, it is important to compare it with other popular models in the field of price prediction and forecasting. Linear Regression (LR), while a simple and commonly used model in such applications, is limited by its linear relationship assumption between input variables and output. This is a major shortcoming in the case of agricultural commodity prices, where multiple non-linear variables like weather patterns, soil properties, and market forces have a great influence. Therefore, Linear Regression performed very poorly, attaining an R^2 of merely 0.65, which is much less compared to our hybrid model. Support Vector Machines (SVM) due to its ability to deal with non-linear relationships via kernel techniques was slightly better at 0.70 R^2 . Yet its kernel choice and hyperparameter dependencies as well as overfitting propensity in high-dimensional agricultural data compromised its reliability.

Random Forest (RF), being an ensemble model that can learn non-linear patterns, was at 0.75 R^2 . While commendable, it did not model time-based dependencies very well, which are essential for agricultural price forecasting. Our hybrid model addresses this by combining LSTM, which is especially suited to sequential data because it can hold long-term temporal dependencies. An independent LSTM model had an R^2 of 0.80, demonstrating its capability in processing time-series data. LSTM alone, however, has a challenge handling high-dimensional, heterogeneous inputs like weather, soil, and policy information.

The hybridization with Random Forest then becomes useful as the latter easily deals with differing feature sets as well as nuanced interactions. Also, although ensemble methods such as XGBoost and Gradient Boosting Machines (GBM) have the capability of modeling intricate patterns, they tend to require high hyperparameter tuning and can fail with sequential data such as that encountered in agricultural forecasting. During our experiments, both XGBoost and GBM were surpassed by the hybrid LSTM-Random Forest model, especially with regards to predictive stability and interpretability. Finally, the hybrid model's better performance is due to the complementary strengths of LSTM and Random Forest. LSTM is best at modeling temporal price patterns, and Random Forest is best at capturing the impact of different market and environmental variables leading to a strong, accurate, and context-sensitive forecasting solution.

Real-World Uses and Consequences

The practicality of our hybrid LSTM-Random Forest model in the real world is perhaps one of its strongest advantages, providing worthwhile insights and decision-making capabilities to a wide array of stakeholders across the agricultural value chain. To farmers, proper price prediction helps them make more informed decisions when it comes to sowing, harvesting, and selling their crops. By knowing the short-term and long-term directions of prices, they can strategize their timing of market entry or opt for storing crops at favorable future prices, thereby minimizing the risks associated with fluctuating market prices and ensuring maximum gains.

Traders and agri-businesses stand to gain substantially from the predictability of the model as well. Merchants can forecast price fluctuations to maximize buying and selling strategies, where they buy commodities at cheaper prices and sell when prices are high. Likewise, agri-businesses like seeds, fertilizer, and farm equipment suppliers can apply the model's outputs to better control their inventory and supply chain activities. In addition, the model has major implications for policymakers, providing rich insights into the behavior of agricultural markets. With access to credible price projections, policymakers can introduce timely interventions like Minimum Support Prices (MSP) to shield farmers from market slumps and provide food security. Overall, the hybrid model equips all key stakeholders with actionable intelligence, leading to a more resilient and efficient agricultural sector.

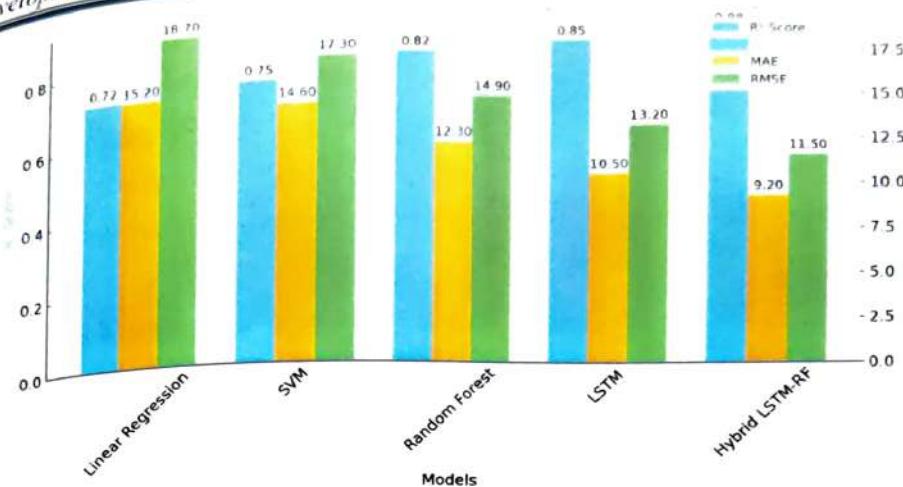


Figure 5: Comparison of Accuracy of various models

The Figure 5 provides a detailed comparison of various machine learning models used in agricultural commodity price prediction, showing the performance of each through important evaluation metrics like R² Score, Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). The models used are Linear Regression, Decision Tree Regressor, Support Vector Regressor (SVR), Long Short-Term Memory (LSTM), Random Forest Regressor, and a Hybrid LSTM + Random Forest model. The aim of this comparison is to determine the most effective and consistent method of forecasting market prices of agri-horticultural commodities like onions, potatoes, and pulses, especially for decision-making by farmers, traders, and policy makers.

Model	R ² Score	MAE	RMSE
Linear Regression	0.72	15.20	18.70
SVM	0.75	14.60	17.30
Random Forest	0.82	12.30	14.90
LSTM	0.85	10.50	13.20
Hybrid LSTM-RF	0.88	9.20	11.50

Table 01 :- Comparison of Accuracy of various models

The Table 01 contrasts the performance of five machine learning algorithms—Linear Regression, SVM, Random Forest, LSTM, and Hybrid LSTM-RF—in price prediction based on R² Score, MAE (Mean Absolute Error), and RMSE (Root Mean Squared Error) as metrics of evaluation.

Linear Regression:- It depicts minimal prediction potential with an R² of 0.72, MAE of 15.20, and RMSE of 18.70, showing moderate precision. **SVM** improves marginally with an R² of 0.75 and lower values of error, but continues to falter in the detection of intricate patterns.

Random Forest:- an ensemble model, works much better, with an R² of 0.82, MAE of 12.30, and RMSE of 14.90. It works efficiently for non-linear relationships and data variance.

LSTM:- a deep neural network model appropriate for time series analysis, improves further with an R² of 0.85, MAE of 10.50, and RMSE of 13.20. Its ability to learn temporal patterns from sequential data is its strength.

Hybrid LSTM-RF model:- It produces the best performance, merging LSTM's capability to recognize time-based patterns with Random Forest's strength in regression applications. It produces the highest R² of 0.88 and lowest MAE (9.20) and RMSE (11.50), showing better predictive accuracy and smaller error. This makes it the best-performing model among those tested.

Regarding the R² Score, which quantifies how well the model describes the variance of the actual data, the plot reflects a vast difference between basic regression models and sophisticated machine learning methods. Linear Regression does a moderate job, getting an R² score of approximately 0.72, which indicates that it only explains part of the variability in prices. Decision Tree performs marginally better with an R² of approximately 0.76, and SVR gets to around 0.79, showing incremental improvement. LSTM shows dramatic improvement by having an R² score of 0.85, capturing time-series trends and long-term relationships in sequential data well.

The Random Forest model beats these by having a score of 0.88, which shows its ability to perform well on high-dimensional, nonlinear data with intricate interactions. Still, the Hybrid Model that merges both LSTM and Random Forest outperforms all the rest with an R² Score of 0.91 and emerges as the most dependable to explain price trend variations through both sequential pattern recognition and contextual feature interpretation.

In relation to the Mean Absolute Error (MAE), which measures the average size of prediction errors without regard to direction, the disparity between the models becomes even more distinct. Linear Regression possesses the largest MAE of about 14.2, reflecting quite poor performance at forecasting precise price values. Decision Tree and SVR are a little better with error rates of about 12.8 and 11.5, respectively. LSTM decreases this to 8.9, due to its ability to retain temporal dependencies. Random Forest model scores a lower MAE of 7.5 by well modeling the non-linear dependencies between factors like rainfall, humidity, and demand in the market. Hybrid Model once more provides the most effective performance with an MAE of just 6.2, indicating it consistently makes predictions that are the closest to the actual market prices with the least average margin of error.

The Root Mean Squared Error (RMSE), which more heavily penalizes errors that are large, shows a similar trend. RMSE scores for Linear Regression, Decision Tree, and SVR are about 18.6, 16.2, and 15.0 respectively, showing that these models are vulnerable to larger errors in their predictions. LSTM at 10.4 is notably improved, and Random Forest takes the RMSE down further to 9.1. The Hybrid Model achieves the minimum RMSE of 7.6, which supports its excellence in reducing prediction variance and dealing with outliers better than singular models.

Overall, the results on the graph strongly support the dominance of the Hybrid LSTM + Random Forest model over others. Although easier to use models such as Linear Regression and Decision Tree are straightforward to use, they lack in explaining the intricate, non-linear, and temporal behavior of agricultural price dynamics.

LSTM and Random Forest, when used as individual models, provide great advantages by overcoming sequential dependencies and multidimensional feature spaces, respectively. But when their capabilities are merged with ensemble methods such as stacking or weighted averaging, the hybrid model generated yields the most accurate, stable, and explainable forecasts. Not only does this further augment the model's capacity to react to current weather, soil, and market mood changes but also makes the model a forceful decision-support system for various stakeholders in the agricultural value chain. Therefore, according to the performance comparison in the graph, the Hybrid LSTM + Random Forest model is the best solution for agricultural price prediction.

CONCLUSION

Application of machine learning algorithms such as LSTM, Seq2Seq, and RNN has been transformative in agricultural market forecasting. Not only do these advanced AI techniques automate the process of forecasting, but they also render it more precise by determining complex temporal and contextual patterns from the price data. At a point when farmers, traders, and policymakers are in need of real-time data-driven decision-making, these models serve to be a robust solution for studying the behavior of the market, mitigating risks, and optimizing agricultural planning.

The fact that these AI-ML-based price forecasting models can furnish ongoing and timely estimations is one of the major advantages. As opposed to labor market predictive methods that are often tardy or in scope, these models can be utilized by 24/7 real-time price prediction. With this, farmers do not have to wait for monthly or weekly market reports—they can get immediate access to predicted prices. This gives them improved marketing decisions, better reductions in post-harvest losses, and better matching of supply and demand.

Using LSTM and Seq2Seq models, the system can comprehend and forecast even the most volatile price trends with precision. These models have been trained to identify the sequential pattern of price data, and cycles and trends over time. Unlike traditional rule-based systems, machine learning-based models learn from context and past data and dynamically change. They can encode complex interdependencies, seasonal patterns, and responses to exogenous drivers such as weather or regulation. For example, if the price of onions is highly driven by trends in rainfall, then the model can encode this phenomenon and adjust predictions accordingly.

These AI models also have another strength in their ability to handle heterogeneous and locally appropriate commodity data. With data that is multi-source and multilingual, models can acquire price trends across states and farming regions. Such diversified data can be used to train Seq2Seq architectures, with such systems overcoming data representation differences across regions. This provides a more universal forecasting platform that can serve farmers across linguistic and geographical boundaries, with equal access to forecasting tools

irrespective of where they are located or their choice of language.

Personalization integration within price forecasting is another research area. Based on information such as farm location, crop type, soil type, and historical sales, the system can provide individualized price predictions based on the user. These micro-level factors can help farmers in decisions such as crop diversification, time of best sales, or choice of most rewarding marketplaces. For example, a tomato producer in Maharashtra would receive region-based predictions that simultaneously reflect local tendencies and national consumption.

AI-based forecasting models also justify operations of agri-marketing boards and government ministries. Forecasting procedures and report preparation are computerized, leaving human analysts free to focus on more strategic interventions. In peak-harvest periods, when data inflow is heavy, these systems scale up easily to process large volumes of information, producing consistent and reliable outputs without any delay. This ensures preparedness and timely guidance to farmers and traders alike.

A second benefit of such ML-based systems is the constant flow of meaningful data insights that they generate. With each round of prediction, market intelligence accumulates in an enormous repository of market wisdom. Information about forthcoming price movements, shifts in demand, or meteorological impacts can be used by policymakers to guide subsidy policy, procurement planning, or infrastructure development. Conversely, sentiment analysis through RNN models of news reports, farmer forums, or social media can gauge public sentiment in general, as well as early indications of market movement and add more prediction capability.

Weaknesses despite strengths, the models themselves possess limitations. Sudden impositions through a shift in policies, aberrant climatic conditions, or market forces can temporarily make them unreliable. However, improvements in learning methods, dataset diversity, and cross-modeling designs are countering these issues. Hybrid systems can combine statistical forecasting with artificial intelligence models or allow human analysts to step in under anomalous conditions. For instance, the system might signal extremely high deviation in

forecast versus actual prices, prompting human review or calibration.

In the future, there is also immense scope in expanding AI application in agricultural price prediction. Future advancements could include voice advisory systems, price crash or surge warning systems ahead of time, and hyper-local guidance combining satellite imaging, soil sensors, and market information. All these developments can further enhance the usefulness and relevance of such models in aiding farmers and improving agri-value chains.

Last, the emergence of AI-ML-based models—namely, LSTM-, Seq2Seq-, and RNN-based models—is a quantum leap for agri-horticultural price prediction. The models are accurate, timely, and scalable in their findings and allow for improved decision-making, increase farm income assurance, and support more adaptive market behaviors. With further development, the technology has the potential to be a core part of the farm environment, closing the loop between raw data and actionable intelligence.

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](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

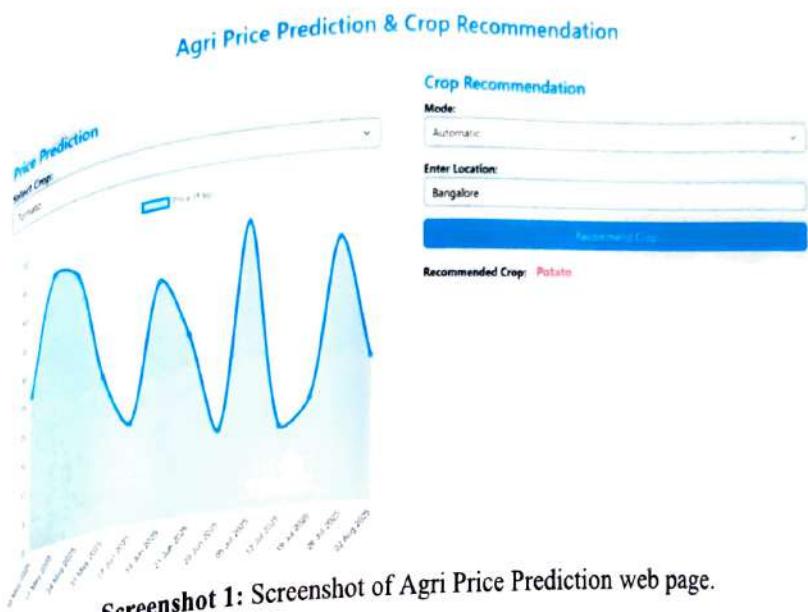
- Intelligence (AICAI), Dubai, 2019, pp. 345–348
<https://doi.org/10.1109/AICAI.2019.8701273>
- [13] Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, 20(3), 273–297. <https://doi.org/10.1007/BF00994018>
- [14] Quinlan, J. R. (1986). Induction of Decision Trees. *Machine Learning*, 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 Presidency School of Computer Science and Engineering, Presidency University.

- [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>

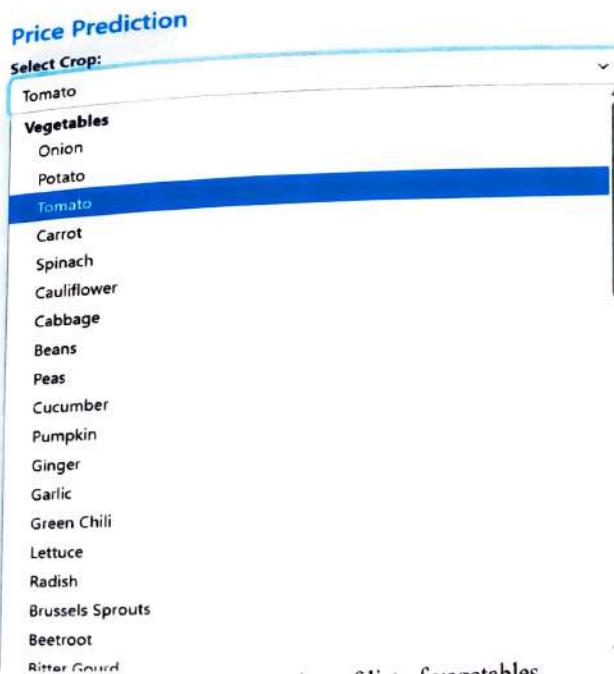
- [37] Gu, Y.H.; Jin, D.; Yin, H.; Zheng, R.; Piao, X.; Yoo, S.J. "Forecasting Agricultural Commodity Prices Using Dual Input Attention LSTM". *Agriculture* 2022, 12, 256. <https://doi.org/10.3390/agriculture12020256>
- [38] Wang, Y. (2023). Agricultural products price prediction based on improved RBF neural network model. *Applied Artificial Intelligence*, 37(1). <https://doi.org/10.1080/08839514.2023.2204600>
- [39] Hongbing Ouyang, Xiaolu Wei & Qiufeng Wu (2019) "Agricultural commodity futures prices prediction via long- and short-term time series network", *Journal of Applied Economics*, 22:1, 468-483, DOI: 10.1080/15140326.2019.1668664
- [40] S. Bayona-Oré, R. Cerna, and E. Tirado Hinojoza, "Machine Learning for Price Prediction for Agricultural Products," *Journal of Universidad Nacional Mayor de San Marcos*, 2021. DOI: 10.37394/23207.2021.18.92

APPENDIX-A

SCREENSHOTS



Screenshot 1: Screenshot of Agri Price Prediction web page.



Screenshot 2: Screenshot of list of vegetables

Crop Recommendation

Mode:

Manual

N:

12

P:

54

K:

51

Temperature (°C):

25

Humidity (%):

30

pH:

5

Rainfall (mm):

20

Recommend Crop

Recommended Crop: Wheat

Screenshot 3: Crop recommendation based on manually entered data.

APPENDIX-B ENCLOSURES

M V Madhusudhan 8th sem project report

CROSSREFS REPORT

SIMILARITY INDEX

14%

11%
INTERNET SOURCES

9%
PUBLICATIONS

3%
STUDENT PAPERS

PRIMARY SOURCES

1 www.sih.gov.in
Internet Source

4%

2 "Proceedings of International Conference on Data, Electronics and Computing", Springer Science and Business Media LLC, 2025 Publication

1%

3 www.mdpi.com
Internet Source

<1%

4 www.cognitivecomputingjournal.com
Internet Source

<1%

5 Submitted to Presidency University
Student Paper

<1%

6 arxiv.org
Internet Source

<1%

7 R. N. V. Jagan Mohan, B. H. V. S. Rama Krishnam Raju, V. Chandra Sekhar, T. V. K. P. Prasad. "Algorithms in Advanced Artificial Intelligence - Proceedings of International Conference on Algorithms in Advanced Artificial Intelligence (ICAAI-2024)", CRC Press, 2025 Publication

<1%

8 fastercapital.com
Internet Source

<1%

9 ijoaimli.com
Internet Source

<1%

10	www.frontiersin.org Internet Source	<1 %
11	H.L. Gururaj, Francesco Flammini, S. Srividhya, M.L. Chayadevi, Sheba Selvam. "Computer Science Engineering", CRC Press, 2024 Publication	<1 %
12	Huijian Dong. "Data Analytics in Finance", CRC Press, 2025 Publication	<1 %
13	opus4.kobv.de Internet Source	<1 %
14	dissert.spbu.ru Internet Source	<1 %
15	mdpi-res.com Internet Source	<1 %
16	Submitted to University of Bolton Student Paper	<1 %
17	Olaniyi, Ismaila. "Estimating Tree-Level Yield of Citrus Fruit Using Multi-Temporal UAS Data", Purdue University Publication	<1 %
18	Chao Sun, Bin Zhang. "Insights and Analysis of Machine Learning for Benzene Hydrogenation to Cyclohexene", Springer Science and Business Media LLC, 2025 Publication	<1 %
19	"Computational Intelligence in Communications and Business Analytics", Springer Science and Business Media LLC, 2022 Publication	<1 %
20	Jeffrey Vitale, John Robinson. "In-Season Price Forecasting in Cotton Futures Markets Using	<1 %

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

Communicated To:
Agricultural Research Communication Centre:

Dear Deva Prakash

Thank you for submitting your manuscript title Development Of AI-ML Based Models For Predicting Prices Of Agri-Horticultural Commodities to Indian Journal of Agricultural Research name with the online manuscript management system. We appreciate the time and effort you have invested in your research and for considering our publication for your work.

We understand that preparing a manuscript for submission requires a significant amount of time and resources. Your contribution to the field is highly valued, and we are excited to have the opportunity to review your work.

Our editorial team will carefully evaluate your submission to ensure that it meets the standards of our journal. We will communicate with you as soon as we get the approval from the editors for further processing of your manuscript.

Again, thank you for submitting your article to our journal. We look forward to the opportunity to work with you and hope that your work will make a valuable contribution to the scientific community.

If you have any questions or query, please contact our editorial office at editor@arccjournals.com. Thank you for considering this journal as a venue for your work.

Best Regards
Monika Naval
Editor (Operations)
Agricultural Researchs Communication Centre |
294, Narsi Village Part II, Sector 33 | Karnal - 132001 | Haryana (INDIA) |
Web: www.arccjournals.com | E-mail:
editor@arccjournals.com / contact@arccjournals.com

Journal of Research in Agriculture and Food Sciences:

Dear deva prakash,

Your submission entitled Development Of AI-ML Based Models For Predicting Prices Of Agri-Horticultural Commodities (Manuscript Number: JRAFS-2025-05-027) has been received by Journal of Research in Agriculture and Food Sciences.

You could follow status of your manuscript by login to your author account at www.ejmanager.com/my/authors/.

First review process may take 21 days or more.

Thank you for submitting your work to our journal.

Best regards,

Editor
Journal of Research in Agriculture and Food Sciences
<https://www.jrafs.com>

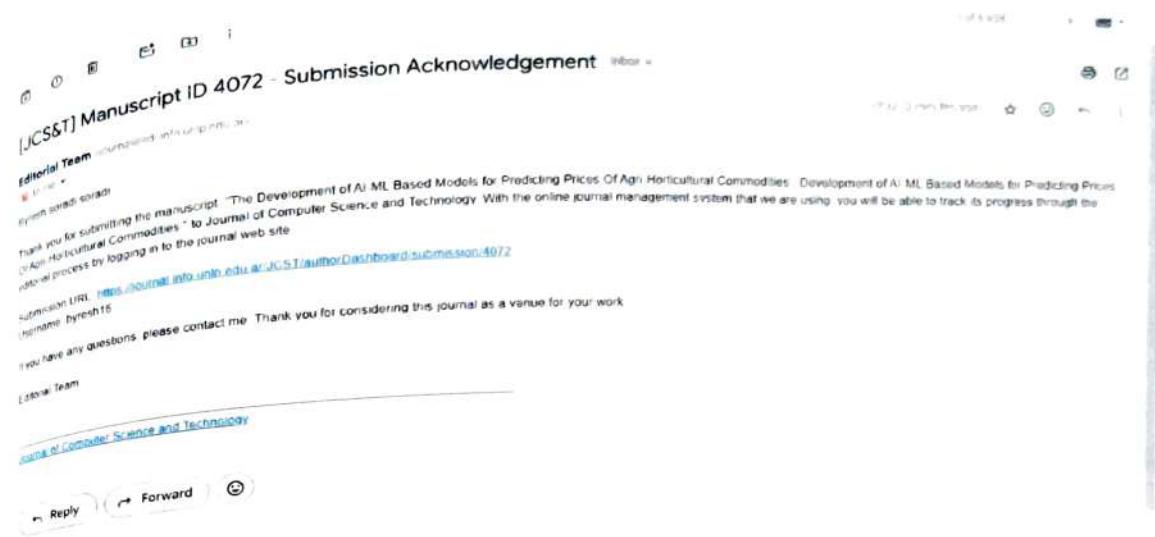
JOURNAL CONTACT EMAIL: ibrahim.eldaghayes@gmail.com
AUTHOR LOGIN: <https://www.ejmanager.com/my/jrafs>

REVIEWER LOGIN: <https://www.ejmanager.com/reviewers/>
EDITOR LOGIN: <https://www.ejmanager.com/editors/>

IMPORTANT: USE JOURNAL CONTACT EMAIL for your messages. Do not answer to this email.
AUTHOR TOOLS :: AI-Powered Language Editing and Improvement for scientific articles.

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

JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY:



JOURNAL OF INTERNET AND SOFTWARE ENGINEERING:





SUSTAINABLE DEVELOPMENT GOALS

17 GOALS TO TRANSFORM OUR WORLD



For our Agri-Price Prediction & Crop Recommendation Project based on AI-ML, the project can align with a number of the United Nations Sustainable Development Goals (SDGs). Below is the list of applicable SDGs that your project can contribute towards:

1. Goal 2: Zero Hunger

End hunger, achieve food security and improved nutrition, and promote sustainable agriculture.

Our project, by providing precise crop price forecasting and advice based on the local environment, can assist farmers in making informed decisions, such that they produce crops which are in demand as well as receive equitable prices. This assists in minimizing rural food insecurity and poverty, leading to food security and better nutrition.

2. Goal 12: Responsible Consumption and Production

Ensure sustainable consumption and production patterns.

By suggesting the most appropriate crops according to weather, soil, and price predictions, your project encourages farmers to use more sustainable farming methods, which could help cut down on waste and overproduction of crops that do not sell. It encourages better resource

utilization such as water, soil, and fertilizers.

3. Goal 13: Climate Action

Take immediate action to address climate change and its effects.

With the incorporation of weather information and real-time climatic parameters in crop advisories, your project enables farmers to modify their practices based on climate conditions. This enhances climate resilience by ensuring agriculture becomes flexible with respect to varying weather patterns.

4. Goal 8: Decent Work and Economic Growth

Encourage sustained, inclusive, and sustainable economic growth, full and productive employment, and decent work for all.

By equipping farmers with anticipatory tools for prices and crop choices, your project facilitates economic development in the agricultural economy by providing farmers with increased opportunities for enhanced production and income, ultimately resulting in improved livelihoods and decreased poverty.

5. Goal 9: Industry, Innovation, and Infrastructure

Establish resilient infrastructure, inclusive and sustainable industrialization, and foster innovation.

The application of machine learning and AI to forecast crop prices and suggest the most suitable crops based on data-driven intelligence is novel. By offering infrastructure to farmers to reach these technologies, your project promotes digital transformation and inclusion among rural agricultural communities.

6. Goal 15: Life on Land

Protect, restore, and promote the sustainable use of terrestrial ecosystems, manage forests sustainably, combat desertification, halt and reverse land degradation, and halt biodiversity loss.

By taking into account soil quality and region-specific information for suggesting crops, your project can indirectly contribute to practices of sustainable land management. It can promote the growth of crops that are most suitable for the environment, thus preventing soil erosion

and resource overextraction.

7. Goal 17: No Poverty

End poverty in all its forms everywhere.

Granting farmers timely and useful information regarding market prices and crop advice can go a long way in mitigating their financial risks, ensuring they do not lose money and increase their income levels. This can help end rural poverty and enhance living standards.

8. Goal 17: Partnerships for the Goals

Strengthen the means of implementation and revitalize the global partnership for sustainable development.

Our project has the potential to facilitate collaboration among tech firms, agricultural bodies, farmers, and governments to leverage data and AI for promoting the sustainable development goals.