GPM: Applying a hybrid model to predict pricing for garlic and farmer's profitability

A PROJECT REPORT

Submitted by,

Mr. Shashank J K – 20211CSD0147
Mr. Parth – 20211CSD0041
Mr. Sagar M – 20221LSD0001
Mr. Teja Reddy M – 20221LSD0003

Under the guidance of,

Dr. Manjunath K V,
Associate Professor

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING (DATA SCIENCE)

At

GAIN MORE KNOWLEDGE
REACH GREATER HEIGHTS

PRESIDENCY UNIVERSITY

BENGALURU

JANUARY 2025

PRESIDENCY UNIVERSITY

SCHOOL OF COMPUTER SCIENCE ENGINEERING

CERTIFICATE

This is to certify that the Project report **“GPM: Applying a hybrid model to predict pricing for garlic and farmer’s profitability”** being submitted by “Shashank J K, Parth, Sagar M and Teja Reddy M” bearing roll number(s) “20211CSD0147, 20211CSD0041, 20221LSD0001, 20221LSD0003” in partial fulfillment of requirement for the award of degree of Bachelor of Technology in Computer Science and Engineering(Data Science)is a bonafide work carried out under my supervision.

Dr. MANJUNATH K V
Associate Professor
School of CSE&IS
Presidency University

Dr. SAIRA BANU ATHAM
Professor & HoD
School of CSE&IS
Presidency University

Dr. MYDHILI NAIR
Associate Dean
School of CSE
Presidency University

Dr. SAMEERUDDIN KHAN
Pro-VC School of Engineering
Dean -School of CSE&IS
Presidency University

PRESIDENCY UNIVERSITY

SCHOOL OF COMPUTER SCIENCE ENGINEERING

DECLARATION

We hereby declare that the work, which is being presented in the project report entitled **GPM: Applying a hybrid model to predict pricing for garlic and farmer's profitability** 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. Manjunath K V Associate Professor, Presidency School of Computer Science & Engineering.**

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

NAME	ROLL NUMBER	SIGNATURE
SHASHANK J K	20211CSD0147	
PARTH	20211CSD0041	
SAGAR M	20221LSD0001	
TEJA REDDY M	20221LSD0003	

ACKNOWLEDGEMENT

First of all, we 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, School of Engineering and Dean, Presidency School of Computer Science Engineering & Information Science, for getting us permission to undergo the project.

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

We are greatly indebted to our guide **Associate Professor Dr. Manjunath K V**, Presidency School of Computer Science 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 project work.

We would like to convey our gratitude and heartfelt thanks to the PIP2001 Capstone Project Coordinators **Dr. Sampath A K and Mr. Md Zia Ur Rahman,** department Project Coordinators **Dr. H M 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.

SHASHANK J K (20211CSD0147)
PARTH (20211CSD0041)
SAGAR M (20221LSD0001)
TEJA REDDY M (20221LSD0003)

ABSTRACT

Project Objective

This project aims to build an intelligent system for forecasting garlic prices in India by utilizing a hybrid approach that blends Random Forest and Long Short-Term Memory (LSTM) models. The objective is to enhance prediction accuracy by using Random Forest to identify important features and LSTM to learn temporal patterns in the data. By analyzing historical pricing information and market dynamics, the system delivers dependable forecasts that can support farmers, decision-makers, and market researchers in planning and decision-making.

Background

Forecasting agricultural prices in India presents considerable challenges due to the influence of diverse factors, including seasonal trends, supply-demand fluctuations, regional market differences, and external variables like changing weather conditions. Garlic, in particular, is a crop of both economic and dietary value, widely cultivated across various Indian states. However, its market prices are known for sharp and unpredictable changes. Farmers frequently struggle with the absence of timely and precise price forecasts, making it difficult to effectively schedule planting and harvesting. This often results in market oversaturation when many producers sell simultaneously or missed opportunities when they delay sales. Small and marginal farmers, in particular, suffer due to limited access to forecasting tools and market intelligence. These difficulties underscore the urgent need for a smart price prediction system that integrates historical trends, market behavior, and climate-related data to help ensure better planning and financial stability for farmers.

Advancements in the Price Prediction System for the Indian Agricultural Market

This research involved testing multiple standalone models—such as ARIMA, SVR, Prophet, XGBoost, and LSTM—to assess their forecasting performance. Among these, XGBoost delivered the most accurate results individually. However, due to the inherent drawbacks of single-model frameworks, a hybrid method was developed. The system combines Random Forest to extract key predictive features and LSTM to effectively capture temporal patterns in the data. To improve result interpretation, the study also incorporates district-level price visualizations and geospatial analysis. Looking ahead, the model could be enhanced by incorporating socio-economic indicators, enabling real-time data integration, and deploying more advanced deep learning techniques to boost forecast accuracy further.

LIST OF TABLES

LIST OF FIGURES

TABLE OF CONTENTS

CHAPTER-1

INTRODUCTION

Forecasting the prices of agricultural commodities has gained critical relevance in today's unpredictable economic landscape, especially in countries like India, where agriculture sustains a large segment of the population. Garlic stands out among these crops due to its extensive use in cooking, health benefits, strong market demand, and export potential. Despite its importance, garlic pricing remains highly erratic, shaped by various interconnected factors including seasonal harvesting patterns, unpredictable weather, logistical inefficiencies, and evolving market demand. These irregularities not only impact farmers' earnings but also pose challenges to the overall agricultural economy, affecting planning, storage infrastructure, and supply chain logistics.

Weather plays a vital role in determining garlic prices in India. The crop thrives under specific climatic conditions, with sowing generally occurring from October to December and harvesting taking place between March and May, varying by region. Adverse weather events such as unexpected rainfall, heatwaves, drought, or frost during key growth stages can significantly impact yields, often leading to sudden drops in supply and subsequent price surges. Areas lacking robust irrigation systems are especially susceptible to these weather-related risks, making accurate prediction more difficult. Additionally, the wide regional variation in garlic production—each with its own climatic and agricultural nuances—further complicates price forecasting by creating localized market dynamics.

Beyond climatic conditions, several other elements contribute to fluctuations in garlic prices—these include changes in consumer demand, the availability and capacity of storage infrastructure, labor shortages, pest infestations, and policy measures such as export restrictions or modifications to the Minimum Support Price (MSP). Traditional forecasting techniques like ARIMA (AutoRegressive Integrated Moving Average) and SVR (Support Vector Regression) have been commonly applied in agricultural economics. Although these models are somewhat effective at capturing short-term or linear trends, they often struggle with the nonlinear, irregular, and seasonal variations typical of agricultural markets. Moreover, their reliance on stationary data makes them less ideal for modeling the inherently dynamic nature of agricultural pricing.

In addition to modeling limitations, issues such as incomplete datasets, irregular data recording practices, and the absence of up-to-date market information pose significant barriers to accurate forecasting. While advanced deep learning models like LSTM are more adept at recognizing sequential trends, they require extensive, high-quality data and are often criticized for their lack of transparency. Their "black box" nature makes it difficult for users—especially farmers and policymakers with limited technical backgrounds—to understand or trust the predictions.

This study tackles the identified challenges by introducing a hybrid forecasting model that integrates the capabilities of Random Forest and Long Short-Term Memory (LSTM) networks. Random Forest, a robust ensemble learning technique, is particularly effective for handling complex datasets with many features and excels at detecting intricate, nonlinear relationships among variables like climatic factors, demand cycles, and seasonal patterns. In contrast, LSTM—an advanced form of recurrent neural network—is well-suited for learning from time-series data due to its ability to capture long-range dependencies. By combining these two approaches, the system uses Random Forest to isolate the most influential variables while LSTM learns temporal behaviors, enhancing predictive accuracy and model adaptability across diverse market conditions.

The model is further refined by integrating external factors such as weather conditions (temperature and rainfall) and cultural events (festival seasons), which play a significant role in influencing both the supply and demand of garlic. Future improvements might incorporate socio-economic data, including minimum support price (MSP) policies, fuel price fluctuations (affecting transportation costs), and farmer sentiment. The ultimate aim is to develop a user-friendly, interpretable forecasting tool tailored for farmers. This tool, ideally available as a mobile or web app, would offer region-specific price predictions in local languages.

In conclusion, this project not only advances the academic field of AI-driven forecasting but also addresses practical challenges faced by farmers. By enhancing the precision, transparency, and accessibility of garlic price predictions, the hybrid model aims to help farmers make informed decisions regarding harvest timing, storage, and sales. This, in turn, can lead to improved price outcomes, decreased market manipulation, and a more robust agricultural sector.

CHAPTER-2
LITERATURE SURVEY

2.1 Introduction

Forecasting garlic prices accurately is an essential area of research due to the economic importance of garlic as a key agricultural product in India. Price volatility, influenced by factors such as weather conditions, pest outbreaks, transportation challenges, demand fluctuations, and government regulations, often impacts both farmer incomes and consumer prices. As a result, effective forecasting helps stabilize farm revenues and supports government agencies and market analysts in taking proactive steps to prevent severe market disruptions.

Seasonality is a key factor in garlic cultivation. Garlic is generally planted in the Rabi season (October to December) and harvested between March and May. During this period, price volatility is common due to factors such as excess production, storage challenges, or unexpected weather events like unseasonal rainfall or extreme temperatures. Additionally, India's agricultural market faces challenges like fragmented supply chains and limited access to timely market data, which heightens the risks for smallholder farmers. Therefore, a reliable and adaptable forecasting system that incorporates these external factors and performs consistently across various districts is of great importance.

Price prediction faces several challenges, including inconsistent or missing data across regions, a lack of real-time updates, absence of standardized quality assessments, and many models' inability to quickly adjust to market disruptions. Additionally, models that rely solely on historical prices often overlook underlying factors, while complex models like deep learning require significant computational resources, making them difficult to implement in rural areas. To tackle these challenges, this project introduces a hybrid model that merges the feature selection and non-linear modeling strengths of Random Forest with the time-series dependency capturing capabilities of LSTM. This approach strikes a balance between interpretability, accuracy, and adaptability, providing a feasible and scalable solution for practical use. The proposed model not only outperforms traditional statistical and machine learning approaches in accuracy but also paves the way for AI-driven, digitized agriculture in India.

2.2 Traditional Time-Series Models

ARIMA (Auto-Regressive Integrated Moving Average) has been a popular choice for time-series forecasting because of its simplicity and ease of interpretation. It is effective at identifying autocorrelations and seasonal trends within short time frames. However, its key drawback is the assumption of linearity and stationarity, which limits its ability to handle sudden price fluctuations caused by factors like policy shifts, climate disruptions, or supply chain issues.

SVR (Support Vector Regression) offers greater flexibility than ARIMA, especially in modeling non-linear relationships between input variables and target outcomes. However, it requires careful tuning of hyperparameters and the selection of appropriate kernels, which can be challenging. Additionally, SVR can be computationally intensive with large datasets, and its performance may decline when faced with sudden market shocks.

Prophet, created by Facebook, is specifically built to address seasonality and gaps in time-series data. It breaks down the time series into components such as trend, seasonality, and holidays, making it a good fit for agricultural data with distinct seasonal patterns. However, it is less responsive to external factors like weather conditions and economic indicators, which limits its ability to adapt to specific contexts.

2.3 Machine Learning Approaches

Random Forest, an ensemble technique built on decision trees, is commonly applied in agricultural data science due to its robustness and capacity to capture non-linear relationships. It is particularly effective in feature selection and handles inconsistent data quality well. However, Random Forest does not have temporal memory, limiting its ability to model time-dependent sequences unless integrated with sequential models.

XGBoost (Extreme Gradient Boosting) improves model performance by reducing bias and variance using boosting methods. Its main advantage is scalability and its ability to handle sparse data, a common characteristic in agricultural datasets. While it has outperformed linear models in various studies, it may struggle to capture the sequential dependencies present in time-series data, such as garlic price trends.

Both Random Forest and XGBoost are typically more interpretable than deep learning models, making them valuable tools for policymakers and stakeholders who require clear and transparent decision-making processes.

2.4 Deep Learning Approaches

Deep learning models, particularly recurrent neural networks (RNNs), have demonstrated strong performance in long-term price forecasting. LSTM (Long Short-Term Memory) networks, specifically, are highly effective at capturing sequential patterns in data. Research indicates that LSTM often outperforms traditional models such as ARIMA and SVR in identifying long-term trends and variations. However, LSTM models demand large datasets and significant computational resources, which can be a limitation for small-scale applications..

Convolutional Neural Networks (CNNs) have been explored for predicting agricultural prices due to their ability to identify localized trends in price changes. However, their inability to capture long-term dependencies typically requires the integration of CNNs with recurrent models like LSTMs to enhance performance.

Model Type	**Description**	**Hyperparameter Tuning**
ARIMA	A statistical model used for time-series forecasting	*p, d, q* (parameters)
XG Boost	A gradient boosting algorithm that improves prediction accuracy by reducing errors iteratively.	Accuracy, RMSE, R^2 Score
SVR	A machine learning algorithm that captures complex patterns in price data.	Kernel Type, C (Regularization), Epsilon
Prophet	A regression model specifically designed to account for seasonality in time-series data.	Growth Type, Changepoint Prior Scale, Seasonality Mode
LSTM	A type of recurrent neural network (RNN) designed to recognize long-term dependencies in sequential data	Number of LSTM Units, Batch Size

Table 2.1 Survey Models

CHAPTER-3
RESEARCH GAPS OF EXISTING METHODS

3.1 Inadequate Handling of Complex Data Patterns

Traditional forecasting models like ARIMA and SVR are based on assumptions of linearity, stationarity, and consistency within time-series data. While these models work well with structured datasets that have low volatility, they face challenges when used for agricultural commodities such as garlic, where prices are influenced by numerous volatile and non-linear factors. Additionally, these models cannot capture long-term historical dependencies, which limits their effectiveness in situations where trends and seasonal patterns change over the course of years rather than just weeks or months.

Even sophisticated deep learning models like CNNs, while effective at extracting local features from data sequences, lack the depth required to capture long-term trends essential for agricultural forecasting. LSTM models address this limitation by preserving memory of previous inputs, but they often demand large, clean, and well-labeled datasets for effective training. This poses a challenge in the agricultural sector, where data inconsistencies, missing values, and limited digital records are prevalent. Furthermore, the need for powerful GPUs and extended training times makes these models less suitable for rapid deployment in rural areas with limited resources.

3.2 Limited Integration of External Influences

A key limitation of many existing models is their exclusive reliance on historical price data, neglecting the influence of external factors that significantly impact market conditions. Agricultural prices, especially for garlic, are highly affected by variables such as weather irregularities (e.g., droughts, heavy rainfall), pest infestations, transportation issues, labor shortages, fuel costs, and government policies (e.g., export restrictions or minimum support prices).

Overlooking these key factors results in forecasts that may be statistically valid but practically unreliable. For example, a sudden change in garlic export tariffs or an unexpected flood in a region can cause significant deviations from historical trends. If models fail to incorporate contextual data such as satellite weather indices, rainfall and temperature predictions, policy changes, and reports on crop diseases, they are unable to adjust to real-world shifts, diminishing their relevance and accuracy.

3.3 Lack of Region-Specific and Real-Time Forecasting

Many studies concentrate on macro-level datasets at the national or state scale, producing generalized forecasting models that fail to account for regional variations. India’s varied agro-climatic zones lead to notable differences across districts in crop yields, storage facilities, soil quality, irrigation methods, and market access. A model trained with data from Madhya Pradesh, for instance, might not yield accurate predictions for farmers in Tamil Nadu due to differences in climate, harvest periods, and trading practices.

Additionally, there is a major gap in real-time forecasting capabilities. Many existing models are trained offline and updated irregularly, rendering them outdated when market conditions shift quickly. For farmers and small-scale traders, even a few days' delay in receiving insights can result in poor decisions about harvesting, storage, or sales timing. This inability to respond in real time restricts the practical usefulness of most current systems.

3.4 Data Scarcity and Quality Challenges

A key challenge in achieving accurate agricultural forecasting is the availability and quality of data. Many publicly available datasets on garlic prices are incomplete, sparsely populated, or updated infrequently, hindering the development of models that can generalize across different time periods and regions. Furthermore, data gathered at the district or market level often lacks consistency due to non-standardized data collection methods, making it more difficult to create scalable forecasting systems.

The problem is further exacerbated by the lack of integration between price data and other critical factors such as weather patterns, soil health, pest outbreaks, or logistical challenges. These data gaps prevent comprehensive model training and diminish the reliability of predictions. To achieve effective forecasting, it is essential to develop strong data pipelines that can handle missing values, filter out noise, and synchronize data from multiple sources to create clean, structured inputs for hybrid modeling approaches.

3.5 Usability and Accessibility for End-Users

Despite the increasing sophistication of predictive models, a major gap remains in making these tools accessible and practical for the primary users—farmers. Many smallholder and marginal farmers in India have limited access to advanced digital infrastructure, often relying on basic mobile phones and inconsistent internet connectivity. Consequently, even the most accurate models fail to provide value if they are not designed with usability in mind. Additionally, existing solutions are frequently presented in complex formats or are English-only, which may not suit the digital literacy or language preferences of rural communities. There is an urgent need to create lightweight, user-friendly platforms that provide insights in local languages, via mobile apps, SMS notifications, or even voice-based services. Only by addressing this usability gap can technology truly empower farmers and encourage widespread adoption.

3.6 Inadequate Hybridization of Models

While hybrid modeling approaches have become more popular in academic studies, their practical application in agriculture is still limited. Most research either focuses on traditional statistical models or standalone machine learning/deep learning methods, with little effort to combine their strengths effectively. Traditional models are appreciated for their interpretability, while machine and deep learning models excel at capturing complex, non-linear relationships. However, many proposed hybrid models tend to be overly complex, computationally demanding, or unsuitable for real-time deployment. Few models manage to balance accuracy, efficiency, and simplicity—key attributes for practical use in critical agricultural settings. Additionally, the lack of a modular structure makes it challenging to integrate hybrid models into larger decision-support systems used by governments or agricultural cooperatives.

The proposed hybrid model combining Random Forest and LSTM overcomes these challenges by integrating strong feature selection and noise resistance with the ability to learn sequential patterns and recognize temporal trends. Designed for flexibility and efficiency, the model can incorporate both historical price data and external factors like weather conditions. This structure not only boosts predictive accuracy but also improves model interpretability and prepares it for real-time deployment, providing clear advantages over traditional forecasting systems.

CHAPTER-4

PROPOSED MOTHODOLOGY

This research employs a combination of machine learning, deep learning, and hybrid models to build a comprehensive framework for real-time prediction and mapping of garlic price trends. The following steps outline the approach used for data collection, model development, and visualization to support informed decision-making by farmers and authorities.

4.1 Data Collection:

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

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

Table 4.1 Types of data

The integration of price data with meteorological variables enables the modeling framework to account for both temporal price dynamics and external climatic influences, enhancing the prediction accuracy and relevance.

4.2 Model Development:

4.2.1 Understanding Model Requirements:

This study recognizes the urgent need for a dependable and accurate model to forecast garlic prices, given the high sensitivity of agricultural markets to seasonal, economic, and environmental factors. To address this, a hybrid prediction system is proposed using Random Forest (RF) and Long Short-Term Memory (LSTM) networks. Random Forest is chosen for its robustness in handling structured, tabular data and its ability to model non-linear relationships efficiently. It performs well even with missing data and complex interactions between variables, making it suitable for diverse agricultural datasets.

LSTM, on the other hand, excels in capturing time-dependent patterns and long-term dependencies in sequential data. Its internal memory mechanism allows it to understand how past events, such as rainfall or previous price spikes, affect future prices.

While LSTM models are computationally more demanding, they offer deep insights into temporal trends that traditional models often miss. By combining the strengths of RF and LSTM, the hybrid approach aims to deliver improved forecasting accuracy, balancing interpretability, speed, and the ability to learn both short-term fluctuations and long-term patterns in garlic prices.

4.2.2 Preprocessing:

The preprocessing phase begins by loading the historical garlic price data and checking for completeness and consistency. Any missing values or duplicates are handled to ensure clean and accurate inputs. Irrelevant or non-informative features are removed to improve model efficiency and reduce noise.

The dataset is then split into training (80%) and testing (20%) sets, allowing for model evaluation on unseen data. All numerical features are normalized using MinMaxScaler to scale the values between 0 and 1, which helps in stabilizing and accelerating model training. Finally, the data is reshaped into a 3D format—comprising samples, timesteps, and features—to meet the input requirements of the LSTM model, enabling it to learn from temporal sequences effectively.

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

Table 4.2 Steps of Data Preprocessing

4.2.3 Implementing the Random Forest Model:

The Random Forest part of the model utilizes a RandomForestRegressor with 100 decision trees (n_estimators=100), a choice made to strike a balance between performance and computational efficiency. This ensemble method works by combining the predictions of multiple decision trees, each trained on different random subsets of the data, to provide more reliable and accurate results. During the training process, the model captures complex, non-linear relationships between input features—such as historical prices, weather data, and seasonal trends—and the target variable, garlic price. Once trained, the Random Forest model generates initial predictions that highlight the underlying patterns in the data, forming one component of the hybrid forecasting system.

4.2.4 Implementing the LSTM Model:

The LSTM-based deep learning model is built with a sequential architecture designed specifically for time-series forecasting. To enhance generalization and reduce the risk of overfitting, a Dropout layer is incorporated, which randomly deactivates a portion of the neurons during training. The final Dense layer serves as the output node, converting the learned patterns into predicted garlic prices. The model is trained using the Adam optimizer, which adjusts the learning rate dynamically, and its performance is evaluated using Mean Squared Error (MSE) as the loss function. Additionally, Early Stopping is employed as a regularization strategy to monitor the validation loss, stopping the training process once the model's performance ceases to improve, thus saving time and preventing unnecessary overfitting.

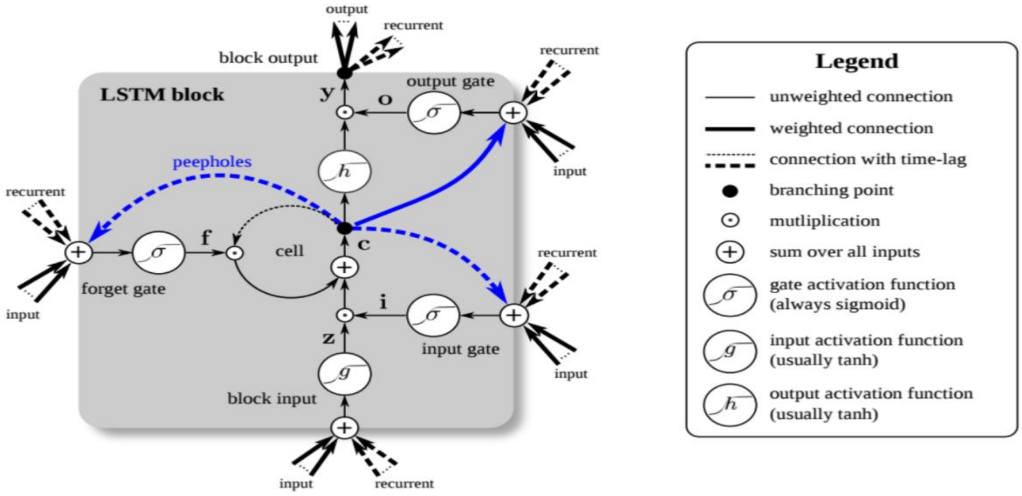

Figure 4.1 LSTM Model

4.2.5 Combining RF and LSTM Predictions (Hybrid Model):

To leverage the complementary strengths of both machine learning and deep learning techniques, a weighted averaging method is used to combine predictions from the Random Forest and LSTM models. This hybrid approach enables the system to take advantage of Random Forest's ability to manage structured data and capture complex non-linear relationships, while also utilizing the LSTM model's proficiency in learning temporal sequences and long-term dependencies in the data. In this study, the Random Forest predictions are assigned a weight of $\alpha = 0.6$, reflecting its strong ability to model static patterns and external variables, while the LSTM predictions are given a weight of 0.4 to preserve key time-dependent insights.

The final garlic price forecast is derived by combining these weighted predictions, producing a more balanced and robust output that improves accuracy over either model individually. This ensemble approach is particularly effective in dynamic agricultural markets, where both external influences and historical price trends must be considered simultaneously.

4.2.6 Model Evaluation and Performance Analysis:

To assess the predictive performance of the proposed hybrid model, a range of standard regression metrics is utilized, including Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). These metrics offer a quantitative evaluation of the difference between predicted and actual garlic prices, enabling a comprehensive analysis of model accuracy. MAE calculates the average of absolute prediction errors, while both MSE and RMSE place greater emphasis on larger errors, making them ideal for identifying significant deviations.

The performance of the individual models—Random Forest and LSTM—are evaluated separately and then compared with the hybrid model to assess the benefits of the ensemble approach. It is expected that the hybrid model will show lower error values, indicating its enhanced capacity to capture both non-linear feature interactions and temporal price trends. In addition to numerical evaluation, visual tools like line graphs displaying actual versus predicted values and error distribution charts are employed. These visualizations assist in demonstrating the model's alignment with real market trends and its potential for effective forecasting in real-world agricultural contexts.

CHAPTER-5

OBJECTIVES

The outcomes of this research aim to provide a comprehensive, data-driven tool for garlic farmers in India, enhancing their decision-making and profitability through predictive forecasting and actionable insights. By leveraging a hybrid model that integrates Random Forest and Long Short-Term Memory (LSTM) networks, the model will address several key aspects influencing garlic cultivation beyond environmental and economic factors. The model will predict not only garlic prices but also help determine the optimal sowing windows by incorporating weather data from previous years. This will enable farmers to plant their garlic at the most advantageous times, maximizing yield and minimizing risks from adverse weather conditions such as droughts or excess rainfall. The integration of historical meteorological data will also help the model forecast the best months for planting, considering how past weather patterns influenced crop growth and market conditions.

Moreover, the model will factor in additional aspects such as soil health, seed quality, and pest management, by considering historical agronomic practices and pest outbreaks that may have affected yields in previous years. These factors will be integrated into the Random Forest model, allowing it to predict how different practices and environmental conditions could impact the farmer's yield and profitability. Additionally, the model will use labor availability and cost data to predict the impact of labor-related challenges on farming costs and profits, helping farmers plan better during critical planting and harvesting periods.

Another critical outcome of the model is its ability to incorporate market access and supply chain efficiency, factors that significantly affect farmers' ability to sell their produce at favorable prices. The model will consider regional differences in infrastructure, such as transportation and storage facilities, to estimate how delays or inefficiencies in the supply chain might affect garlic prices and farmer profitability. Market demand, both domestic and export, will be factored into the model, helping predict fluctuations in garlic prices based on shifts in consumer preferences and global trade dynamics.

Government policies and subsidies will also play a role in shaping the model’s predictions. The model will include data on minimum support prices (MSPs), subsidy schemes, and export regulations, ensuring that farmers benefit from up-to-date information about available support systems. The system will also account for regulatory factors, such as changes in pesticide usage or export tariffs, that could influence market prices.

Incorporating technological advancements, such as the adoption of efficient irrigation systems, farm mechanization, and data-driven tools for pest detection and weather forecasting, the hybrid model will predict how these technologies impact garlic farming costs and potential profits. By using such tools, farmers can optimize water usage, reduce labor costs, and increase productivity.

By analyzing climate change effects and long-term weather patterns, the model will help farmers anticipate potential shifts in planting and harvesting cycles, providing insights on how changing climate conditions may affect garlic cultivation. In regions prone to extreme weather events, the model will factor in disaster risk assessments, adjusting profitability predictions to account for the likelihood of floods, droughts, or cyclones.

The hybrid model will combine these multiple data inputs to produce predictions that are not only more accurate but also relevant to real-world farming challenges. Using established regression metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE), the model will be rigorously evaluated to ensure its effectiveness. Ultimately, the tool aims to empower garlic farmers with a more comprehensive, holistic, and actionable forecast that supports better decision-making, improves profitability, and reduces risks, all while integrating advanced data science with practical, on-the-ground agricultural needs.

In conclusion, the outcomes of this project will help farmers understand the timing and optimal conditions for garlic cultivation, enabling them to make informed decisions about planting, irrigation, pest management, and harvesting, all while accounting for external factors like weather conditions, market demand, and government policies. The result will be a more resilient and profitable garlic farming ecosystem, supported by data-driven insights tailored to the needs of smallholder farmers in India.

CHAPTER-6

SYSTEM DESIGN & IMPLEMENTATION

6.1 System Architecture

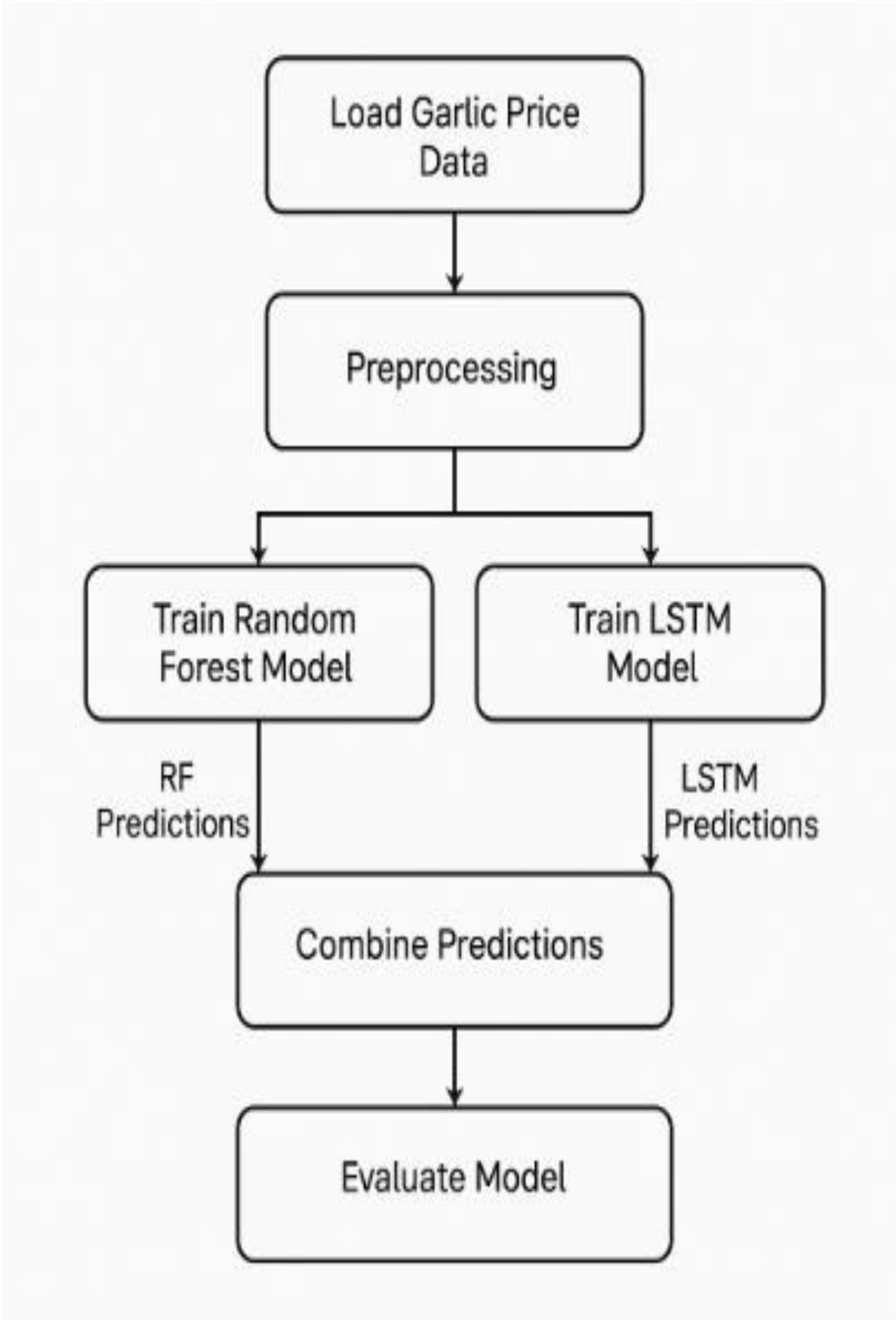

Figure 6.1 System Architecture

The garlic price prediction project begins with loading the garlic price data, which serves as the foundation of the entire analysis. This data is collected from a reliable source, such as government databases, online market platforms, or local records, and typically includes historical garlic prices, timestamps, and additional relevant variables, such as regional factors, weather conditions, or agricultural inputs that may influence garlic prices. Proper data loading ensures that all the necessary features are captured for modeling.

Following the data loading, the next critical step is preprocessing the data. In this phase, the raw data undergoes cleaning and preparation to ensure it is suitable for training predictive models. Preprocessing may include handling missing or null values by either imputing them with reasonable estimates or removing them if necessary. Additionally, data normalization or scaling techniques are applied to standardize the features, especially when using algorithms that are sensitive to the magnitude of input features. Feature engineering might also be performed, where new features are derived from existing ones—such as creating moving averages or extracting seasonal patterns from the time-series data. Lastly, the data is split into training and testing sets, ensuring that the models can be effectively validated against unseen data.

Once the data is ready, the next step is to train the models. In this case, two powerful models are employed: the Random Forest model and the LSTM (Long Short-Term Memory) model. The Random Forest model, an ensemble machine learning method, is used to capture complex, non-linear relationships between garlic prices and external variables. By constructing multiple decision trees and aggregating their outputs, Random Forest ensures robust predictions, reducing the likelihood of overfitting. On the other hand, the LSTM model, a deep learning architecture designed for time-series forecasting, is employed to learn temporal dependencies and patterns in historical garlic prices. LSTM networks are particularly effective at modeling sequential data and can capture trends over long periods, making them ideal for predicting future garlic prices based on past values.

After training both models, the project moves on to combine the predictions from the Random Forest and LSTM models. Since both models bring unique strengths to the prediction process, combining their outputs using techniques like averaging, weighted averaging, or stacking can provide a more accurate and reliable prediction of future garlic prices. The ensemble approach benefits from the individual models' strengths and compensates for their weaknesses, ensuring that the final predictions are well-rounded and robust.

Finally, the trained models are subjected to an evaluation phase. During evaluation, the performance of each model is assessed using appropriate metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2), among others. These metrics help quantify how closely the predicted prices align with the actual observed garlic prices. A thorough evaluation ensures that the best model is selected based on predictive accuracy and performance, and allows for adjustments to be made if necessary. The final goal of this project is to have a reliable system that can accurately forecast garlic prices, helping farmers and businesses make informed decisions.

CHAPTER-7

TIMELINE FOR EXECUTION OF PROJECT

(GANTT CHART)

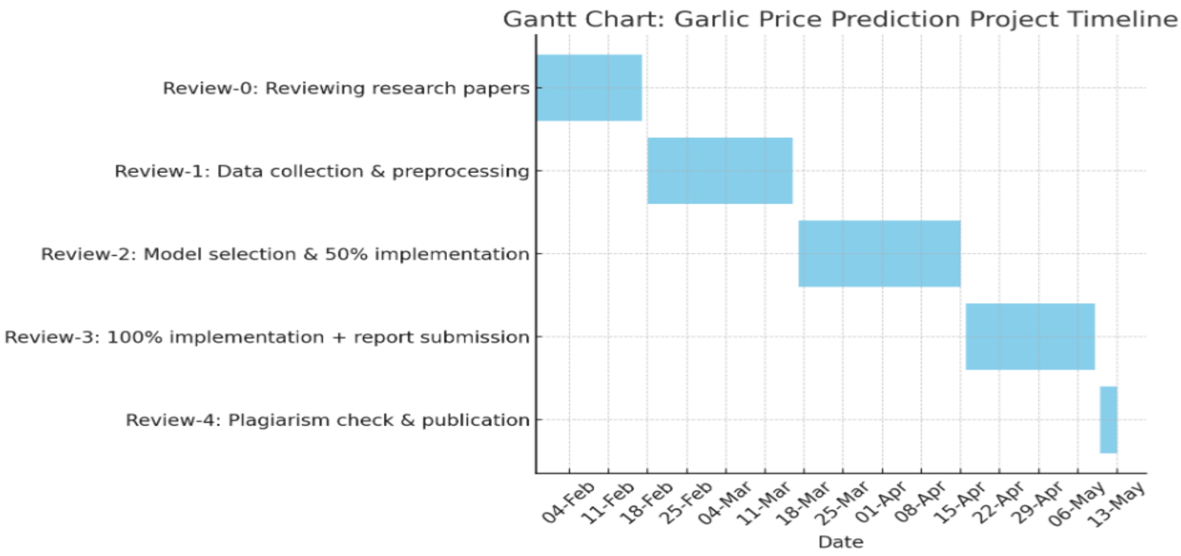

Figure 7.1 Project Timeline

This is a simple project timeline in a Gantt chart format showing tasks completed between January, February, March, April and May 2025. Key points

Tasks: Includes steps like:

1. Reviewing research papers
2. Collecting and preprocessing data
3. 50% implememtation
4. 100% implentation
5. Plagiarism check and publication

Timeline: Tasks are spread across November and December, with each task represented by a colored bar.

Progress: All tasks are 100% complete.

CHAPTER-8

OUTCOMES

The implementation of the hybrid garlic price forecasting system resulted in several significant outcomes, demonstrating its technical validity and practical importance within India’s agricultural landscape. By combining Random Forest and Long Short-Term Memory (LSTM) models, the system successfully addressed the limitations of traditional forecasting methods. While conventional models like ARIMA, SVR, and Prophet are often limited by assumptions of linearity or by their inability to capture temporal dependencies effectively, the hybrid approach harnessed the strengths of both ensemble tree learning and deep sequential modeling. Random Forest excelled in modeling non-linear relationships and feature interactions, while LSTM captured the temporal dependencies and long-term trends present in the garlic price data. The ensemble design led to improvements in predictive accuracy, as reflected in better performance on key evaluation metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). This enhanced accuracy is particularly vital in the garlic market, where price fluctuations can significantly impact farmers' income and market stability..

A key innovation of this project was the successful integration of external agro-climatic factors, such as rainfall and temperature data, into the forecasting model. These variables directly influence garlic production cycles, harvest outcomes, and ultimately, market supply and prices. In contrast to many traditional models that rely solely on historical price data, the hybrid system employed a multi-dimensional approach by incorporating climatic indicators that better reflect the real agricultural environment. This made the model more resilient and contextually aware, enabling it to react to irregular seasonal patterns, extreme weather events, or climate-related supply disruptions. Consequently, the system offers more accurate forecasts, helping stakeholders adjust their strategies in response to emerging trends and potential market disturbances.

Another notable feature of the system is its ability to provide region-specific predictions. Rather than offering generalized forecasts, the model was designed to incorporate district-level inputs and unique regional factors. This localization ensures that the model reflects local variations in climate, soil conditions, agricultural practices, and market behaviors, all of which are crucial in the Indian agricultural landscape. For farmers, this level of detail provides valuable insights that guide decisions related to planting, irrigation, harvesting, and market timing. Additionally, the system’s near real-time forecasting capabilities enable proactive planning and timely responses. For example, if the model predicts a price drop in a particular area, farmers can adjust their strategies, such as postponing sales or implementing storage techniques, to mitigate potential losses.

To ensure the system is practical and user-friendly for its target audience, a visualization dashboard was created. This interface converts complex forecasting data into easy-to-understand visual representations, such as line charts, bar graphs, and interactive geospatial heatmaps. These visual tools cater to users with varying levels of technical expertise, ranging from smallholder farmers to agricultural analysts and policymakers. The dashboard not only improves the clarity of the prediction process but also supports data-driven decision-making. For example, a farmer can use a district-specific map of predicted garlic prices to determine the best time and location to sell, while a policymaker can pinpoint regions at risk of price drops and plan interventions or subsidies accordingly.

The success of this hybrid model highlights the scalability and versatility of AI-driven solutions in agriculture. With minimal adjustments, the same framework can be extended to other crops like onions, tomatoes, or chilies, which experience similar price fluctuations. The model’s modular design, along with its capacity to integrate various datasets, enables easy adaptation and expansion to encompass broader agricultural forecasting projects. This flexibility paves the way for future developments in crop recommendation systems, yield prediction tools, and market risk assessment platforms, thereby contributing to a data-driven agricultural ecosystem.

In conclusion, this research not only tackled the challenge of garlic price prediction with robust technical methods but also showcased a scalable solution that contributes to the larger objectives of agricultural sustainability, food security, and economic resilience. By incorporating AI alongside real-world agricultural factors, the system bridges the gap between data science and the practical realities of farming, providing a valuable tool for empowering India's farming communities and enhancing market efficiency.

CHAPTER-9

RESULTS AND DISCUSSIONS

9.1 Results:

9.1.1. Predictive Accuracy of models

Accuracy is a widely used performance metric in classification problems, in regression tasks like price prediction, accuracy is typically assessed using error-based measures such as Mean Squared Error (MSE), Mean Absolute Error (MAE), or Root Mean Squared Error(RMSE). Here Accuracy = 1 minus the ratio of the Mean Squared Error (MSE) to the Variance (VAR) of the actual values, multiplied by 100.

This form of accuracy gives an indication of how well the model performs in comparison to a baseline model that always predicts the mean of the actual values. A higher accuracy percentage indicates better performance.

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

Table 9.1 Accuracy Scores

The model accuracy comparison graph illustrates the performance differences among various predictive models, including ARIMA, SVR, Prophet, XGBoost, LSTM, and the Hybrid Model that integrates Random Forest with LSTM. Among these, the Hybrid Model demonstrates a clear advantage, achieving the highest accuracy of 89%, while the remaining models show accuracy levels ranging from 63% to 69%.

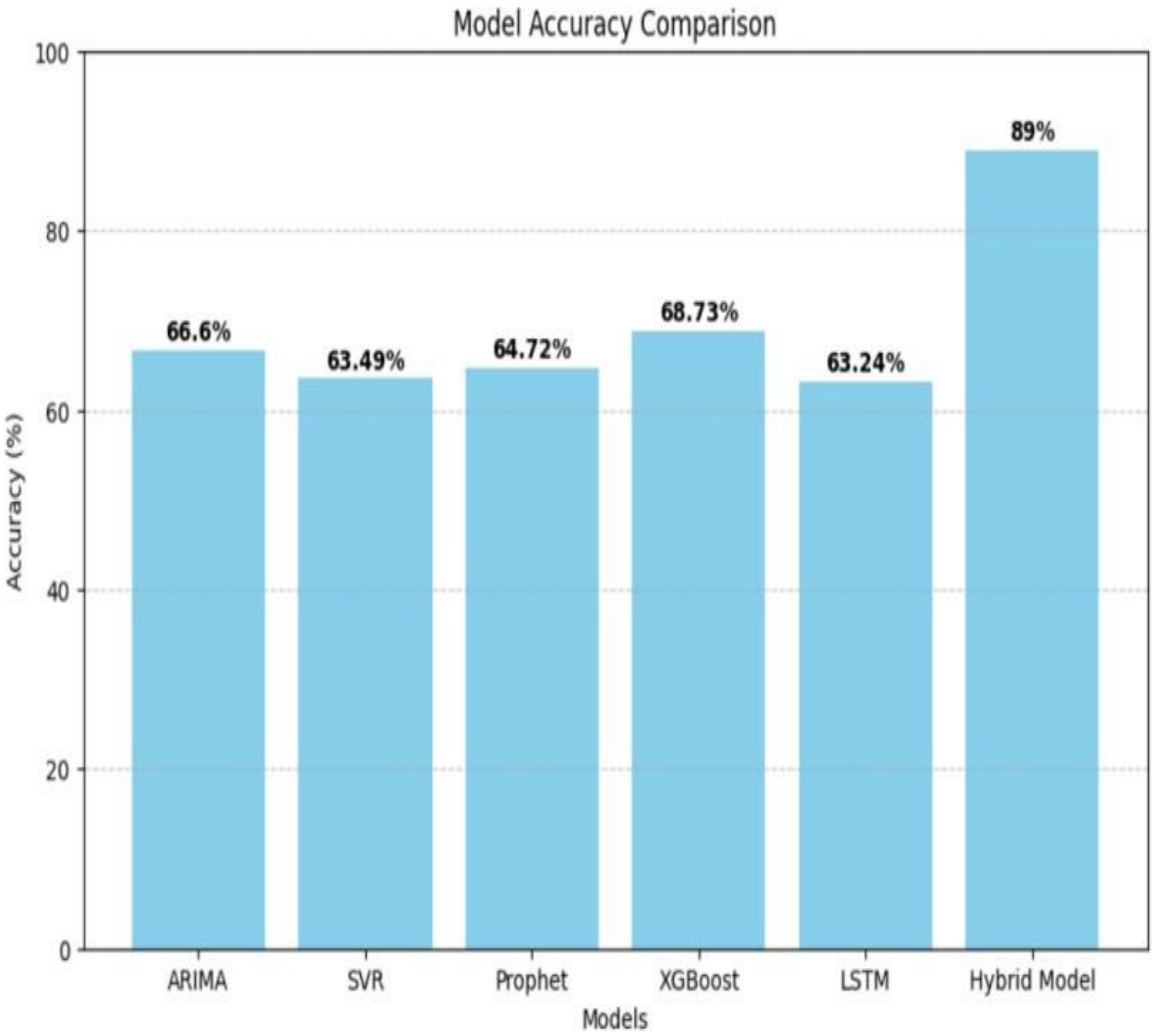

Figure 9.1 Accuracy Score Bar Graph

The evaluation of each individual model reveals both their capabilities and limitations in forecasting garlic prices. ARIMA, with an accuracy of 66.6%, effectively models linear trends but is less reliable during abrupt market shifts. SVR, scoring 63.49%, falls short in capturing long-term dependencies. Prophet, with 64.72% accuracy, performs well in handling seasonality but lacks flexibility when faced with unexpected changes. XGBoost achieves a relatively higher accuracy of 68.73% but has difficulty managing time-based dependencies. LSTM, although designed for modeling long-term trends, attains an accuracy of 63.24% due to its dependence on large datasets and complex tuning. In summary, while each model offers distinct advantages, none alone addresses all aspects of the forecasting challenge.

9.1.2 Visualization

Visualization is essential for interpreting and understanding complex datasets, as it converts raw data into graphical formats that simplify the identification of trends, patterns, and key insights. In the context of garlic price prediction, two primary types of visualizations have been employed to effectively present and analyze the results.

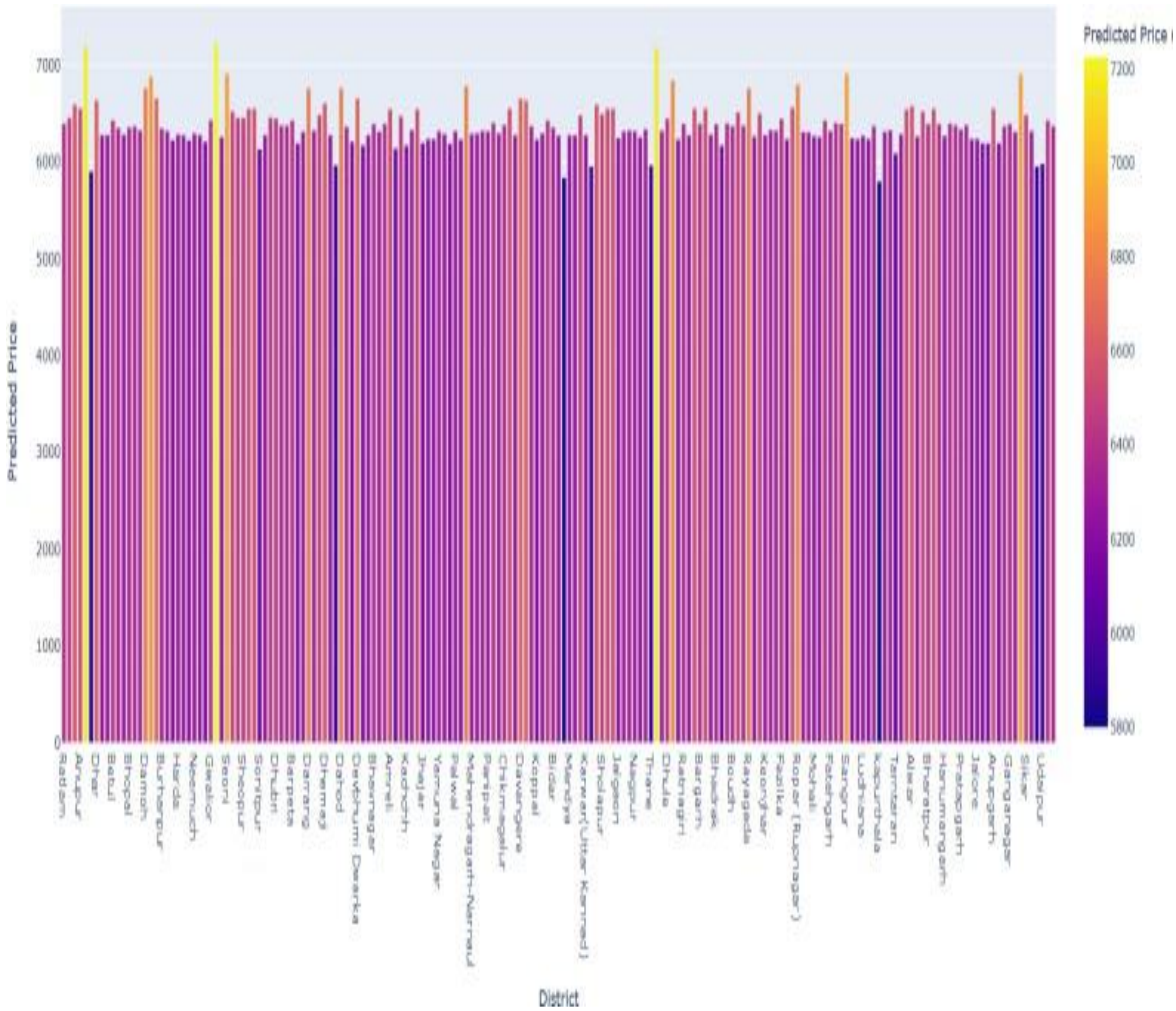

Figure 9.2 District Wise Price Prediction Bar Chart

This bar chart illustrates the predicted garlic prices across various districts, where each bar corresponds to a specific district and its height indicates the forecasted price. A color gradient ranging from purple to yellow is applied to enhance visual clarity, making it easier to distinguish between lower and higher price values. This visual representation allows for quick and effective comparisons between regions, enabling stakeholders to identify districts with the highest and lowest projected garlic prices at a glance.

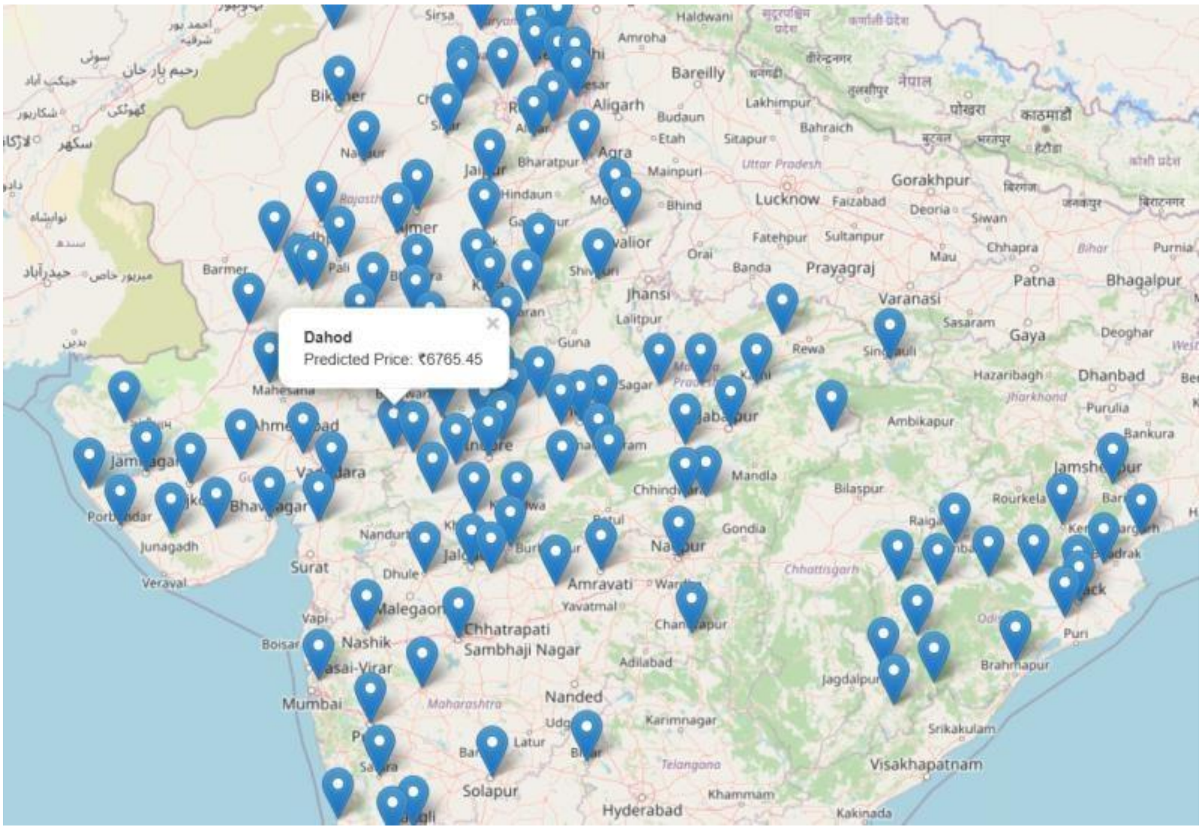

Figure 9.3 Geographical Map for Spatial Price Distribution

The interactive map displays the spatial distribution of predicted garlic prices across India, with each blue pin representing a specific district. This geographic visualization aids traders, policymakers, and farmers in understanding regional price variations, allowing them to make more informed decisions related to pricing trends, transportation logistics, and market selection.

Figure 9.4 Year with highest Prices

Figure 9.4 illustrates the annual peak prices of garlic from the year 2015 to 2025. It visually highlights how garlic prices have fluctuated over the years and provides insight into trends, potential volatility, and recent market surges. Overall, this graph reflects a sharp increase in garlic prices in the last three years, suggesting rising market demand, possible supply constraints, or external economic pressures. This escalating trend underlines the importance of accurate forecasting models to help stakeholders prepare for price volatility and make informed agricultural and financial decisions.

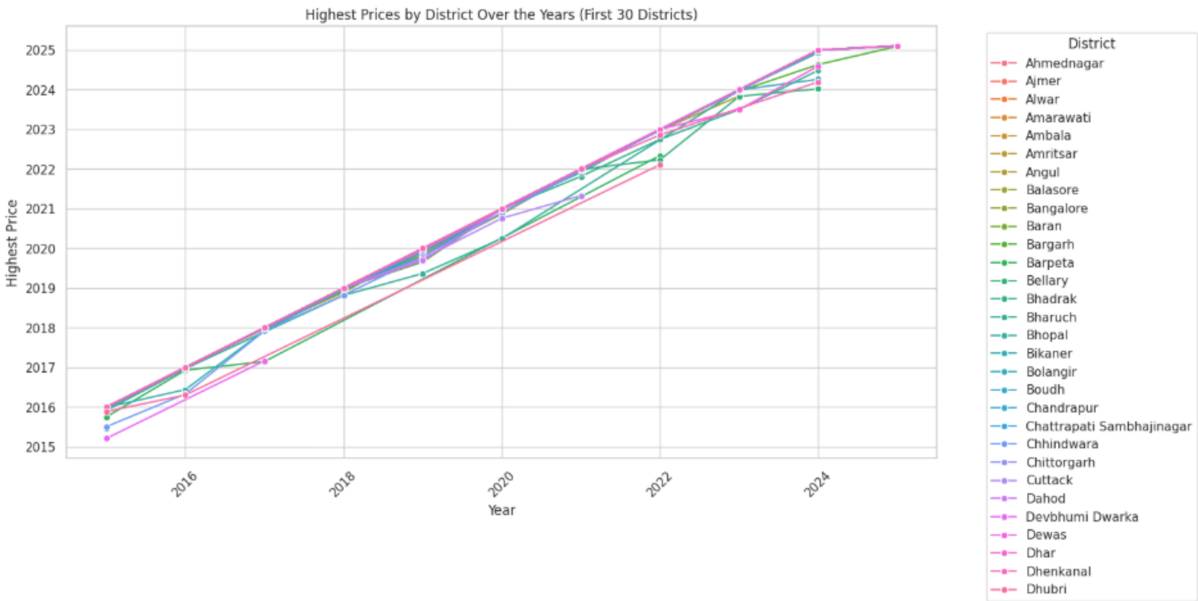

Figure 9.5 Highest Prices by District

Figure 9.5 illustrates the trend of garlic's peak prices across 30 different districts from 2015 to 2024. Each line represents a district, and the y-axis shows the highest recorded garlic price, while the x-axis represents the corresponding year. Overall, the graph indicates a consistently upward trend in garlic prices across nearly all districts over the years. From 2015 to 2020, the price increase is steady but moderate. However, after 2020, the rate of increase becomes more pronounced, with most districts experiencing sharper growth in garlic prices between 2021 and 2024. This suggests a national pattern of rising garlic prices, potentially driven by growing demand, supply shortages, inflation, or changes in agricultural policy.

In summary, the graph highlights a clear and widespread increase in garlic prices across multiple regions, emphasizing the importance of region-specific forecasting tools to support local agricultural and economic planning.

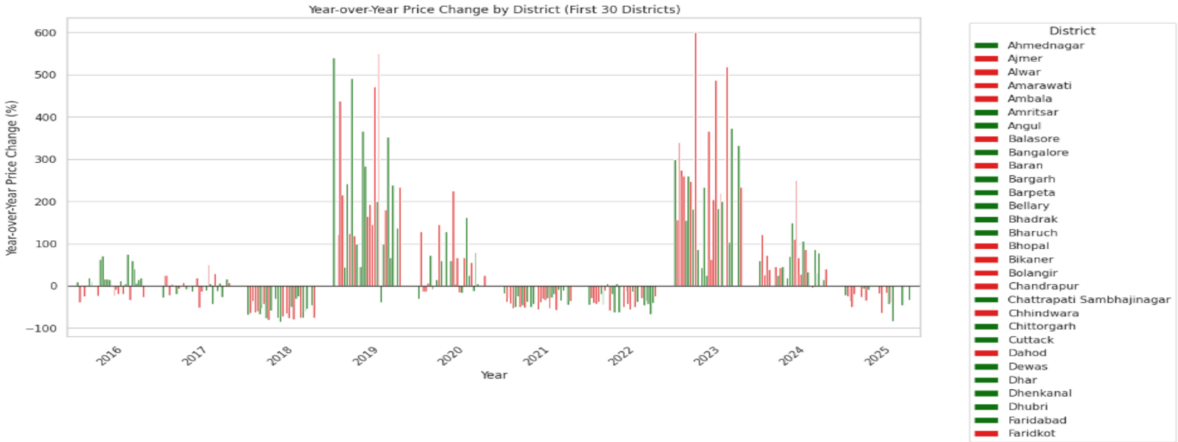

Figure 9.6 Year over year price change by district

Figure 9.6 depicts the percentage change in garlic prices from one year to the next across various districts, spanning from 2015 to 2025. Each bar represents the change in price for a particular district in a given year, with green and red bars used to distinguish between districts. The graph reveals high volatility in garlic prices, especially between 2018 and 2020, where several districts experienced extreme spikes—some with price increases exceeding 500%, indicating sharp surges likely due to supply disruptions, climate-related crop failures, or market anomalies. Similarly, another cluster of steep changes is visible around 2023, showing how unpredictable and reactive the market can be. Conversely, multiple bars dip below the 0% line, showing negative year-over-year changes, especially around 2021 and 2022. This suggests that prices dropped significantly in many districts during these years, potentially due to market corrections, surplus supply, or reduced demand.

Overall, this graph emphasizes the volatile nature of garlic pricing across regions and time, reinforcing the necessity of predictive tools and hybrid forecasting models to help mitigate risk and guide informed agricultural and financial decisions.

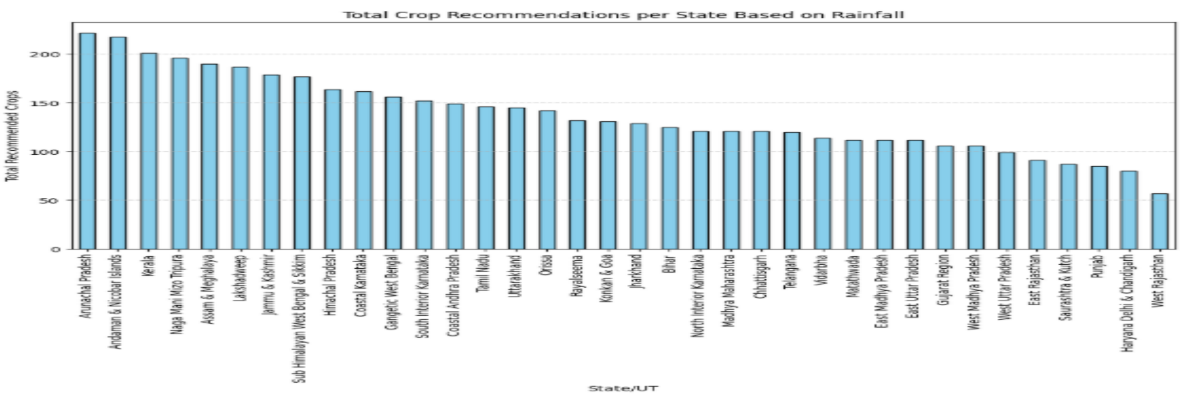

Figure 9.7 Total Crop Recommendation per state based on rainfall

Figure 9.7 illustrates the number of crop types recommended for each Indian state or union territory, taking into account regional rainfall patterns. This type of analysis is essential in aligning agricultural planning with climatic suitability, helping farmers optimize yield and sustainability. From the graph, Arunachal Pradesh leads with the highest number of crop recommendations (above 210), closely followed by Andaman & Nicobar Islands and Kerala. These regions receive abundant rainfall, making them suitable for a diverse range of crops, including both staple and high-value commercial varieties. In contrast, states like Punjab, Haryana, Delhi, and West Rajasthan have fewer recommended crops (less than 100). These areas receive less rainfall or face water resource challenges, which limits crop diversity and requires a more strategic selection of drought-resistant or irrigated crops.

Overall, this graph provides a useful snapshot for agriculture policy, regional crop planning, and climate adaptation, enabling decision-makers to recommend crops that align with environmental conditions and maximize efficiency in resource usage.

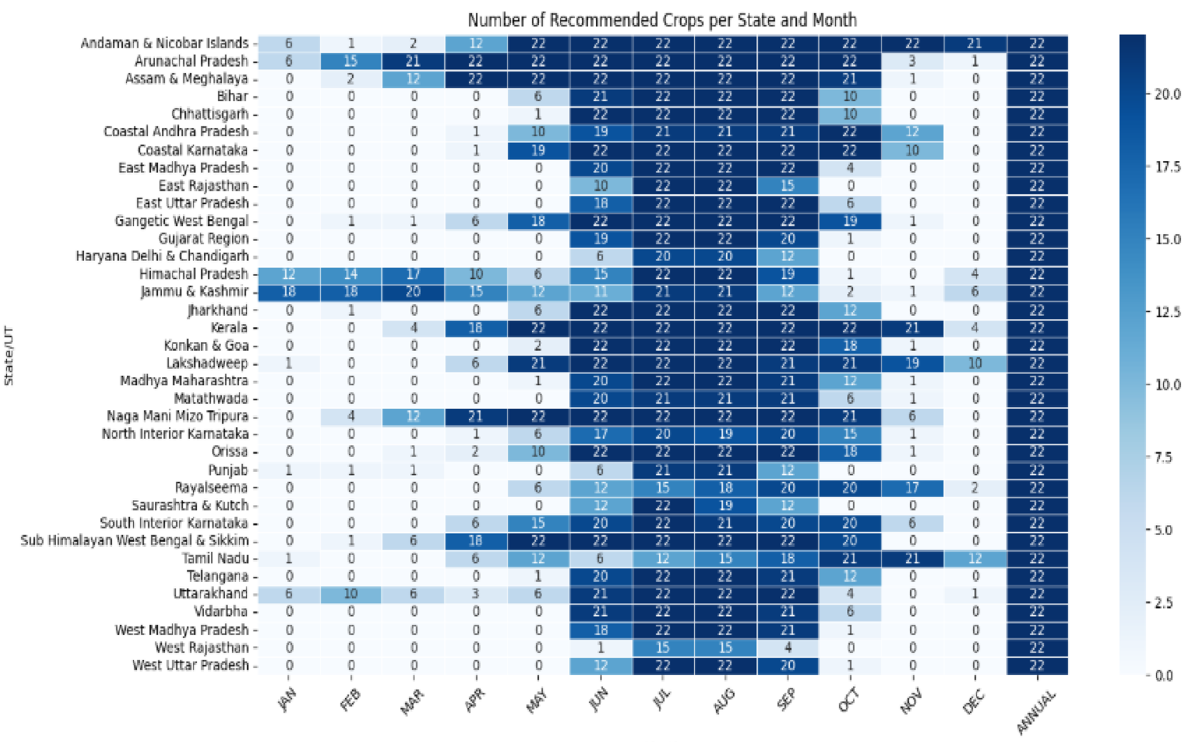

Figure 9.8 Number of recommended crops per state and month

The heatmap illustrates the number of recommended crops per state and month across various regions of India. It reveals significant seasonal variation in crop recommendations, with the months of June to September showing the highest number of crop suggestions in nearly all states. This pattern aligns with the Kharif season, during which rainfall is abundant and agricultural activity peaks. States such as Arunachal Pradesh, Jammu & Kashmir, and Assam & Meghalaya display consistently high crop recommendations during this period, often reaching the maximum of 22 crops per month, highlighting their rich agro-climatic potential during the monsoon. In contrast, the winter months—particularly December and January—show a notable decline in crop recommendations across most states. This suggests a seasonal reduction in agricultural variety, likely due to colder temperatures and reduced rainfall during the Rabi season. Some regions, such as Punjab, Haryana, and Rajasthan, show minimal crop activity in these months, which could be attributed to water availability issues or more limited crop cycles.

Interestingly, a few states like Jammu & Kashmir, Arunachal Pradesh, and Sub-Himalayan West Bengal & Sikkim maintain a relatively high number of crop recommendations throughout the year. This reflects their diverse microclimates and the ability to support agricultural activities beyond typical monsoon cycles. Overall, the heatmap emphasizes the critical influence of climate, particularly rainfall, on crop planning and underscores the need for region-specific agricultural strategies to optimize productivity year-round.

9.2 Algorithms:

9.2.1 Random Forest

Random Forest is an ensemble machine learning algorithm that builds multiple decision trees and merges their outputs to improve the accuracy and stability of predictions. It operates by training several decision trees on random subsets of the data and features, and then aggregating their results—typically through averaging in regression tasks. This approach minimizes the risk of overfitting that is often associated with single decision trees and offers robustness in handling high-dimensional and noisy datasets. In the context of garlic price prediction, Random Forest is particularly effective in modeling non-linear relationships among various input features such as date, district, rainfall, and past prices. It also provides valuable feature importance scores, which help identify the most influential variables affecting garlic prices. Additionally, it performs well with structured data, requires minimal preprocessing, and handles missing values efficiently. However, Random Forest does not consider the sequence or order of the data over time, making it less suitable for capturing temporal dependencies or trends that evolve sequentially—an essential aspect in time-series forecasting.

9.2.2 Long Short Term Memory (LSTM)

LSTM, or Long Short-Term Memory, is a type of recurrent neural network (RNN) specifically designed to learn and remember patterns in sequential data. Unlike traditional neural networks, LSTMs have a memory cell structure and gating mechanisms that enable them to retain information over long periods, making them highly effective for time-series analysis. In the context of garlic price forecasting, LSTM networks are used to model the temporal behavior of prices by analyzing historical trends and identifying seasonal or recurring patterns. The LSTM architecture includes input, forget, and output gates that control the flow of information, allowing the model to focus on relevant past values while discarding less important data. This makes LSTM particularly powerful in dealing with fluctuations in garlic prices caused by market cycles, climate changes, or other time-sensitive factors. However, LSTMs typically require a large volume of clean, sequential data and are computationally intensive to train. Moreover, they often require careful tuning of hyperparameters, such as the number of units, batch size, and learning rate, to achieve optimal performance.

9.2.3 Hybrid Model: Random Forest + LSTM

To overcome the limitations of individual models and leverage their respective strengths, a hybrid approach combining Random Forest and LSTM is employed in this study. This hybrid model integrates the capabilities of Random Forest in handling structured, non-sequential features and the strength of LSTM in capturing time-series dependencies. The process begins with Random Forest being used for feature selection and to generate initial predictions based on historical market and environmental data. LSTM is then applied to model the temporal dynamics in the garlic price trends, learning patterns over time from sequences of past values. The final predictions are produced using a weighted averaging technique, where a weight of 0.6 is assigned to the Random Forest output and 0.4 to the LSTM output. This weighting reflects the complementary contribution of both models—Random Forest providing structured, feature-rich context and LSTM offering deep temporal learning. The hybrid model improves overall prediction accuracy, enhances generalization, and provides a more robust solution to the complex and dynamic nature of agricultural price forecasting. By combining these two models, the system is better equipped to deliver accurate, timely, and region-specific garlic price predictions that can support informed decision-making for farmers, traders, and policymakers alike.

9.3 Evaluations:

9.3.1 MAE

Mean Absolute Error (MAE) quantifies the closeness of predicted values to the actual values by taking the absolute differences and averaging them. Lower MAE values indicate better model accuracy

$$\textbf{MAE=(1/n)} \sum \textbf{(yi - } \hat{y}\, i)$$

Where

N = Total number of observations

yi = Actual value of the iiith observation

$\hat{y}\, i$ = Predicted value of the iiith observation

$\sum$ = Summation symbol, summing over all observations.

9.3.2 MSE

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

$$\textbf{MSE= (1/n)} \sum \textbf{(yi - } \hat{y}\, i)^2$$

Where

N represents the total number of observations

Yi denotes the actual value of the i-th observation

$\hat{y}\, i$ represents the predicted value of the i-th observation

CHAPTER - 10

CONCLUSIONS AND FUTURE DIRECTIONS

10.1 Conclusion

The garlic price prediction system developed in this study marks a significant step forward in the field of agricultural price forecasting in India. By leveraging a hybrid model that integrates Random Forest and Long Short-Term Memory (LSTM) networks, the system successfully combines feature selection capabilities with deep temporal learning to enhance forecasting precision. The outcomes clearly demonstrate the potential of machine learning in supporting farmers, traders, and policymakers with data-driven insights that mitigate the financial risks posed by market price fluctuations. Moreover, the inclusion of district-level visualizations and a user-friendly interface enhances the system's practicality and accessibility for real-world agricultural decision-making.

Accurate forecasting of crop prices plays a vital role in stabilizing agricultural markets, ensuring fair income for farmers, and maintaining affordable pricing for consumers. While traditional models such as ARIMA and SVR often struggle to capture the dynamic and nonlinear behaviors of agricultural markets, deep learning models like LSTM have shown greater flexibility in identifying temporal patterns. Nevertheless, each technique has its limitations. The hybrid model proposed in this project serves as a balanced and effective solution, merging the strengths of both machine learning and deep learning paradigms to offer more reliable predictions.

10.2 Future Direction:

This project lays a robust foundation for the future development of AI-powered agricultural intelligence systems. Moving forward, several enhancements can significantly improve the performance, reach, and impact of the system:

10.2.1 Integration of Real-Time and External Data Sources

To enhance forecasting precision and adaptability, future versions of the model should incorporate real-time inputs such as current market prices, weather forecasts, and economic indicators like fuel costs or inflation rates. Unlike static historical data, these dynamic inputs will allow the system to respond quickly to emerging trends or disruptions, such as sudden rainfall during harvest or political changes affecting trade policies. Real-time integration will make the model more agile and practical, providing up-to-date forecasts that can support on-the-ground decision-making with minimal lag.

10.2.2 Development of Farmer-Centric Platforms and Accessibility Features

A significant step toward real-world adoption involves the creation of mobile and web-based platforms tailored to farmers and local market stakeholders. These platforms should support offline functionality, push notifications, and region-specific alerts to ensure timely and actionable information. Further, local language support, voice-based interactions, and simplified visual elements like graphs and district heatmaps will be critical to reaching rural users, many of whom may lack formal education or internet fluency. Such tools will help bridge the digital divide and empower farmers with technology that feels intuitive and trustworthy.

10.2.3 Explainable AI (XAI) for Trust and Transparency

As AI adoption grows in sensitive sectors like agriculture, ensuring transparency and interpretability of model outputs becomes increasingly important. Implementing Explainable AI (XAI) techniques can help users understand why the model forecasts a price rise or fall, by highlighting key influencing factors such as climate changes or regional production surpluses. By demystifying the prediction process, XAI will foster greater user trust, enhance decision-making, and potentially improve collaboration between AI developers and agricultural stakeholders.

10.2.4 Socio-Economic and Policy-Aware Forecasting

While current models primarily rely on agro-climatic and price data, incorporating socio-economic factors such as government subsidies, Minimum Support Prices (MSP), labor availability, and export-import regulations can greatly improve forecasting accuracy and realism. These factors often have substantial influence on market behavior but are typically excluded from technical models. By reflecting the broader policy and economic environment, future models can become holistic decision-support systems, guiding not only farmers but also agri-business planners and policy designers.

10.2.5 Expansion to Yield Forecasting and Multi-Crop Systems

An impactful direction for future development is the integration of crop yield prediction alongside price forecasting. Using tools such as satellite imagery, remote sensing, and vegetation indices (like NDVI), the model could estimate expected production levels and correlate them with market prices. This dual-view approach would enable farmers to assess both supply potential and profitability, facilitating smarter choices regarding crop selection, investment, and harvesting. Additionally, scaling the system to include other high-volatility crops—such as onions, tomatoes, or chilies—can widen its impact and support a comprehensive agricultural intelligence platform.

CHAPTER-11

REFERENCES

- Singh, R., Gupta, P., & Sharma, S. (2023). A comprehensive study of machine learning techniques for disaster prediction. Journal of Computational Science and Engineering, 11(4), 234-247. doi: 10.1016/j.jcse.2023.01.012.
- Lee, H., & Choi, S. (2022). Forecasting natural disasters with deep learning: A case study on typhoon prediction. Computational Environmental Science, 8(2), 56-67. doi: 10.1080/23456789.2022.2045698.
- Automated Agriculture Commodity Price Prediction System with Machine Learning Techniques - Zhiyuan Chen*1, Howe Seng Goh1, Kai Ling Sin1, Kelly Lim1, Nicole Ka Hei Chung1, Xin Yu Liew
- Support Vector Machines for Regression and Forecasting – Vapnik, V.
- Prophet: Forecasting at Scale – Taylor, S. J., & Letham, B.
- XGBoost: A Scalable Tree Boosting System – Chen, T., & Guestrin, C.
- Long Short-Term Memory Networks for Time Series Prediction – Hochreiter, S., & Schmidhuber, J.
- Hybrid Machine Learning Models for Price Prediction – Zhang, G., Eddy Patuwo, B., & Hu, M. Y.
- Impact of Climate Factors on Agricultural Prices in India – Kumar, R., & Singh, P.
- Comparative Study of Machine Learning Approaches for Market Forecasting – Patel, J., Shah, S., Thakkar, P., & Kotecha, K.
- Random Forest Algorithm for Agricultural Price Prediction – Breiman, L.
- Deep Learning in Agricultural Forecasting – LeCun, Y., Bengio, Y., & Hinton, G.
- An Overview of Time Series Forecasting Techniques – Hyndman, R. J., & Athanasopoulos, G.
- Integrating Machine Learning for Agricultural Market Intelligence – Chakraborty, S., & Roy, D.
- Challenges in AI-Based Agricultural Forecasting – Sharma, V., & Mehta, A.
- Big Data and AI in Agricultural Supply Chain Management – Jha, S., & Agrawal, R.
- A Review of Hybrid Models in Predicting Commodity Prices – Gupta, A., & Verma, S.
- Evaluating the Performance of Machine Learning Models in Price Forecasting – Bose, K., & Sen, A.
- Leveraging LSTM for Market Trend Analysis – Wang, Y., & Li, H.

- The Role of AI in Enhancing Agricultural Productivity – Mishra, P., & Tiwari, R.
- Feature Engineering for Agricultural Price Prediction – Das, A., & Mukherjee, S.
- Exploring the Influence of Macroeconomic Factors on Market Prices – Banerjee, N., & Singh, A
- Neural Networks for Time Series Forecasting: A Review – Adhikari, R., & Agrawal, R. K.
- Machine Learning Applications in Agricultural Economics – Fernandez, S., & Martin, J.
- The Use of LSTMs in Financial and Commodity Market Forecasting – Huang, J., & Zhang, T.
- Bayesian Methods in Agricultural Price Prediction – O'Hagan, A., & West, M.
- Explainable AI in Agricultural Forecasting Models – Ribeiro, M. T., Singh, S., & Guestrin, C.
- A Comparative Analysis of Time Series Models for Price Prediction – Box, G. E. P., & Jenkins, G. M.
- AI-Driven Decision Support Systems in Agriculture – Bhattacharya, R., & Chattopadhyay, S.
- Understanding Price Volatility in Agricultural Markets – Nelson, G. C., & Robertson, R.
- Predictive Analytics in Agri-Business: Challenges and Opportunities – Kapoor, V., & Malhotra, R.
- The Role of Reinforcement Learning in Agricultural Price Forecasting – Sutton, R. S., & Barto, A. .

APPENDIX-A
PSEUDOCODE

1. Setup and Perform EDA

- **Import Libraries:** Essential Python libraries are loaded for data handling (Pandas), numerical computations (NumPy), visualization (Matplotlib, Seaborn), and machine learning/deep learning (Scikit-learn, TensorFlow).

- **Load Dataset:** The garlic price dataset is imported from a CSV file containing price information from various markets.

- **Preview and Summarize Data:**

1. First few rows are displayed using `.head()` to understand data structure.
2. Data types, missing values, and memory usage are checked using `.info()`.
3. Descriptive statistics are generated using `.describe()`.

- **Check Data Quality:**

1. Missing values in numerical columns are handled using mean imputation.
2. Duplicate rows are identified and removed.

- **Analyze Data Distribution:**

1. Unique values in categorical columns are counted (e.g., market names).
2. Mean values are calculated for numeric columns.

- **Verify Time-Series and Dataset Dimensions:**

1. Extract and display unique years from the dataset.
2. Print total number of rows and columns

2. Data Transformation

- **Feature Engineering:**

1. Extract date components (Year, Month, Day) from the 'Price Date' column.
2. Drop irrelevant columns such as index numbers.
3. Encode categorical features using Label Encoding.

- **Data Scaling:** Normalize numerical values using MinMaxScaler to ensure consistent input scaling, especially for LSTM.

3. Splitting and Modeling Preparation

- **Define Features and Target:**

1. The primary features selected include minimum price, maximum price, and time-based components like year, month, and day. These factors are expected to influence the modal price significantly.

2. The 'Modal Price' column is set as the target variable, representing the price to be predicted.

- **Handle Data Imbalance (if applicable):** Prior to modeling, price value distributions are checked for any extreme imbalance or outliers which may skew predictions.

- **Train-Test Split:**

1. The dataset is split into training and test sets using an 80:20 ratio to ensure that model performance is validated on unseen data.

2. The split is randomized but controlled using a fixed seed to ensure reproducibility.

- **Scaling:**

1. Features are scaled using MinMaxScaler to the [0, 1] range for compatibility with LSTM input.

2. The target variable is also scaled separately to maintain prediction consistency.

4. Model Development

- **Random Forest Model:**

1. Trained using 100 estimators.

2. Captures non-linear relationships in pricing data.

3. Provides a baseline prediction and feature importance.

- **LSTM Model:**

1. Designed with 50 LSTM units.

2. Trained using normalized time-series data reshaped into 3D arrays.

3. Incorporates early stopping to prevent overfitting.

5. Hybrid Model Formation

- **Rationale:**

1. A hybrid model is used to leverage the strengths of both machine learning and deep learning.

2. While Random Forest is robust with structured tabular data, LSTM excels at capturing temporal patterns. Combining both helps smooth predictions and account for both spatial and time-based variability.

- **Implementation:**

1. Predictions from both the Random Forest and LSTM models are computed independently.

2. The final hybrid output is calculated as the average of both models' predictions for each test sample.

- **Benefits of Hybridization:**

1. Reduces variance and bias associated with single-model predictions.

2. Improves robustness and generalization of the forecast.

3. Enhances model stability, especially in the presence of seasonal or location-based fluctuations.

6. Model Evaluation

- **Metrics Used:**

1. MAE (Mean Absolute Error)
2. MSE (Mean Squared Error)
3. RMSE (Root Mean Squared Error)
4. R^2 Score (Coefficient of Determination)

- **Comparison:**

1. All three models (RF, LSTM, and Hybrid) are evaluated on test data.
2. Hybrid model shows improved performance over standalone models.

7. Visualization

- **Random Forest Plot:**

1. Visualized using Plotly to show predicted vs actual prices.
2. Enhances understanding of model accuracy across samples.

- **District-Wise Prediction Map:**

1. Interactive map displays predicted garlic prices across various Indian districts.

2. Color-coded markers provide a visual comparison between high and low price regions.

- **Crop Recommendation Map:**

1. Users can select Indian states and months to view optimal crops.

2. Designed for farmers and agricultural planners to make informed sowing decisions.

3. Each marker on the map includes crop name recommendations for that location.

APPENDIX-B

ENCLOSURES

IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Ref No : IJCRT/Vol 13/ Issue 1 / 194

To,
Shashank J K

Subject: Publication of paper at International Journal of Creative Research Thoughts.

Dear Author,

With Greetings we are informing you that your paper has been successfully published in the International Journal of Creative Research Thoughts - IJCRT (ISSN: 2320-2882). Thank you very much for your patience and cooperation during the submission of paper to final publication Process. It gives me immense pleasure to send the certificate of publication in our Journal. Following are the details regarding the published paper.

About IJCRT	:	Scholarly open access journals, Peer-reviewed, and Refereed Journals, Impact factor 7.97 (Calculate by google scholar and Semantic Scholar | AI-Powered Research Tool) , Multidisciplinary, Monthly, Indexing in all major database & Metadata, Citation Generator, Digital Object Identifier(DOI) | UGC Approved Journal No: 49023 (18)
Registration ID	:	IJCRT_275402
Paper ID	:	IJCRT2501194
Title of Paper	:	Proactive disaster detection
Impact Factor	:	7.97 (Calculate by Google Scholar) | License by Creative Common 3.0
Publication Date	:	09-January-2025
DOI	:	
Published in	:	Volume 13 | Issue 1 | January 2025
Page No	:	b635-b646
Published URL	:	http://www.ijcrt.org/viewfull.php?&p_id=IJCRT2501194
Authors	:	Shashank J K, Parth, Sagar M, Teja Reddy M, Dr. Manjunath K V
Notification	:	UGC Approved Journal No: 49023 (18)

Thank you very much for publishing your article in IJCRT.

Editor In Chief
International Journal of Creative Research Thoughts - IJCRT
(ISSN: 2320-2882)

International Journal of Creative Research Thoughts (IJCRT) ISSN 2320-2882 IJCRT

PLAIGARISM REPORT

SUSTAINABLE DEVELOPMENT GOALS

1. Goal 1: No Poverty

By providing accurate garlic price forecasts, the system empowers farmers to make informed selling decisions, thereby stabilizing their income and reducing vulnerability to market shocks. This contributes to poverty alleviation in rural communities where agriculture is a primary livelihood.

2. Goal 2: Zero Hunger

Stable garlic pricing helps ensure fair returns for farmers and affordability for consumers. This balance supports food security by sustaining crop production and access to nutritious food at reasonable prices.

3. Goal 8: Decent Work and Economic Growth

The project promotes sustainable agricultural practices and enhances economic resilience among farmers by introducing data-driven decision-making tools. It encourages productivity, reduces exploitation by middlemen, and facilitates more efficient supply chain planning.

4. Goal 9: Industry, Innovation and Infrastructure

This project exemplifies innovation in agriculture through the use of hybrid AI models (Random Forest + LSTM), integrating climatic and regional data for more precise forecasting. It contributes to smart agricultural infrastructure and opens avenues for agri-tech growth in rural areas.

5. Goal 10: Reduced Inequalities

With district-level predictions and localized insights, the system ensures that farmers across different regions, especially those in less-developed districts, receive tailored support. This helps bridge the rural-urban and regional divide in market access and technological benefits.

6. Goal 12: Responsible Consumption and Production

Price forecasting allows for better planning in both farming and distribution. This minimizes overproduction and wastage, ensuring resources like water, fertilizers, and energy are used more efficiently.

7. Goal 13: Climate Action

The integration of climate variables (rainfall, temperature) into the model aligns the project with climate-smart agriculture. It promotes adaptation to changing weather patterns and supports farmers in planning their cropping activities accordingly.

8. Goal 17: Partnerships for the Goals

The project encourages collaboration between academic researchers, agricultural experts, and governmental or non-governmental institutions. Such partnerships are vital for scaling AI-driven solutions across India and beyond.