

Mini Project Report
on

**MACHINE LEARNING BASED
SHORT-TERM FORECASTING OF
ORANGE AND COTTON CROP
PRICES IN CONTEXT OF INDIAN
MARKET**

**Submitted
BY**

**Humanshu Gajbhiye - 147
Susrut Patole - 169
Geetika Mahant - 102
Neha Thakur - 110**

**Under the Guidance of
Prof. (Dr.) Nileshsingh V. Thakur**



October, 2023-24

**Department of Computer Technology
YESHWANTRAO CHAVAN COLLEGE OF
ENGINEERING, Nagpur**

(An Autonomous Institution Affiliated to Rashtrasant Tukadoji Maharaj Nagpur University)

**YESHWANTRAO CHAVAN COLLEGE OF ENGINEERING
NAGPUR**

(An Autonomous Institution Affiliated to Rashtrasant Tukadoji Maharaj Nagpur University)

Department of Computer Technology

(2023-24)

Certificate

This is to certify that the Mini Project Report titled “ **title of the Mini Project** ” is submitted towards the partial fulfillment of requirement of Mini Project course in VII Semester, B.E.(Computer Technology).

Submitted by:

Ms Neha Thakur (RollNo: 110)

Ms Geetika Mahant (RollNo: 102)

Mr Susrut Patole (RollNo: 169)

Mr Humanshu Gajbhiye (RollNo: 147)

is approved.

Project Guide

Prof. (Dr.) Nileshsingh V. Thakur

Project Coordinator

Smita R. Kapse

Head, Department of Computer Technology

Dr. R.D.Wajgi

Date: _____

Place: _____

ACKNOWLEDGEMENT

We would like to thank our guide Dr Nileshsingh V. Thakur for thorough guidance in the project. We are extremely grateful and indebted to them for their expert, sincere, valuable guidance and encouragement which was of immense help to us.

We would like to express our sincere gratitude to Dr R.D.Wajgi, Head, Department of Computer Technology for her constant encouragement towards the successful completion of our work.

We wish to express our sincere thanks to Dr U.P. Waghe, the Principal of our college for providing us with all the necessary facilities and the infrastructure without which we would not have been able to complete our project successfully.

We would also like to thank our Project Coordinator Prof. S. R. Kapse for their continuous guidance owing to which the project could take shape.

We would like to thank technical assistant Mrs. B. H. Kulkarni for providing necessary technological support. Last, but not the least, we would like to thank all the faculty members and non-teaching staff members who helped us despite of their busy schedule

Abstract

This project endeavors to construct a Machine Learning-based model tailored to the short-term forecasting of orange and cotton crop prices in the Indian market. Leveraging an extensive dataset provided by the Indian Government, the project initiates with a pivotal phase of Exploratory Data Analysis (EDA). This process serves to address the challenges posed by missing data and, importantly, to unveil latent patterns inherent in the dataset.

A key aspect of the project is data preparation and partitioning. The dataset is meticulously split into training, testing, and validation sets to facilitate the subsequent model development and evaluation phases.

The forecasting methodology encompasses a spectrum of regression-based algorithms, including Artificial Neural Networks (ANN), Auto Regressive Integrated Moving Average (ARIMA), Logistic Regression (LR), and Support Vector Regression (SVR). Additionally, an ARIMA model is incorporated for comparative assessment. These algorithms serve as the workhorses for our price prediction model.

The analysis extends beyond mere algorithm selection. Regional and seasonal dynamics are incorporated into the model's framework. The dataset is categorized into urban and rural segments to account for regional variations. Notably, model performance and accuracy are evaluated at different geographical granularities, including the national, state, and city levels, to provide a comprehensive understanding of the price prediction capability.

Table of Contents

Title	Page No.
1.0 Introduction.....	
2.0 Aim & Objectives.....	
3.0 Literature Review.....	
4.0 Proposed methodology	
5.0 Results and Discussion.....	
6.0 Conclusion and Future Scope.....	
References.....	

1.0 Introduction

I. Background:

1. Agricultural Significance in India:

The agricultural sector in India is the backbone of the economy, supporting the livelihoods of a substantial portion of the population. The pricing of key crops, such as oranges and cotton, not only impacts farmers directly but also influences the overall economic landscape. Understanding and predicting the fluctuations in crop prices is paramount for informed decision-making and sustainable agricultural practices.

2. Dataset Overview:

The project leverages a comprehensive dataset provided by the Indian Government, offering a historical perspective on crop prices across diverse regions and seasons. This dataset serves as the bedrock for the development and validation of machine learning models for short-term forecasting.

II. Exploratory Data Analysis (EDA):

1. Data Cleaning and Imputation:

Prior to analysis, a meticulous EDA is conducted to address missing data and ensure the integrity of the dataset. Techniques such as imputation and data cleaning are employed to enhance the quality of the dataset, laying the groundwork for accurate and reliable forecasting models.

2. Pattern Identification:

The EDA phase aims to uncover inherent patterns within the data. Through statistical analyses and visualization techniques, the project seeks to discern trends, cyclical patterns, and any anomalies that may be indicative of the underlying dynamics governing crop prices.

III. Geographic Segmentation:

1. Urban and Rural Tiers:

Recognizing the diversity of the Indian agricultural landscape, the dataset is segmented into tiers representing urban and rural cities. This segmentation allows for a nuanced examination of how regional disparities influence pricing dynamics, providing valuable insights for targeted interventions.

2. State-Level Analysis:

The project delves into state-level dynamics, acknowledging the varied agricultural practices and market conditions across different states in India. This analysis contributes to a more localized understanding of the factors shaping crop prices.

3. City-Specific Insights:

Further granularity is achieved by focusing on specific cities, allowing the project to unravel localized trends and intricacies. This micro-level analysis facilitates a fine-tuned evaluation of model accuracy and aids in tailoring strategies to address city-specific challenges.

IV. Model Development and Selection:

1. Regression-Based Algorithms:

Four candidate algorithms—Artificial Neural Network (ANN), AutoRegressive Integrated Moving Average (ARIMA), Logistic Regression (LR), and Support Vector Regression (SVR)—are chosen for their capacity to handle regression tasks. The inclusion of ARIMA serves as a benchmark, offering a basis for comparison against machine learning approaches.

2. Training-Testing-Validation Sets:

The dataset is strategically partitioned into training, testing, and validation sets to enable rigorous evaluation of model performance. This ensures that the selected algorithms are not only accurate in their predictions but also generalize well to unseen data.

V. Comparative Analysis:

1. National Overview:

The project begins with a broad national analysis, providing a holistic understanding of crop price forecasting across the entire country. This sets the stage for subsequent, more detailed examinations at lower geographic levels.

2. State-Level Performance:

The effectiveness of the forecasting models is assessed at the state level, allowing for the identification of regional variations and the validation of model robustness in diverse agricultural contexts.

3. City-Specific Accuracy:

The project concludes with an exploration of city-specific insights, shedding light on how well the models perform in distinct urban and rural environments. This localized evaluation is crucial for tailoring strategies to address specific market conditions.

In summary, this project is a comprehensive exploration of machine learning-based short-term forecasting for orange and cotton crop prices in the Indian market. By incorporating geographic segmentation and seasonal considerations, the aim is not only to provide accurate predictions but also to offer actionable insights for stakeholders at different levels of the agricultural supply chain. Through this multifaceted approach, the project contributes to the ongoing efforts to enhance the resilience and sustainability of India's agricultural sector.

2.0 Aim & Objectives

1. Aim :

To devise a Machine Learning based short-term forecasting mechanism of orange and cotton crop prices in context of Indian market.

2. Objectives :

The objective of this project is to create a machine learning-driven model for predicting the prices of orange and cotton crops within the Indian market.

The model should be able:

1. Conduct exploratory data analysis to address missing data in the dataset.
2. Identify patterns within the data, if any, that can inform the price prediction model.
3. Devise a method to determine the appropriate number of data records for training the model.
4. Test and validate the model to ensure its accuracy and reliability.
5. Explore various training and learning algorithms to identify the most suitable ones for the task.
6. Build a machine learning-based model for predicting the prices of orange and cotton crops in the Indian market.

3.0 Literature Review

1. General Work

Casper Solheim Bojer. (2022) [1] The research paper titled "Understanding machine learning-based forecasting methods: A decomposition framework and research opportunities" by C.S. Bojer offers insights into the growing interest in machine learning-based forecasting methods. It highlights their successful applications in forecasting competitions like the M4, M5, and Kaggle, challenging traditional statistical methods. The paper underscores the need for a systematic understanding of these methods due to their complexity, with diverse components including preprocessing, feature engineering, and ensembling. It introduces a framework for regression-based machine learning forecasting, aiming to standardize terminology and abstraction for researchers. The framework is applied to map and compare winning solutions from the M5 Uncertainty competition, emphasizing the complexity of strategies employed. Additionally, the paper identifies research areas such as cross-learning, feature engineering, and cross-validation, signaling potential advancements in the field of machine learning-based forecasting.

Bontempi et al. [2] The research paper titled "Machine Learning Strategies for Time Series Forecasting" authored by Gianluca Bontempi, Souhaib Ben Taieb, and Yann-Aël Le Borgne provides a comprehensive overview of the role of machine learning techniques in the field of time series forecasting. The paper addresses the growing demand for accurate forecasting across various domains and the transition from traditional linear statistical methods, such as ARIMA models, to machine learning models. It highlights the importance of formalizing one-step forecasting problems as supervised learning tasks and discusses the use of local learning techniques for handling temporal data. The paper explores various strategies for multi-step forecasting, including the Recursive, Direct, and Multiple Output (MIMO) strategies, each with its advantages and limitations. Additionally, the paper introduces the concept of the DIRMO strategy, which combines elements of both Direct and MIMO strategies, providing a flexible approach to multi-step forecasting. The authors also delve into the use of local learning techniques for improving long-term predictions within these strategies.

Li et al. (2023) [3] The ARIMA (autoregressive integrated moving average) model is a widely respected and traditional linear model for time series prediction, known for its effectiveness in various domains. However, its vulnerability to noisy data, leading to instability and reduced performance, has not received sufficient attention. In this study, we introduce the spARIMA framework to enhance the reliability of time series prediction. We implement a sequential training approach in batches, taking into account noise levels and their impact on accurate modeling, with the goal of minimizing noise-related disruptions during training. To leverage the advantages of self-paced learning (SPL), spARIMA integrates a differential prediction model into the ARIMA framework.

Tealab et al. (2017) [4] Since basic linear time series models often fall short in fully explaining certain aspects of economic and financial data, it becomes crucial to categorize time series based on their linearity characteristics when conducting

forecasts. This approach ensures that linear time series continue to be the focal point of both academic and practical research. Predicting nonlinear time series, especially those containing inherent moving average components, using computational intelligence methods like neural networks can be challenging, given that real-world time series often exhibit dynamic behavior, autoregressive elements, and inherited moving average terms. Research focusing on the prediction of nonlinear time series with moving average components is relatively rare. In this study, we demonstrate that conventional neural networks struggle to effectively capture the behavior of nonlinear time series with moving average terms, resulting in subpar forecasting performance.

Yang a et al. [5] It explores the domain of network traffic forecasting, emphasizing the challenges faced by traditional linear and non-linear models. The proposed method integrates Simulated Annealing (SA)-optimized AutoRegressive Integrated Moving Average (ARIMA) with Back Propagation Neural Network (BPNN) to enhance prediction accuracy. The research highlights the contribution of the approach in extracting both linear and non-linear features from network traffic data, addressing limitations in existing models. Notable advances include improved network congestion control mechanisms and enhanced accuracy in fields relying on precise network traffic predictions. Real-world datasets validate the superiority of the proposed method, showcasing its effectiveness against single and hybrid prediction models. The study also conducts sensitivity analysis, demonstrating the method's stability across varying temporal spans. Future directions involve considering user behavior impact, exploring additional factors, and applying the method in Software-Defined Networking (SDN) systems for automated adjustments based on predicted trends. In conclusion, the proposed SA-optimized ARIMA–BPNN method proves promising for various applications, offering superior accuracy in network traffic forecasting and addressing non-linear complexities.

Castle et al. [6] The paper investigates robust approaches to forecasting, particularly focusing on a new class of robust devices in contrast to equilibrium-correction models. The authors explore the forecasting properties of these robust methods in the face of empirical challenges such as measurement errors, impulses, omitted variables, unanticipated location shifts, and incorrectly included variables. The study applies these robust approaches to forecast US GDP using autoregressive models and models with factors extracted from a large dataset of macroeconomic variables. The empirical analysis spans both the Great Recession and an earlier, more stable period, providing insights into the performance of robust forecasting devices under different economic conditions. The authors compare the proposed robust devices with traditional models like autoregression and factor models, evaluating their effectiveness in mitigating forecast failures caused by shifts in the economic environment. The results suggest that the robust devices, especially those with smoothing techniques, show promise in improving forecast accuracy, particularly in the face of significant economic shifts. The empirical findings contribute to the understanding of robust forecasting methods, emphasizing their potential advantages in handling real-world forecasting challenges.

Peter Zhang (2002) [7] The research paper explores time series forecasting by

proposing a hybrid methodology that combines the Autoregressive Integrated Moving Average (ARIMA) model with Artificial Neural Networks (ANNs). ARIMA has been a popular linear model in time series forecasting, while recent studies suggest that ANNs offer a promising alternative with flexible nonlinear modeling capabilities. The hybrid approach aims to leverage the strengths of both models to enhance forecasting accuracy. The motivation behind this combination arises from the challenges in determining whether a time series is linear or nonlinear, the presence of both linear and nonlinear patterns in real-world data, and the recognition that no single method is universally superior. The paper argues that the hybrid model addresses these issues and presents experimental results using real datasets to support its effectiveness. The literature review highlights the importance of model combination for improved forecasting accuracy, drawing on the well-known M-competition and emphasizing the benefits of combining dissimilar models to reduce generalization variance and model uncertainty. The authors position their work within the context of ongoing efforts to enhance time series forecasting methodologies.

2. Related Work

Zhanga et al. [8] In this study, a novel neural network model called the QR-RBF model is introduced, which combines the principles of quantile regression and radial basis functions. This model offers two key capabilities:

1. It characterizes the distribution of soybean price ranges through the use of quantile regression models.
2. It approximates the nonlinear aspects of soybean prices using RBF neural networks.
3. To optimize the parameters of the QR-RBF neural network model, the research proposes a hybrid algorithm called GDGA. GDGA combines the global search capabilities of the genetic algorithm with the local search capabilities of gradient descent.

The analysis of monthly domestic soybean price data in China led to the following conclusions:

1. The hybrid GDGA algorithm performs effectively in optimizing the model.
2. The factors influencing soybean prices vary depending on the price level.
3. Money supply and port distribution prices for imported soybeans are significant factors across multiple quantiles.
4. Domestic soybean production and the consumer confidence index are significant primarily for lower quantiles.
5. Soybean import volume and the consumer price index are significant mainly for higher quantiles.

Tella et al. [9] This paper delves into the forecasting of food prices, focusing on the interdependence of individual models for corn, soybeans, and wheat. The authors address the strong historical correlations among food prices, examining whether the accuracy of individual models can be enhanced by considering their cross-dependence. The study covers the period of 2008–2014, marked by instability, and employs robust methodologies and recursive schemes to deal with potential breaks in the data. The investigation is grounded in the understanding that these three food prices exhibit significant correlations, and the paper explores whether joint

modeling could lead to improved forecasting. The research is particularly timely given the common behaviors observed in the prices of these commodities, influenced by factors such as the growth of emerging economies, biofuel demand, and macroeconomic developments. The paper aims to contribute insights into conditional forecasting models for food prices, considering the effects of external variables.

Sabu et al. (2020) [10] In this research, a combination of time-series and machine learning models is employed to predict monthly arecanut prices in the Kerala region. The dataset comprises price data spanning from 2007 to 2017, serving as a benchmark for evaluating the performance of three models: SARIMA, Holt Winters Seasonal approach, and the LSTM neural network. The model that exhibited the best fit to the data turned out to be the LSTM neural network model. Fluctuations in the prices of agricultural products have a detrimental impact on a country's GDP. These price changes also carry emotional and financial consequences for farmers who invest years of hard work, only to see their efforts sometimes yield disappointing results. The agricultural supply chain may benefit from price prediction by using it to better manage and minimize the risk of price changes. Predictive analytics is anticipated to address issues facing the average person as a result of the decline in agricultural productivity brought on by uncertain climatic conditions, global warming, etc. India produces a significant amount of arecanuts, with Kerala coming in second place. Due to price volatility and climatic change, farmers in Kerala have recently shifted from arecanut production to other crops.

Pantazi et al. (2016) [11] The study focuses on predicting wheat yield variations within a field by leveraging high-resolution multi-layer information from on-line proximal soil sensing and satellite imagery. Traditional linear models for yield prediction based on limited soil samples have shown variable spatial and temporal results. The paper explores the application of computational intelligence, including artificial neural networks (ANNs) and self-organizing maps (SOMs), to integrate data from soil sensors and crop growth characteristics for more accurate predictions. Previous non-linear modeling approaches have been limited by low accuracy and the lack of fusion of high-resolution soil data with remotely sensed crop growth. The research introduces an algorithm based on supervised SOMs to address these limitations, aiming to provide a flexible and accurate framework for predicting wheat yield and visualizing correlations between soil parameters, crop characteristics, and yield. The results indicate that the supervised Kohonen Network (SKN) model outperforms other models, achieving an overall accuracy of 81.65% in cross-validation. The study emphasizes the need for a unified framework, incorporating high-resolution data fusion for improved yield predictions in precision farming.

Kyriazi et al. [12] They present an innovative forecasting approach termed "adaptive learning forecasting," which supports both forecast averaging and learning from forecast errors. They delve into its theoretical characteristics and illustrate that, in certain scenarios, it enhances the mean squared error (MSE) a priori. The research also reveals that the learning rate, contingent on previous forecast errors, follows a nonlinear pattern. This methodology finds wide-ranging applications and can even enhance the MSE when compared to the most basic benchmark models. To exemplify the strategy's applicability, the researchers employ data on agricultural prices for

various agricultural commodities and real GDP growth figures for relevant countries. They consider numerous forecasting models, including both univariate and bivariate ones related to output and productivity. This is particularly important because agricultural price time series are typically short and display irregular cyclical patterns that correlate with economic performance and productivity. Their findings confirm the effectiveness of the new approach and the predictability of agricultural price movements.

Shiferaw(2022) [13]They present an innovative forecasting approach termed "adaptive learning forecasting," which supports both forecast averaging and learning from forecast errors. They delve into its theoretical characteristics and illustrate that, in certain scenarios, it enhances the mean squared error (MSE) a priori. The research also reveals that the learning rate, contingent on previous forecast errors, follows a nonlinear pattern. This methodology finds wide-ranging applications and can even enhance the MSE when compared to the most basic benchmark models. To exemplify the strategy's applicability, the researchers employ data on agricultural prices for various agricultural commodities and real GDP growth figures for relevant countries. They consider numerous forecasting models, including both univariate and bivariate ones related to output and productivity. This is particularly important because agricultural price time series are typically short and display irregular cyclical patterns that correlate with economic performance and productivity. Their findings confirm the effectiveness of the new approach and the predictability of agricultural price movements.

Scott Armstrong et al. [14]This research paper presents the Golden Rule of forecasting as its central and unifying concept. Being conservative when forecasting is the Golden Rule. A cautious forecast is in line with our collective understanding of the past and present. To exercise prudence, forecasters should actively search for and utilize all relevant information pertaining to the matter, including knowledge about techniques that have demonstrated effectiveness in the specific context. The Golden Rule logically leads to the deduction of twenty-eight rules. An assessment of the evidence turned up 105 publications with experimental comparisons; 102 of them are in favor of the recommendations. The average forecast error increased by more than two-fifths when a single guideline was disregarded. The Golden Rule should always be followed, especially in complex and unclear situations when bias is likely.

Yuehjen E. Shao , Jun-Ting Dai(2018) [15] The supply of three major food crops, namely rice, wheat, and corn, is rapidly diminishing due to global climate change, limited land availability, and a rapidly growing population. Predicting the prices of these essential food crops has become a subject of considerable attention. While numerous feature selection techniques (FSMs) have been developed for integrated forecasting models, a significant challenge arises from the absence of future values for these critical factors, making predictions using these factors impossible. In this study, an Autoregressive Integrated Moving Average (ARIMA) model is employed as the foundational statistical method (FSM) within computational intelligence (CI) models to forecast the prices of these three vital food crops. The ARIMA model is chosen because it can identify important self-predictor variables with calculable future values. The proposed integrated forecasting models encompass not only

ARIMA but also include Support Vector Regression (SVR), Multivariate Adaptive Regression Splines (MARS), and Artificial Neural Networks (ANNs). The research conducts a comparative analysis and discussion on the prediction accuracies of the proposed integrated model in contrast to ARIMA, ANN, SVR, MARS, and other existing models.

Dimpfla et al. [16] This paper delves into the dynamic relationship between spot and futures prices of various agricultural commodities, such as corn, wheat, soybeans, soybean meal and oil, feeder and live cattle, as well as lean hogs. The primary focus is on determining which markets play a predominant role in price discovery for these commodities. The authors employ a unique information share methodology and find that the spot market overwhelmingly leads price discovery, with the futures market contributing less than 10% in the Hasbrouck sense. This study challenges the notion of adverse effects of futures speculation on commodity prices in the long run, emphasizing the robustness of spot market dynamics. The paper contributes to the broader literature on price discovery, challenging the prevailing view that futures markets typically dominate this process across various assets. The research provides insights into the specific context of agricultural commodity markets and their susceptibility to speculation, contributing valuable perspectives to both academic and practical discussions on market dynamics and regulation.

Zoua et al. [17] The paper conducts a comprehensive investigation into the predictive abilities of ARIMA, artificial neural networks (ANNs), and a combined model in forecasting Chinese wheat prices. Time series forecasting is underscored as a critical domain, especially in the absence of a well-defined data generating process or explanatory model. The literature review reveals divergent findings in the comparison of traditional models with ANN, with some studies favoring neural networks' superiority in forecasting, while others report mixed or conflicting results. Emphasizing the effectiveness of combining different models, the paper notes the M-competition's conclusion that forecast combination often leads to improved performance. The study's twofold objective includes assessing the forecasting performance of the mentioned models for Chinese food grain markets and examining their ability to predict turning points, providing valuable insights into nonlinear methods' efficiency in enhancing forecasting outcomes.

Cheung et al. [18] The research paper proposes the application of a novel Clustered 3D-CNN model to enhance the prediction of crop future prices. Traditional statistical methods, such as ARIMA, often fall short in capturing the non-stationary patterns and multidimensional influences on crop prices. The paper highlights the complex nature of factors affecting crop production and prices, including environmental changes and economic factors. The proposed Clustered 3D-CNN model is introduced as a solution, leveraging its ability to handle non-stationary data and learn non-linear patterns. The study compares the performance of the proposed model with the ARIMA model through experiments using real-world datasets. The results indicate that the Clustered 3D-CNN model outperforms ARIMA in terms of Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE). The research emphasizes the significance of considering various factors, including environmental and economic variables, for more accurate crop

price predictions. The implications suggest that the proposed model could aid decision-makers in predicting crop price trends and developing strategic plans to address food insecurity issues. The paper concludes by discussing the limitations of traditional statistical models and the potential of the Clustered 3D-CNN model in agricultural price forecasting.

V. Sneha; V. Bhavana (2023) [19] Sugarcane is a critical commercial crop in India, and the agricultural sector has benefited significantly from advancements in technology, particularly in the realm of machine learning (ML). This cutting-edge technology has proven invaluable to farmers by providing comprehensive recommendations and insights to enhance crop quality and productivity while reducing farming losses. Before planting sugarcane, accurate estimations of yield and market prices are essential for making profitable decisions.

Machine learning techniques such as Decision Tree Regressor, Multi Linear Regression, Random Forest, Adaboost Regressor, and Lasso Regression are employed to predict sugarcane yield. Additionally, an ARIMA model is used to forecast sugarcane prices. Several factors, including historical sugarcane yields in specific locations, rainfall patterns, and the state in which sugarcane is cultivated, all contribute to the accurate prediction of sugarcane yields. Meanwhile, sugarcane price predictions are based on time series analysis of historical price data.

B Chaitra,K Meena,(2023) [20] From the agricultural industry's perspective, the market price of a specific crop reflects its current demand. To facilitate monitoring of agricultural prices, the Council of Agriculture (COA) has established an official website that provides open data on daily market prices across more than 15 local markets in Taiwan, encompassing over 100 different crops. In parallel, the Institute for Information Industry (III) has introduced the smart agri-management platform (S.A.M.P.), a comprehensive cloud-based solution for agri-business. This research paper focuses on the development of a crop price forecasting service within S.A.M.P., inspired by the availability of open data on crop prices. The service automatically retrieves historical price data from the official website, utilizing it as a training dataset, and employs well-established time series analysis algorithms for price forecasting. The study employs three techniques: Partial Least Square (PLS), Artificial Neural Network (ANN), and Autoregressive Integrated Moving Average (ARIMA). Moreover, to explore nonlinear relationships within historical prices, the researchers introduce Response Surface Methodology (RSM) into PLS, resulting in a novel approach called RSMPLS. The performance of these four algorithms is compared using price data obtained from the First Fruit and Vegetable Wholesale Market in Taipei, focusing on crops like cauliflower, watermelon, bok choy, and cabbage. Based on experimental data, PLS and ANN exhibit smaller percentage errors in their forecasting accuracy.

Ndunagu et al. [21] India's economy relies heavily on agriculture. India has a sizable chunk of arable land. Numerous crops are in high demand abroad. Other than software, one of the main exports from India is food. But because of unpredictable weather patterns, a lack of human resources, and poor crop selection, farmers are unable to turn a profit. The agriculture profession is gradually losing relevance as a result of urbanization as well. For accurate Yield and Crop Price Prediction. This

issue has led numerous researchers to identify the challenges and employ a diverse range of machine learning methods, including Autoregressive Integrated Moving Average (ARIMA), Decision Trees, Long Short-Term Memory (LSTM), K Nearest Neighbors (KNN), and others. These techniques assist farmers in making informed decisions regarding crop selection, ultimately enabling them to achieve optimal yields and profits.

Xiong et al. [22] This paper addresses the challenging task of interval forecasting for agricultural commodity futures prices, particularly focusing on the Chinese futures market. With the significant growth in China's agricultural commodity futures markets, accurate forecasting is crucial for risk management and speculation. The study proposes a novel method, VECM-MSVR (Vector Error Correction Model-Multi-Output Support Vector Regression), within the established 'linear and nonlinear' modeling framework. This method aims to capture both linear and nonlinear patterns in futures prices by combining VECM and MSVR. The interval forecasting approach provides a range of values, considering the inherent volatility and complexity of agricultural commodity prices. The authors compare the proposed method with several benchmarks, including VECM, SSVR, MSVR, ARIMA-MSVR, and ARIMA-ANN, using Chinese cotton and corn futures prices as experimental data. The results demonstrate that VECM-MSVR consistently outperforms other methods, showcasing its potential as a promising alternative for forecasting interval-valued agricultural commodity futures prices. The paper contributes to the literature by exploring interval forecasting, an underexplored area in agricultural commodity futures, and providing empirical evidence from the Chinese market.

Xiaoli L et al. [23] This paper investigates the dynamics of corn, soybean, and wheat futures markets from 2004 to 2013, a period marked by significant price volatility. Contrary to prevailing hypotheses, the study finds that explosive price episodes or "bubbles" occurred only around two percent of the time, and when they did, they were short-lived and of modest magnitude. Notably, commodity index trader positions did not significantly influence positive bubble occurrences, challenging the widely held "Masters Hypothesis." The study also examines the impact of speculation, revealing that it either has minimal or negative effects on price explosiveness. The research sheds light on the complex interplay of fundamental factors, such as inventories, exports, and economic growth, in driving grain futures markets. In summary, this work provides nuanced insights into the factors contributing to price explosiveness and challenges prevailing narratives about the role of speculation in agricultural futures markets.

Ref. No	Work Carried Out	Methodology Used	Evaluation Parameters	Datasets Used	Claims By Author	Our Findings
1	Mapping and comparing ML methods	Framework development, method mapping, performance assessment.	Mapping, comparison, ablation testing for evaluation.	Kaggle	Common language, improved forecasting methods.	Complexity understanding, needs further exploration.
2	Comparative analysis of time series forecasting methods.	Application of various machine learning strategies.	Accuracy, robustness, and computational efficiency assessment.	Diverse time series datasets from different domains.	Superior forecasting accuracy with machine learning approaches.	Insufficient Local Learning Exploration ,Limited Adaptive Model Integration
3	Extended ARIMA, introduced spARIMA, diversity selection.	Sequential training, diversity selection in spARIMA.	Generalization, efficiency on noisy data.	12 datasets for time series forecasting.	spARIMA enhances time series forecasting.	Improvements for complex trend information.
4	Assessment of time series forecasting methods.	Evaluate neural networks for time series.	Efficiency in forecasting nonlinear time series.	Nonlinear time series simulation dataset.	Common neural networks lack efficiency.	Need models for nonlinear time series.
5	Integration of ARIMA and BPNN, Implementation of SA optimization, Prediction of network traffic.	SA-optimized ARIMA–BPNN hybrid, ARIMA captures linear features, BPNN refines predictions, handling non-linearities.	MAE, RMSE, MAPE metrics, Comparative analysis with existing models, Real-world network traffic datasets.	Historical network traffic data, Two sampling points in WIDE project.	Improved prediction accuracy, Outperforms traditional models, Enhanced network congestion control.	Impact of user behavior, Exploration of additional factors, Application in SDN systems.
6	New adaptive learning forecasting methodology	Forecast averaging and error learning, with the learning rate based on past forecast errors being non-linear. MZ Regression	Mean Squared Error (MSE). Mean Absolute Error (MAE)	Data on European economies obtained from Eurostat	Adaptive learning enhances forecasting accuracy for wide range of scenarios. Improving upon bench mark models.	Learning rate parameter time-varying not utilized. Joint optimization of averaging weights and the learning rate.
7	Proposed hybrid ARIMA-ANN methodology.	Combined linear ARIMA with nonlinear ANN.	Assessed forecasting accuracy improvement.	real-world datasets.	Hybrid model enhances accuracy.	Addressing time series complexity gaps.
8	Forecasting soybean price in China.	QR-RBF neural network with GDGA.	Forecasting performance, convergence.	Monthly domestic soybean price in China.	Improved soybean price forecasting model.	Influential factors stability, comprehensive exploration needed.

Ref. No	Work Carried Out	Methodology Used	Evaluation Parameters	Datasets Used	Claims By Author	Our Findings
9	Examining food price correlations.	Utilizing joint modeling approaches.	Assessing forecast improvements.	Quarterly prices of corn, soybeans, and wheat.	Improved accuracy with joint models.	Limited updating schemes explored.
10	Prediction of Arecanut Prices in Kerala	SARIMA, Hot-Winters Seasonal Method, LSTM Neural Network	Root Squared Mean Error (RMSE)	Arecanut price data spanning from 2007 to 2017.	LSTM neural network model outperformed other methods.	LSTM most effective model for arecanut price prediction. Data availability for LSTM training is a limitation.
11	Predicting wheat yield variations using sensors, satellite imagery.	Utilized self-organizing maps (SOMs), artificial neural networks (ANNs).	Assessed accuracy through cross-validation and independent validation.	Combined on-line soil sensing, satellite imagery, crop growth indicators.	SKN model outperforms, achieving 81.65% overall accuracy.	Need for unified data fusion, high-resolution soil-crop correlation.
12	Forecasting approaches and contrasted them with equilibrium correction models.	Developed a framework. Mapped and assessed forecasting methods.	Forecast performance through mapping, comparison, and ablation testing.	Dataset from Stock and Watson	Robust devices outperformed equilibrium correction models in the presence of location shifts.	Recursive updating and combining information not utilized.
13	Modeling the dynamic volatility of East African tea crop prices	Error distributions and scedastic functions for in-sample value-at-risk(VaR). MCMC approach.	GARCH, EGARCH models (volatility and VaR). Factors fat tails, asymmetry, and regime switching.	Monthly tea auction prices in USD from the Mombasa auction.	EGARCH model exhibited superior performance in estimating volatility.	Lack of high-frequency time series data. Multivariate GARCH models not explored.
14	Formulation of Golden Rule of forecasting guidelines.	Deduction of guidelines, evidence review.	Accuracy, adherence to Golden Rule.	Not Specified	Golden Rule of forecasting enhances forecast accuracy.	Unexplored Method Validation, Non-expert Decision Evaluation
15	Integrated ARIMA with CI for food crop price prediction.	ARIMA as a feature selector in CI models.	MAE, RMSE, MAPE used for comparison.	Real dataset: rice, wheat, corn prices (1990-2015).	Proposed model outperforms in food crop price forecasting.	Lack of focus on feature selection in crop price prediction, Limited research addressing unknown variables in crop price prediction.

Ref. No	Work Carried Out	Methodology Used	Evaluation Parameters	Datasets Used	Claims By Author	Our Findings
16	Analyzing agricultural commodity price discovery.	Variance decomposition.	Information share and variance decomposition.	Specific datasets not mentioned.	Spot markets dominate agricultural price discovery.	Limited exploration of speculative impact.
17	Comprehensive model comparison	ARIMA, artificial neural networks	Forecasting accuracy, turning points	Chinese wheat prices	Improved forecasting with combination	Limited turning point analysis
18	Developed CNN model for crop prices.	Clustered 3D-CNN for price prediction.	MAPE, RMSE, MAE for model comparison.	Data will be made available on request.	CNN outperforms ARIMA in prediction.	Variable choice, computational cost optimization.
19	Sugarcane yield and price forecasting.	Machine learning algorithms, ARIMA model.	Accuracy, precision, and model comparison.	Historical yield, rainfall, price data.	ML improves sugarcane yield forecasting accuracy.	Limited discussion on model limitations.
20	Crop prediction using machine learning models.	ARIMA model along with NARY, Decision Tree, Linear Regression, Random Forest..	Accuracy, Precision, Recall, F1 Score.	Historical data of various agricultural parameters.	Improved crop yield and profit predictions.	Limited mention of climate impact.
21	Comparative price analysis of food items.	ARIMA model for price forecasting.	Forecast accuracy and price trends.	National Bureau of Statistics data.	ARIMA is effective for price forecasting.	Unclear measures to address high prices.
22	Interval forecasting method development.	Combined VECM and MSVR.	Quantitative assessment, benchmark comparisons.	Chinese cotton and corn futures.	VECM-MSVR promising for forecasting.	Limited interval research, Chinese focus, model integration, profitability analysis
23	The study analyzes grain futures markets.	Employed Phillips, Shi, Yu Bubble Test.	Assessed price explosiveness factors.	Utilized daily futures prices (2004–2013).	Commodity index traders don't drive bubbles.	Understanding speculative impact, market dynamics.

Prior Art (Patent Search) :

Patent Application No.	Title of Patent	Existing Solutions (Abstract of Patent)
CN103577581B	Agricultural product price trend forecasting method	The technique of agricultural product price trend forecasting described in the invention includes the following steps. Step 1: computer-acquired article relevant to agricultural commodity price and with a forecasting standpoint;Step 2: Duplicate elimination is completed on the collected articles;Extract and save the article's important element in step three.Step 4: The location of the agricultural product-related area mentioned in the article is found;Step 5: Quantify and preserve agricultural products according to the given predictability viewpoint after analyzing expert opinions using text mining technology;Step 6: Utilizing the model established so that agricultural product price is carried out trend prediction, the trend prediction viewpoint delivering time, agricultural product affiliated area, agricultural product sort, and quantization according to article is carried out using microcomputer modeling.
CN105205099B	A kind of agricultural product price analysis method	The steps included in the technique of agricultural product price analysis that the present invention relates to are as follows:Information about the types of agricultural products is gathered using one assembled classifier of pre-trained search engines;The default commodities trading website is crawled to obtain the geographic location information of the supplier for each category of agricultural commodity as well as the pricing data.It is divided based on the area where agricultural goods are cultivated , with each agricultural product kind carrying out agricultural production and obtaining information on the area where agricultural goods are cultivated for each kind.
WO2018232845A1	Smart agriculture management method and system	The present invention relates to a smart agriculture management system and method, the system comprising: monitoring soil nutrient data and moisture data; acquiring historical weather data and predicted weather data; acquiring

		historical price data; comparing the nutrient data with a nutrient content standard value and comparing the moisture data with a moisture content standard value; processing and analyzing the historical price data of the crop and the historical weather data. The system consists of a control center, a soil monitoring module, an information-gathering module for the weather, and an acquisition module for prices.
--	--	---

3. Gaps Identified

1. Integration of ML(Machine Learning) and Time Series: While ML(machine learning) and time series forecasting methods are commonly used separately, there may be a gap in combining these two approaches effectively for crop price forecasting. Your project aims to bridge this gap by integrating both methods and comparing their performance.
2. Lack of Short-term Forecasting Models: Many existing studies focus on long-term crop price forecasting. However, short-term forecasting is essential for farmers and traders to make immediate decisions. Your project addresses this gap by specifically targeting short-term forecasts.
3. Model Comparison: While ARIMA is a broadly used benchmark for time series forecasting, there may be a gap in comparing its performance with machine learning models in the context of crop price forecasting. Your project fills this gap by conducting a comparative analysis.

4. Scope of improvement and limitations

Scope of Improvement :

1. The scope of this project is limited to the forecasting of orange and cotton crop prices in the Indian market using machine learning algorithms.
2. The system will be developed using publicly available data and will not involve any primary data collection.

Limitations :

1. Much of the external factors cannot be accommodated into the model like unexpected climatic conditions and marketing conditions like hoarding.
2. Inferences will be very specific to the kind and specific variety of the crop chosen which means that a generalized model is not possible.

4.0 Proposed Methodologies / Proposed Approach

Block schematic of the procedure adopted to carry out is depicted in 4.1.

The procedure adopted to carry out proposed work with details of the procedure is as follows:

4.1 Data Collection

4.1.1 Source

The dataset is acquired from the Government of India's Open Government Data Platform and Agricultural Market.

4.1.2 Period

We have taken the data for five years to be able to sufficiently generalize our model and a moderate data size as per our computing infrastructure.

4.2 Preliminary overview of data

- The dataset has eight attributes, namely - state, district, market, commodity, variety,
- arrival_date, min_price, max_price. Our target variable is modal_price.
- This dataset had some missing values for min_price and max_price which we had to preprocess before moving on to building a model based on the dataset.

4.3 Exploratory Data Analysis

- Studying the existing Exploratory Data Analysis techniques.
- Applying selected techniques on the collected dataset.
- Based on the performed Exploratory Data Analysis technique generate indicators.
- Try to generalize the data based on patterns observed and verify it using indicators.
- Accordingly make changes in dataset.

4.4 Preprocessing

- Imputing the missing values using various strategies like - mean, median, mode or nearest neighbor.
- Convert all the categorical values into numerical values by using encoding techniques like
- one-hot, label, ordinal, etc.
- Scaling the values to make computation uniform across each iteration and epochs.
- Normalizing the data points and removing outliers based on statistical tests to avoid overfitting and high model complexity.

4.5 Feature Engineering

- After using encodings like one-hot encoding our attribute size grows proportional to the distinct values in the original attribute.
- We use feature selection techniques like Principal Component Analysis to capture the essence of all the attributes and compress the variability in k-features.

4.6 Model Selection

- Study regression based models that can potentially be good in forecasting.
- Select a subset of these models to form candidate algorithm set.
- Identify performance evaluation metrics for these implementations.
- Execute vanilla implementation of each candidate algorithms and select the

most promising subset for the next steps.

4.7 Model Training

- Split the dataset into distinct training and testing set. From the training set keep subset as validation set.
- Utilise different strategies for validation like k-fold validation.
- Cycle between various training validation cycles to tune the parameters to obtain the most 20 optimum set of performance evaluation metrics.

4.8 ARIMA implementation

- Apply the ARIMA model on the refined dataset.
- Check all the parameters and tune the parameters to obtain the most optimum set of performance evaluation metrics.

4.9 Model Evaluation

- The best performing model will be our final model for forecasting. Apply all the trained models on the testing set.
- Select the best model based on all the performance evaluation metrics.

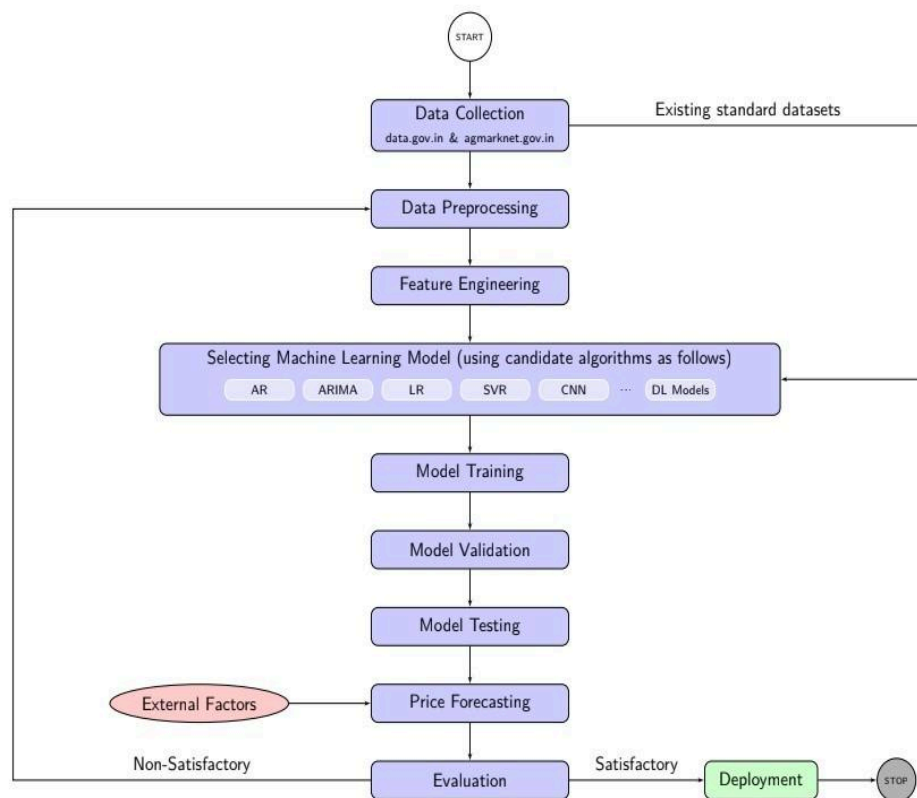


Figure 1: Methodology Block Diagram for Machine learning based short-term forecasting of orange and cotton crop prices in the context of the Indian market

5.0 Results and Discussion

Results

Time series forecasting was conducted in this study to predict the prices of oranges and cotton. The intervention period was from 2018 to 2022. Figure 5.1 shows the time series plot of cotton prices in India, while Figure 5.2 shows the time series plot of orange prices in India.

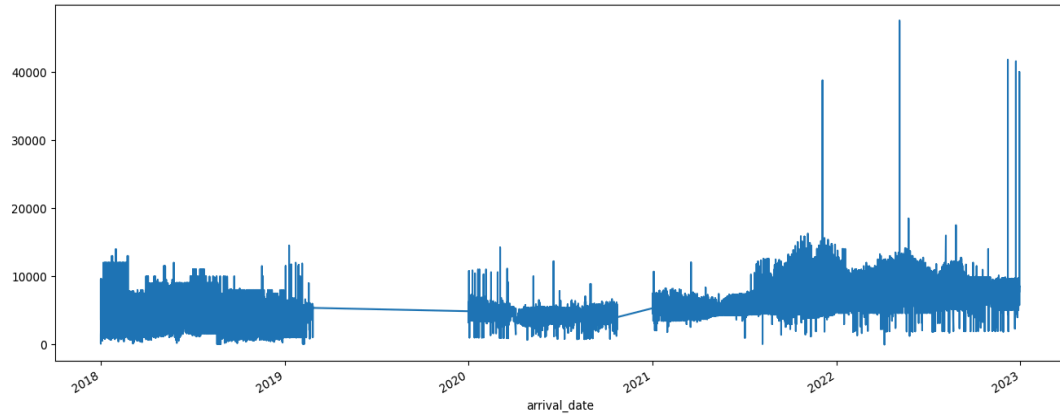


Fig 5.1. Time Series plot of Cotton Prices in India

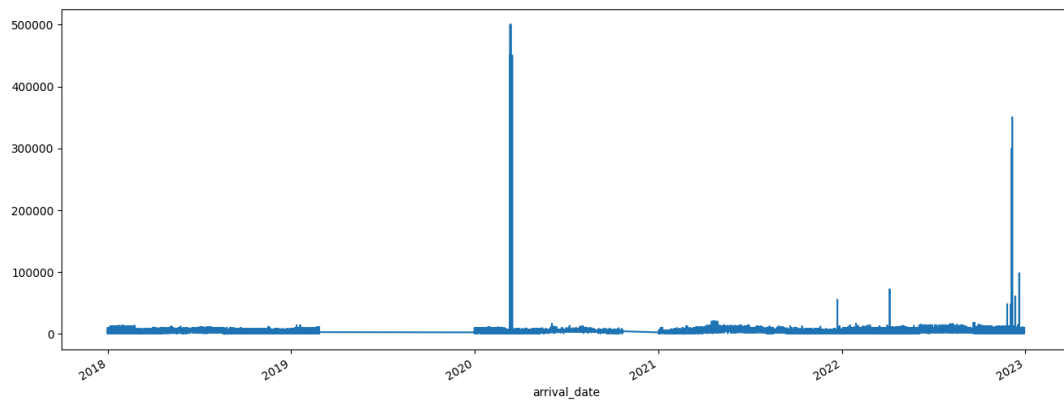


Fig 5.2. Time Series plot of Orange Prices in India

We have imported the Augmented Dicky-Fuller(ADF) test from ‘stattools’ and performed the ADF test on a dataset.

It determines whether the time series is stationary or not.

Table 5.1. Results of ADF Test

ADF	-19.709403365352486
P-Value	0.0
Number Of Lags	59
Number of observations used for ADF regression	58269
Critical Values	1% : -3.430462230999349 5% : -2.861589603951926 10% : -2.56679640251604

If the ADF statistic is more negative than the critical value at a given significance level, then the null hypothesis of non-stationary is rejected. It exhibits trends or patterns that vary overtime. Hence, We can determine that ‘modal_price’ is likely stationary.

Figure 5.3 shows a summary statistics table for the principal components (PC1 to PC10) of the dataset. It includes statistics such as count, mean, standard deviation, minimum, 25th percentile, median (50th percentile), 75th percentile, and maximum values for each principal component. These statistics give insights into the distribution and characteristics of the principal components in the dataset.

	count	mean	std	min	25%	50%	75%	max
PC1	58095.0	0.002111	0.681346	-0.705121	-0.539274	-0.498282	0.768234	1.140785
PC2	58095.0	0.000931	0.418753	-0.770957	-0.275533	-0.097021	0.287222	1.304029
PC3	58095.0	-0.000229	0.399492	-0.741450	-0.301095	-0.061363	0.244472	0.829785
PC4	58095.0	0.001482	0.357057	-0.810879	-0.162991	-0.045996	0.123058	0.874225
PC5	58095.0	-0.001337	0.337317	-0.724099	-0.240971	0.019507	0.223303	0.827871
PC6	58095.0	-0.001017	0.327201	-0.822308	-0.228749	0.009565	0.220706	0.942050
PC7	58095.0	-0.000033	0.302162	-0.687368	-0.155780	-0.021336	0.110331	1.071474
PC8	58095.0	0.001648	0.276641	-0.837648	-0.062432	-0.028280	0.040898	1.265809
PC9	58095.0	0.002757	0.262201	-0.684558	-0.130565	0.051126	0.165191	0.767326
PC10	58095.0	-0.004253	0.238921	-0.924919	-0.021179	0.001328	0.082865	1.055913

Fig 5.3. Summary Statistics Table of Principal Components

5.1. Results Of ARIMA Model

Table 5.2 shows the results of ARIMA implementation. The ARIMA model with order (1,1,1) has the lowest MAE, MSE, and RMSE values among the three configurations, indicating better predictive performance compared to (1,1,2) and (1,2,2).The model with order (1,2,2) has the highest MAE, MSE, and RMSE values, suggesting that it may not be as effective in capturing the price trends accurately.In summary, based on the provided results, the model with order (1,1,1) appears to have relatively better accuracy compared to the other configurations.

Table 5.2. Results of ARIMA Model

Parameters	MAE	MSE	RMSE
order(1,1,1)	1492.94	3223544.63	1795.42
order(1,1,2)	1514.52	3393028.97	1842.01
order(1,2,2)	1900.71	5475048.05	2339.88

MAE : Mean Absolute Error

MSE : Mean Squared Error

RMSE : Root Mean Squared Error

5.2. Results Of Linear Regression

Table 5.3 shows the results of Linear Regression. implementation.The model performs significantly better with the intercept included (fit_intercept = True) compared to without it (fit_intercept = False). This is evident from the much lower MAE, MSE, and RMSE values when the intercept is included.The extremely high

MAE, MSE, and RMSE without the intercept suggest that the model without the intercept is not capturing the underlying relationships in the data effectively. The model with the intercept has relatively low error metrics, indicating better predictive accuracy.

Table 5.3. Results of Linear Regression

Parameters	MAE	MSE	RMSE
fit_intercept = True	189.33	108074.73	328.74
fit_intercept = False	3621.35	1323066.88	3637.39

5.3. Results Of Support Vector Regression

Table 5.4 shows the results of Support Vector Regression. Among the kernels tested, the sigmoid kernel has the lowest MAE and RMSE, indicating relatively better predictive performance compared to the others. The RBF (Radial Basis Function) kernel and Linear kernel show similar MAE and RMSE values, suggesting comparable performance. The polynomial (Poly) kernel has the highest MAE and RMSE, indicating poorer predictive accuracy compared to other kernels. Overall, the choice of kernel significantly affects the performance of the SVR model. The sigmoid kernel seems to be the most suitable for this dataset based on the provided metrics.

Table 5.4. Results of Support Vector Regression

Parameters	MAE	MSE	RMSE
kernel = 'sigmoid'	398.65	565297.84	751.86
kernel = 'rbf'	585.98	1139148.92	1067.30
kernel = 'poly'	948.92	2881658.33	1697.54
kernel = 'linear'	585.98	1139148.92	1067.30

The accuracy of SVR can be considered acceptable based on the MAE, MSE, and RMSE values. Lower values of these metrics indicate better model performance. While ARIMA serves as a benchmark for forecasting, SVR demonstrates superior performance in this context.

Discussion

In this study, we aimed to implement machine learning-based models for short-term forecasting of orange and cotton crop prices in the Indian market. The agricultural sector, being a significant part of India's economy, faces challenges due to the volatile nature of crop prices influenced by factors like weather, supply-demand dynamics, and government policies. Farmers and traders often struggle to make informed decisions regarding the timing of product sales and purchases amidst these price fluctuations. We utilized a dataset obtained from India's Open Government Data Platform and Agricultural Market, covering five years from 2018 to 2022, with 2023 data serving as the training and validation set. The dataset included attributes such as state, district, market, commodity, variety, arrival_date, min_price, max_price, and the target variable, modal_price. Our methodology involved exploratory data analysis (EDA) to handle missing data and identify patterns, followed by splitting the

dataset into training, testing, and validation sets. We employed machine learning algorithms, specifically ARIMA, Logistic Regression (LR), and Support Vector Regression (SVR), for price forecasting. The SVR models were evaluated using different kernels: sigmoid, rbf (Radial Basis Function), poly (Polynomial), and linear. The results showed that SVR with the sigmoid kernel outperformed other models, exhibiting the lowest MAE and RMSE values, indicating better predictive accuracy. In conclusion, our study highlights the effectiveness of machine learning models, particularly SVR with the sigmoid kernel, in forecasting crop prices in the Indian market. The findings suggest the potential for improved decision-making and risk management for farmers and traders in the agricultural sector. Further research could focus on refining model parameters and exploring additional features to enhance predictive capabilities.

6.0 Conclusion and Future Scope

Conclusion

In conclusion, this project successfully establishes a Machine Learning-based short-term forecasting mechanism for orange and cotton crop prices in the Indian market. Leveraging a comprehensive dataset provided by the Indian Government, the study incorporates Exploratory Data Analysis (EDA) to address data challenges and reveal underlying patterns. The model development phase involves the careful selection of regression-based algorithms, including Artificial Neural Networks (ANN), Auto Regressive Integrated Moving Average (ARIMA), Logistic Regression (LR), and Support Vector Regression (SVR). Importantly, the inclusion of an ARIMA model serves as a benchmark for comparative analysis. The geographic segmentation of the dataset into urban and rural tiers, state-level analysis, and city-specific insights enriches the model with regional and seasonal dynamics, providing a nuanced understanding of pricing variations. The model's performance is evaluated at national, state, and city levels, offering a comprehensive assessment. This multifaceted approach not only contributes to accurate short-term price predictions but also furnishes actionable insights for stakeholders in the agricultural supply chain, thereby enhancing the resilience and sustainability of India's agricultural sector.

Future Scope

1. **Integration of Additional Factors:** Future iterations of the model could consider incorporating additional factors such as climate data, market demand, and government policies to enhance predictive accuracy and comprehensiveness.
2. **Dynamic Model Updating:** The development of a dynamic updating mechanism would ensure that the model adapts to changing market conditions, further improving its applicability and relevance over time.
3. **Extension to Other Crops:** The methodology developed in this project can be extended to other key crops, broadening the scope of the model's applicability and providing a more holistic understanding of India's agricultural pricing dynamics.
4. **Incorporation of Advanced ML Techniques:** The integration of advanced machine learning techniques, such as ensemble models or deep learning

architectures, could be explored to push the boundaries of forecasting precision.

5. Stakeholder Engagement: Future research could involve greater stakeholder engagement, ensuring that the developed model aligns with the practical needs and perspectives of those directly involved in the agricultural sector.
6. Real-time Predictions: Building real-time prediction capabilities would be a valuable enhancement, allowing stakeholders to access up-to-the-minute insights for more agile decision-making.

References

- [1] Casper Solheim Bojer. (2022) "Understanding machine learning-based forecasting methods: A decomposition framework and research opportunities". <https://doi.org/10.1016/j.ijforecast.2021.11.003>
- [2] Gianluca Bontempi , Souhaib Ben Taieb, and Yann-A`el Le Borgne (2013) "Machine Learning Strategies for Time Series Forecasting". https://doi.org/10.1007/978-3-642-36318-4_3
- [3] Yitong Li, Kai Wu, Jing Liu (2023) "Self-paced ARIMA for robust time series prediction". <https://doi.org/10.1016/j.knosys.2023.110489>
- [4] Ahmed Tealab, Hesham Hefny, Amr Badr (2017) "Forecasting of nonlinear time series using ANN". <https://doi.org/10.1016/j.fcij.2017.05.001>
- [5] Hanyu Yang a, Xutao Li a, Wenhao Qiang a, Yuhao Zhao a, Wei Zhang b, Chang Tang c (2021) "A network traffic forecasting method based on SA optimized ARIMA–BP neural network". <https://doi.org/10.1016/j.comnet.2021.108102>
- [6] Jennifer L. Castle , Michael P. Clements, David F. Hendry (2014). "Robust approaches to forecasting ". <https://doi.org/10.1016/j.ijforecast.2014.11.002>
- [7] G. Peter Zhang (2002) "Time series forecasting using a hybrid ARIMA and neural network model". [https://doi.org/10.1016/S0925-2312\(01\)00702-0](https://doi.org/10.1016/S0925-2312(01)00702-0)
- [8] Dongqing Zhanga, Guangming Zangb, Jing Lia, Kaiping Maa, Huan Liu. (2018) "Prediction of soybean price in China using QR-RBF neural network model". <https://doi.org/10.1016/j.compag.2018.08.016>
- [9] Di Tella University, Argentina and Buenos Aires University, Argentina(2016)

“Forecasting food prices: The case of corn, soybeans and wheat”.

<https://doi.org/10.1016/j.ijforecast.2016.01.002>

- [10] Kiran M. Sabu, T. K. Manoj Kumar (2020).“Predictive analytics in Agriculture: Forecasting prices of Areca Nuts in Kerala”
<https://doi.org/10.1016/j.procs.2020.04.076>
- [11] Pantazi, X. E., Moshou, D., Alexandridis, T., Whetton, R. L., & Mouazen, A. M. (2016). “Wheat yield prediction using machine learning and advanced sensing techniques”.
<https://doi.org/10.1016/j.compag.2015.11.018>
- [12] Foteini Kyriazi , Dimitrios D. Thomakos , John B. Guerard (2019).“Adaptive learning forecasting, with applications in forecasting agricultural prices”.
<https://doi.org/10.1016/j.ijforecast.2019.03.031>
- [13] Yegnanew A. Shiferaw(2022) “An analysis of East African tea crop prices using the MCMC approach to estimate volatility and forecast the in-sample value-at-risk”.
<https://doi.org/10.1016/j.sciaf.2022.e01442>
- [14] J. Scott Armstrong , Kesten C. Green , Andreas Graefe(2015) “Golden rule of forecasting: Be conservative”.
<http://dx.doi.org/10.1016/j.jbusres.2015.03.031>
- [15] Yuehjen E. Shao , Jun-Ting Dai(2018) “Integrated Feature Selection of ARIMA with Computational Intelligence Approaches for Food Crop Price Prediction”
<https://doi.org/10.1155/2018/1910520>
- [16] Thomas Dimpfla, Robert C. Jungb, Michael Fladc(2017) “Price discovery in agricultural commodity markets in the presence of futures speculation”
<https://doi.org/10.1016/j.jcomm.2017.01.002>
- [17] H.F. Zoua, G.P. Xia, F.T. Yang, H.Y. Wanga(2007) “An investigation and comparison of artificial neural network and time series models for Chinese food grain price forecasting”
<https://doi.org/10.1016/j.neucom.2007.01.009>
- [18] Liege Cheung , Yun Wang , Adela S.M. Lau , Rogers M.C. Chan (2022) “Using a novel clustered 3D-CNN model for improving crop future price prediction.
<https://doi.org/10.1016/j.knosys.2022.110133>
- [19]V. Sneha; V. Bhavana (2023) “Sugarcane Yield and Price Prediction Using Forecasting Models”
<https://doi.org/10.1109/ICECONF57129.2023.10084094>
- [20]B Chaitra,K Meena,(2023)“Forecasting Crop Price using various approaches of Machine Learning”
<https://doi.org/10.1109/ICIET57285.2023.10220616>

- [21]Juliana Ngozi Ndunagu, Eyiyemi.Helen Aderemi, Rasheed Gbenga Jimoh, Joseph Bamidele Awotunde(2022) “Time Series: Predicting Nigerian Food Prices using ARIMA Model and R-Programming”
<https://doi.org/10.1109/ITED56637.2022.10051516>
- [22]Tao Xiong ,Chongguang Li,Yukun Bao , Zhongyi Hu , Lu Zhang(2015) “A combination method for interval forecasting of agricultural commodity futures prices”
<https://doi.org/10.1109/ITED56637.2022.10051516>
- [23]Xiaoli L. Etienne, Scott H. Irwin, Philip Garcia (2014) “Price Explosiveness, Speculation and Grain Futures Prices”
<https://doi.org/10.1093/ajae/aau069>