CROP YIELD ESTIMATION USING ARIMA REGRESSION

S. Renuka Devi
Artificial Intelligence and Data Science
Rajalakshmi Engineering College
Chennai, India
renukadevi.s@rajalakshmi.edu.in

Swetha P
Artificial Intelligence and Data Science
Rajalakshmi Engineering College
Chennai, India
221801054@rajalakshmi.edu.in

Sebrina Catherine Rose.S
Artificial Intelligence and Data Science
Rajalakshmi Engineering College
Chennai, India
221801048@rajalakshmi.edu.in

***Abstract*—This paper discusses the prediction of crop prices using ARIMA (Auto-Regressive Integrated Moving Average) regression, a statistical model well-suited for time series analysis. By examining historical crop price data, ARIMA regression captures trends, seasonal variations, and cyclical patterns, enabling accurate price forecasts. In the context of agriculture, where prices fluctuate due to factors like seasonal harvest cycles, demand changes, and market conditions, ARIMA's ability to model these dynamics is especially valuable. Its integration within modern crop prediction systems enhances decision-making for farmers, agribusinesses, and policymakers, who rely on precise price forecasts to optimize storage, transportation, and sales strategies. Through the accurate prediction of crop prices, ARIMA contributes significantly to addressing the challenges of price volatility and planning, ultimately supporting more resilient and efficient agricultural practices.**

***Keywords*—Crop Price Prediction, Seasonal Patterns, Agricultural Forecasting, ARIMA Regression, Agricultural Productivity.**

INTRODUCTION

The agricultural industry has always struggled with unpredictable market conditions, fluctuating crop prices, and the challenge of meeting the food demands of a growing global population. To address these ongoing challenges, this project leverages advanced data analytics, machine learning, and ARIMA (Auto-Regressive Integrated Moving Average) regression to provide accurate and precise crop yield estimation. This technology enables farmers to make data-driven decisions, optimize their resources, and manage risks associated with climate variability and market volatility. By predicting crop yields and prices more accurately, the system enhances market competitiveness, allowing farmers and agribusinesses to better anticipate supply and demand shifts, plan their harvest and storage strategies, and set competitive prices that reflect real-time market conditions.

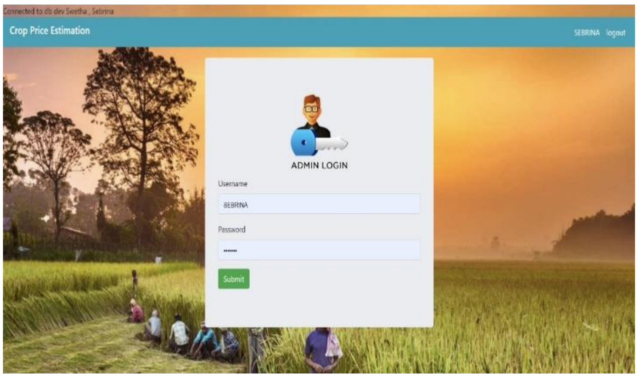
Additionally, these predictive insights empower policymakers to make timely interventions, such as price stabilization and subsidy allocations, that help protect farmers' incomes and stabilize food markets. The integration of such technologies into the agricultural sector not only contributes to economic resilience but also promotes sustainable farming practices by minimizing waste and improving resource allocation. Ultimately, this approach supports food security on a global scale, creating a more efficient, sustainable, and resilient agricultural demand.

MOTIVATION

The motivation is to address the pressing need for innovative technologies, including advanced statistical methods like ARIMA regression, to estimate crop yield in the face of a rapidly growing population and increased demand for food. Such technologies can boost farm output and enhance economic returns for farmers, but also provide them with the economic security essential for long-term food security and environmental health.By leveraging techniques like ARIMA (AutoRegressive Integrated Moving Average) regression, which excels at analyzing time series data, farmers can gain accurate, data-driven insights into crop yield patterns and crop price patterns. Ultimately, the integration of ARIMA regression and other modern methods promotes agricultural sustainability by increasing productivity and resilience, enabling the agricultural sector to meet growing demands without placing additional strain on the environment

DATASET SUMMARY

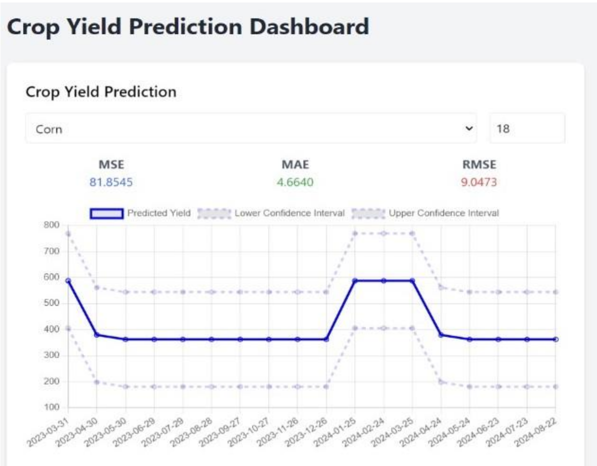
The dataset consists of 500 rows and 7 columns, providing information on various crop yields over time. Key columns include `id`, a unique identifier for each record, and `crop_type`, indicating the type of crop (e.g., Corn, Wheat, Barley). The `yield_amount` column records the yield output for each crop instance. The `date` column specifies the date of each yield record, allowing for time-series analysis essential for ARIMA regression. Additionally, `area` represents the land area (in acres or hectares) used for cultivation, while `rainfall` records the precipitation (in mm) that the area received. The `temperature` column gives the average temperature (in Celsius) at the time of yield measurement. Data types vary, with integers, floats, and strings. There are no missing values in this dataset, indicating complete data for all records. A sample entry includes crop types like Barley and Coffee, each with unique values for area, rainfall, and temperature.



ADMIN LOGIN PAGE

The admin login page for your crop price prediction project has a clean, functional layout. The header shows "Crop Price Estimation" and session details like database, developer, and user. A farming background adds agricultural context. The central login panel has an "Admin Login" label, fields for username and password, and a "Submit" button.

CROP YIELD PREDICTION



The image shows a "Crop Yield Prediction Dashboard" that provides a forecast for crop yield, specifically for "Corn," using ARIMA regression. Here is an overview of the key components and metrics:

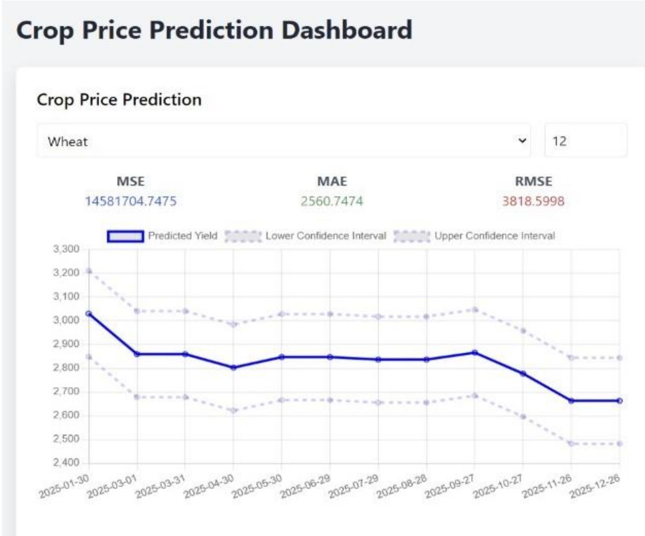
Model Performance Metrics:

MSE (Mean Squared Error): 81.8545 — This is a measure of the average squared difference between predicted and actual values. Lower values indicate better accuracy.

MAE (Mean Absolute Error): 4.6640 — This represents the average absolute error of predictions, showing the average magnitude of errors.

RMSE (Root Mean Squared Error): 9.0473 — This is the square root of MSE and provides an overall measure of prediction accuracy

CROP PRICE PREDICTION



The Crop Price Prediction Dashboard provides a forecast for wheat prices using an ARIMA regression model. Key performance metrics are displayed, including a high Mean Squared Error (MSE) of 14,581,704.75, Mean Absolute Error (MAE) of 2,560.75, and Root Mean Squared Error (RMSE) of 3,818.60, indicating moderate prediction accuracy. The line chart shows a gradual decline in predicted wheat prices over time, with minor fluctuations. Dashed lines represent the upper and lower confidence intervals, highlighting the range of uncertainty in predictions. This suggests potential price decreases and market variability. The dashboard helps users visualize trends and plan ahead based on forecasted price changes.

CROP DISTRIBUTIONS DETAILS

Crop Distribution Details

Crop Type	Record Count	Avg Yield	Min Yield	Max Price Prediction
Banana	28	7333.75	2232	11873
Barley	29	7319.9	2327	11784
Carrot	30	6562.63	1932	11489
Coffee	28	6682.57	1558	11994
corn	19	159.03	65.5	1800
Cotton	22	7112.05	2008	11769
Garlic	25	6572.08	1509	11888
Grapes	21	6372.48	2112	11824
Maize	24	6386.46	1506	11923
Mango	22	6384.73	1800	11773
Millet	30	6456.67	1513	11903

The crop distribution table summarizes key data for various crops, including record count, average yield, minimum yield, and maximum price prediction. Each crop type, like Banana, Barley, and Carrot, has specific entries, showing the number of data records and yield statistics. The average yield and minimum yield give insights into production, while the maximum price prediction offers an upper estimate of market prices. This information helps analyze yield trends and potential pricing for each crop.

Mathematical Representation of ARIMA Regression

Autoregressive (AR): This means the current value depends on past values of itself.

Integrated (I): If the data isn't stable (like a stock price that's constantly increasing), ARIMA can make it stable by taking differences between values.

Moving Average (MA): This means the current value depends on past errors in the forecast.

$$Yt = \alpha + \beta_1 Y_{1}\text{-}_{1} + \beta_2 Y_{1}\text{-}_{2} + ... + \beta p Y_{1}\text{-}p + \varepsilon_1$$

Yt: The current value of the dependent variable.

$\beta_1, \beta_2, ..., \beta p$: Autoregressive coefficients.

$Y_1\text{-}_1, Y_1\text{-}_2, ..., Y_1\text{-}p$: Past values of the dependent variable (lags)

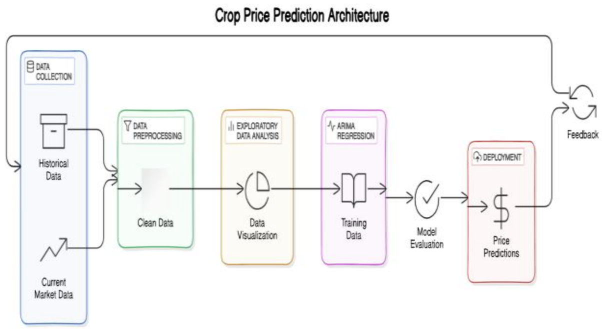
Application of ARIMA Regression in Crop pirce prediction

In this crop prediction system, prediction of crop prices forms an important part of the system and ARIMA regression forms a useful tool. Through analysis of historical prices, ARIMA is able to model trends, cycles and characteristics that affect crop prices to generate important information for all the players involved. For example, in the area of monthly or yearly price forecasting, it is boosted by its capability to capture seasonal fluctuations due to regular crop production and changing weather conditions, cyclical demand patterns.The advantages of ARIMA Regression:

Trend and Seasonality: ARIMA can model both the long-term trend (increasing or decreasing prices over time) and seasonal fluctuations (cyclical patterns that repeat at fixed intervals, such as yearly harvest seasons).

Autocorrelation: It accounts for the autocorrelation in the data—i.e., how current values are correlated with past values, which is common in financial and commodity markets like crop prices.

Flexibility: Can incorporate external factors using **ARIMAX** when relevant variables (e.g., weather, government policies) impact crop prices.



MODEL BUILD CRITERION

The system for predicting crop prices using ARIMA regression is structured around several key layers. First, data is collected from various sources, including historical crop prices, external factors like weather data and market conditions, through APIs or data scraping. This data is then preprocessed to handle missing values, normalize scales, and align time-series data. Feature engineering involves decomposing price data into seasonal, trend, and residual components, while also incorporating external factors as exogenous variables where necessary. The ARIMA model is built using these prepared datasets, with the optimal parameters selected through techniques like AIC and grid search. The model is then evaluated and fine-tuned, followed by using it to forecast future crop prices, with prediction intervals to quantify uncertainty. Once deployed, the model continuously generates updated predictions, monitored for accuracy and drift over time, and retrained as needed with new data. A visualization dashboard displays insights, trends, and forecasts, making the results accessible to stakeholders such as farmers or market analysts. Additionally, an API layer allows integration with other systems, ensuring the predictions are available for use in various applications. This architecture enables an end-to-end solution for accurate, actionable crop price forecasting using ARIMA regression, authenticates the files and orders them to store them in the Data Storage component for proper storage.

CONCLUSION

In conclusion, the crop prediction using ARIMA regression that has been discussed in the paper provide a reliable and effective system for crop price prediction based on the past data collected. It helps the various players within the agriculture value chain to make wiser and better decisions by accurately depicting past, current and expected patterns of production, consumption and demand. From farmers and raw material suppliers to consumers and policy makers, insurers and investors' end, this ability strengthens the capacity to predict market shifts, design and develop optimal supply chains, control and mitigate risks..

In conclusion crop price forecasting made possible by the ARIMA model establishment provides for a significantly enhanced agricultural business model that is capable of coping with price changes and therefore, planning for stable future growth.

REFERENCES

1. Box, G. E. P., & Jenkins, G. M. (1976). Time Series Analysis: Forecasting and Control. San Francisco, CA: Holden-Day.

- This foundational book introduced the ARIMA model and provided a comprehensive guide to time series analysis and forecasting techniques.

2. Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and Practice. OTexts. Retrieved from https://otexts.com/fpp3/

- This online book is an accessible resource for time series forecasting methods, including ARIMA, with practical applications and examples.

3. Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (1998). Forecasting: Methods and Applications. John Wiley & Sons.

- This comprehensive reference covers a wide range of forecasting techniques, including ARIMA, and discusses practical applications and challenges in forecasting.

4. Chatfield, C. (2004). The Analysis of Time Series: An Introduction (6th ed.). CRC Press.

- Chatfield provides a detailed introduction to time series analysis, covering topics such as stationarity, trend, seasonality, and forecasting models, including ARIMA.

5. Agricultural Marketing Information System (AMIS). (2024). Crop Price Data API Documentation. Retrieved from https://www.amis.org

- AMIS provides crop price data through its API, which can be used to access historical and real-time market prices for various crops worldwide.

6. Adhikari, R., & Agrawal, R. K. (2013). "An Introductive Guide to Time-Series Modeling and Forecasting." arXiv preprint arXiv:1302.6613.

- This paper provides an introduction to time series modeling, including ARIMA, and discusses practical considerations and common challenges in forecasting.

7. McKinney, W. (2010). Data Structures for Statistical Computing in Python. Proceedings of the 9th Python in Science Conference.

- This work discusses the use of Python’s pandas library for data manipulation, which is instrumental in the preprocessing phase of time series forecasting projects.

8. Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time Series Analysis: Forecasting and Control (5th ed.). John Wiley & Sons.

- An updated edition of the original ARIMA model book, including new methods for model building and a focus on forecasting applications.

9. Brockwell, P. J., & Davis, R. A. (2016). Introduction to Time Series and Forecasting (3rd ed.). Springer.

- This text offers an introduction to time series analysis with a focus on forecasting models, including ARIMA and seasonal variations.

10. Tang, B., & Li, D. (2020). "Time Series Forecasting with Advanced Machine Learning Models." Journal of Computational and Applied Mathematics, 376, 112857.

- This paper explores the application of advanced machine learning models for time series forecasting, providing insights into potential alternatives to traditional ARIMA models.

11. Sarker, I. H. (2021). "Machine Learning: Algorithms, Real-World Applications and Research Directions." SN Computer Science, 2(3), 160.

- This article discusses the application of machine learning techniques in various fields, including time series forecasting, and provides guidance on selecting appropriate algorithms.