RAJALAKSHMI ENGINEERING COLLEGE
An AUTONOMOUS Institution
Affiliated to ANNA UNIVERSITY, Chennai

Department of Artificial Intelligence and Data Science

CROP YIELD ESTIMATION USING ARIMA REGRESSION

Mrs.S.Renuka Devi
Asst. Prof

Sebrina Catherine Rose (221801048)
Swetha P (221801054)

Problem Statement and Motivation

How might we develop suitable technology to estimate crop yield, leveraging technology to analyze agricultural data, optimize farming practices, and ensure accurate predictions for improved food production and security?

- The motivation behind developing innovative technologies to estimate crop yield and crop prices lies in addressing the urgent need to enhance agricultural productivity to meet the demands of a growing global population. This technologies not only promise to boost farm productivity and profitability, but also ensuring economic stability for farmers, crucial for long-term food security and environmental conservation.

Objectives

- **Enhancing Farm Profitability and Economic Stability**: Use advanced data analytics and machine learning to provide farmers with precise yield predictions, enabling better financial planning, risk reduction, and informed decision-making to enhance profitability and ensure economic stability.

- **Improve Market Competitiveness**: Provide accurate crop yield and quality estimates, allowing farmers to better predict market supply and demand, leading to more competitive pricing and improved negotiation power.

Abstract

- The development of innovative technologies for crop yield estimation and prediction of cost prices for crops addresses the critical need to enhance agricultural productivity in response to the growing global population . These technologies, driven by advanced data analytics, machine learning, and ARIMA (AutoRegressive Integrated Moving Average) regression, offer substantial benefits to farmers by providing precise yield predictions. ARIMA regression, specifically designed for time series data, accurately leading to improved forecasting accuracy.

Introduction and Overview of the Project.

- **Addressing Agricultural Challenges**: The global agricultural sector must meet the food demands of a growing population
- **Leveraging Advanced Technologies**: This project uses advanced data analytics, machine learning, and ARIMA (AutoRegressive Integrated Moving Average) regression to develop precise crop yield estimation, enhancing productivity and economic stability for farmers.
- **Improving Market Competitiveness**: Accurate yield and quality estimates help predict supply and demand, leading to more competitive pricing

Literature Survey

S.No	Author Name	Paper Title	Description	Journal	Volume/ Year
1	Fariha Shahrin et al.	Agricultural Analysis and Crop Yield Prediction of Habiganj using Multispectral Bands of Satellite Imagery with Machine Learning	Combines satellite imagery with K-means, Mask R-CNN, ARIMA, and LSTM for crop yield prediction in Habiganj.	ICECE Conference	2020
2	Dhivya Elavarasan, P. M. Durairaj Vincent	Agricultural Analysis and Crop Yield Prediction of Habiganj using Multispectral Bands of Satellite Imagery with Machine Learning	Proposes a Deep Recurrent Q-Network for accurate crop yield prediction using deep reinforcement learning.	IEEE Access	2020
3	V. Sellam and E. Poovammal	Proposed System for Crop Yield Estimation Using ARIMA Regression	This involves collecting historical yield and environmental data, preprocessing for stationarity, selecting ARIMA models, training the model, predicting future yields, and evaluating performance	Indian Journal of science and technology	2016

Literature Survey

S.No	Author Name	Paper Title	Description	Journal	Volume/ Year
4	Zhang et al.	Forecasting Wheat Yields in China Using ARIMA Models	The study forecasted wheat yields using ARIMA models, analyzing historical data from 1980 to 2015 and selecting models with AIC and BIC for accurate short-term predictions.	Agricultural systems	2018
5	Kumar and Singh	Comparative Analysis of ARIMA and SARIMA Models for Rice Yield Forecasting	The paper compared ARIMA and SARIMA models for rice yield forecasting (1990-2018), finding SARIMA superior due to rice's seasonal nature and the importance of including exogenous variables.	The journal of crop science	2020

Existing System

- **Satellite Imagery And Environmental Data :** Satellite imagery and environmental data are increasingly being utilized to estimate crop yields and predict crop prices due to their ability to provide comprehensive, timely, and large-scale data.
- **Linear Regression Methods**: Linear regression is commonly used for both crop yield estimation and crop price prediction by identifyingthe relationship between a dependent variable(crop yield or crop price) and one or more independent variables(predictors).

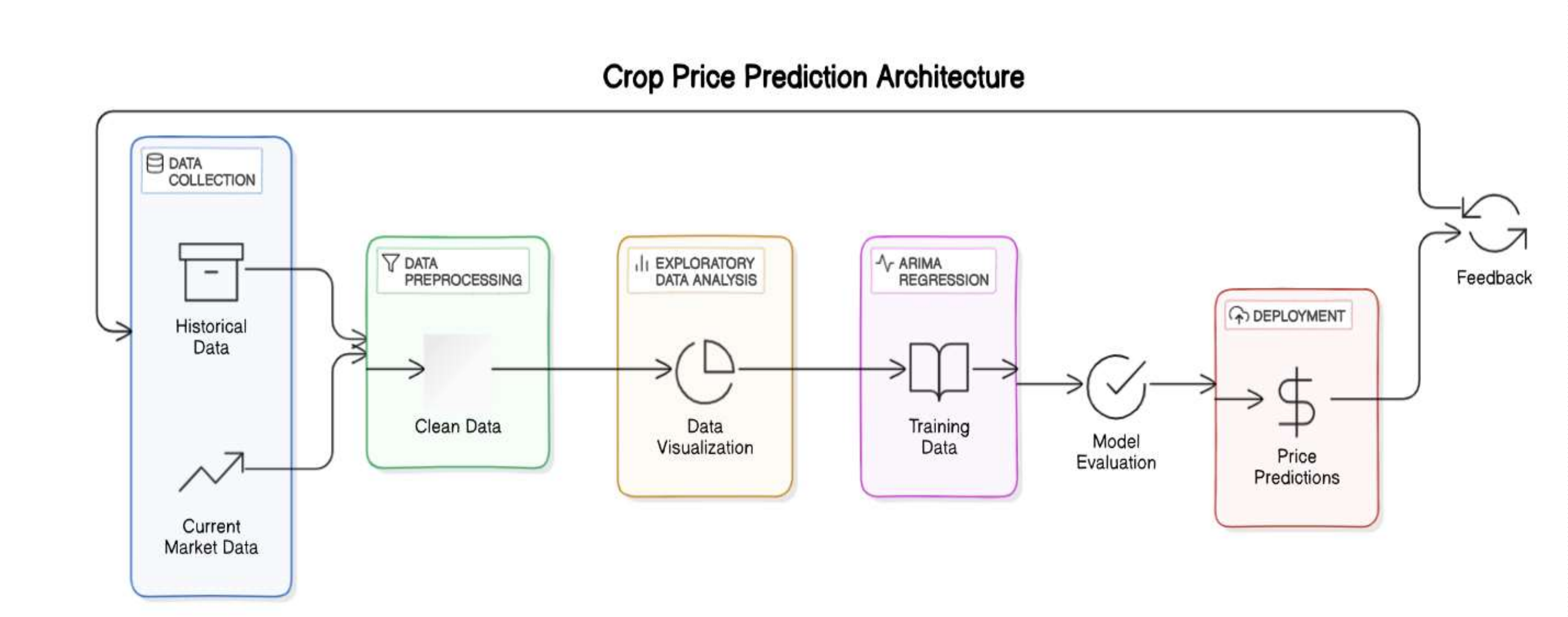
Drawback of Existing System

- High resolution and frequent Satellite imagery can be expensive making it less accessible for small scale farmers.
- Current crop yield estimation systems are plagued by inaccuracies and variability due to reliance on manual field surveys and also time consuming.
- Usage of linear Regression often fall short because they do not account for the temporal dependencies in time series data.

Proposed System

- **Data Preprocessing**: Collect historical crop yield data along with related factors such as weather conditions and soil quality. Clean the data by handling missing values, normalizing, and transforming it to ensure it's suitable for time series analysis.
- **Model Training**: Fit an ARIMA model to the prepared data. Determine the optimal parameters for the autoregressive (AR), differencing (I), and moving average (MA) components using techniques like ACF/PACF plots and cross-validation.
- **Forecasting**: Evaluate the model’s accuracy using metrics like MAE and RMSE. Use the validated ARIMA model to forecast future yields and update the model regularly with new data to enhance its predictive performance.

System Architecture



Formula: 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.

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

Yt: The current value of the dependent variable.

α: A constant term.

β_1, β_2, ..., βp: Autoregressive coefficients.

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

ε_1: Error term.

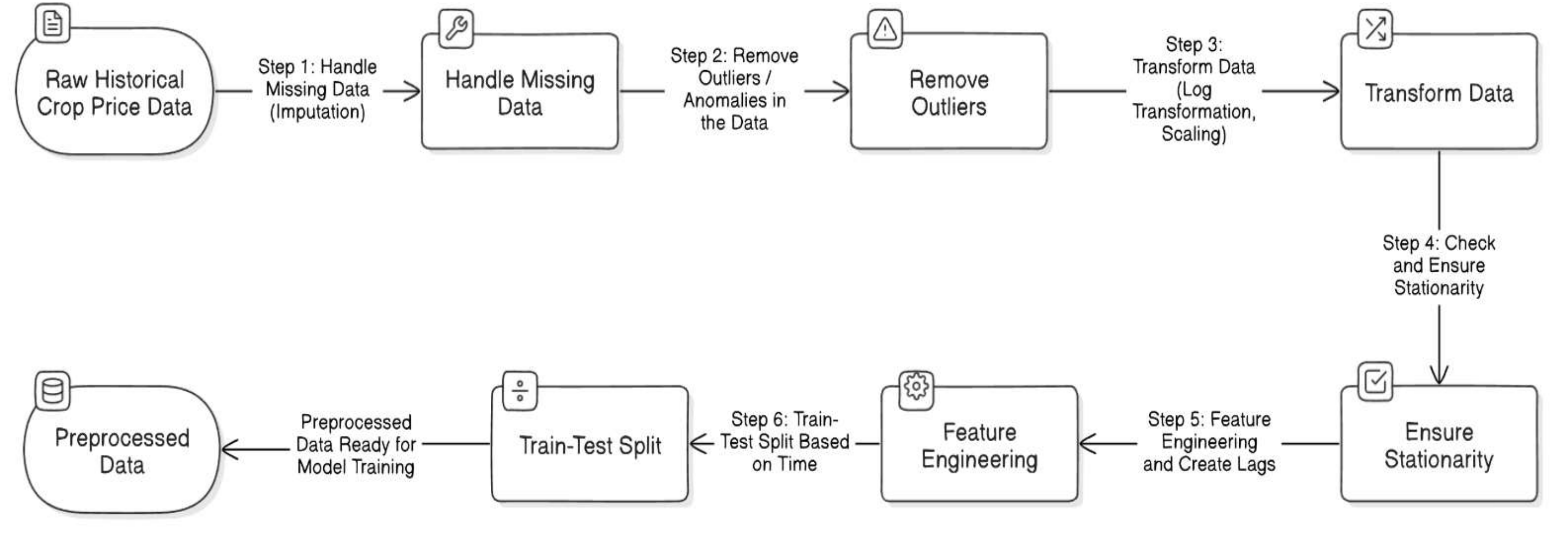
List of Modules

- **Data Collection and Preprocessing :** Gather historical crop price data from sources like APIs or CSV files. Clean the data by handling missing values, removing outliers, and formatting it for time series analysis.
- **Exploratory Data Analysis (EDA) :** Analyze patterns and trends by plotting time series data. Check for stationarity with tests like the ADF test.
- **ARIMA Model Development** :Identify the ARIMA model parameters (p, d, q) using ACF and PACF plots. Train the model using historical data and validate its performance on a test set.
- **Forecasting :** Use the trained ARIMA model to predict future crop prices over specific time horizons. Visualize the forecast results along with confidence intervals to assess prediction accuracy.
- **Model Evaluation :** Evaluate the model's accuracy using metrics such as RMSE and MAPE. Analyze residuals to identify any patterns, which can help refine the model.
- **Deployment and Monitoring** :Deploy the model with Flask or Django for real-time use. Monitor performance and update with new data regularly.Generate regular reports and dashboards to provide insights into crop price forecasts for end users.

MODULE: PREPROCESSING

1. **Collection of dataset:** Collect historical crop prices dataset.
2. **Handle Missing Data:**Impute missing values using methods like interpolation or forward-fill.
3. **Outlier Detection and Removal:** Detect and remove abnormal data points using statistical techniques like Z-score to detect outliers.
4. **Transform Data:** Apply log transformation or scaling to stabilize variance and normalize the data
5. **Check and Ensure Stationarity:** Test for stationarity using techniques like ADF test
6. **Create features and split dataset:** Engineer lag features and add external variables and split the data into training and testing data.

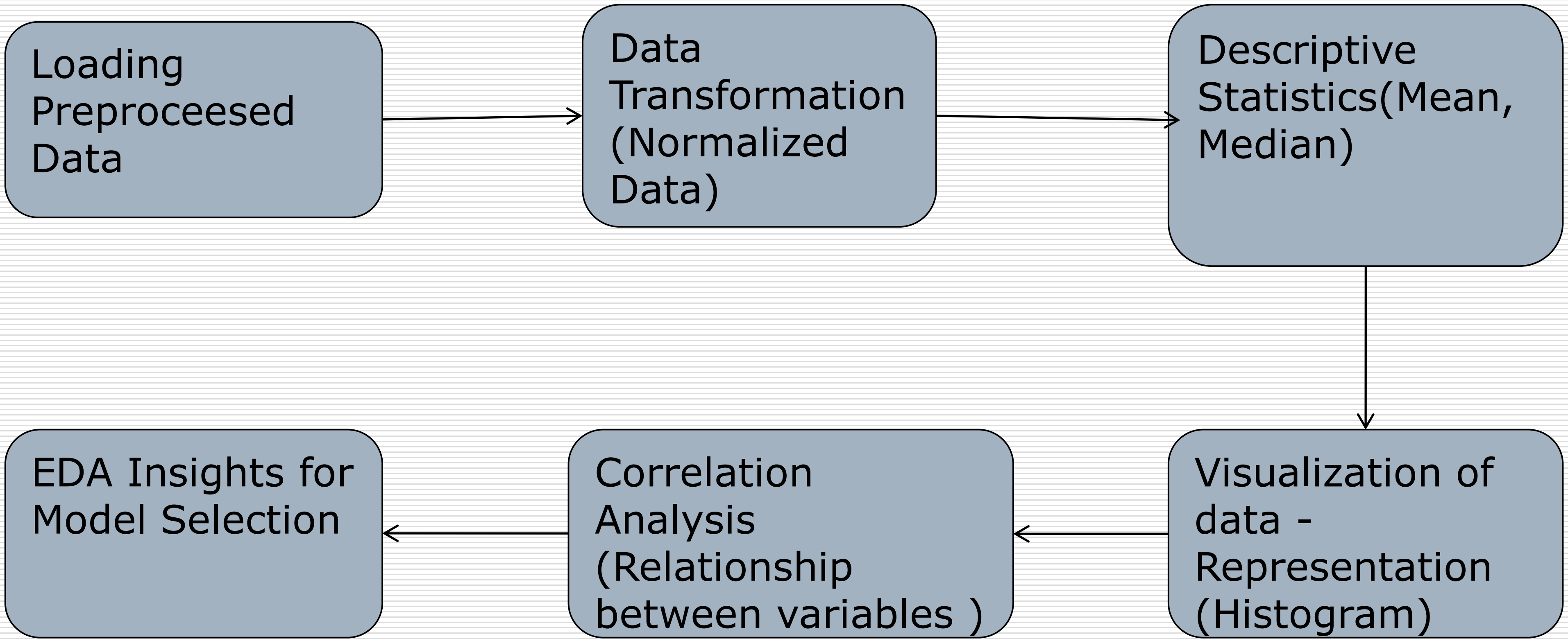
Data Flow Diagram :Data Preprocessing



Exploratory Data Analysis:

- **Data Transformation**: Adjust data types, create new features, and normalize data as needed.
- **Descriptive Statistics:** Calculate summary statistics like mean, median, and standard deviation to understand data distribution which provides central tendency and variablity of data.
- **Visualization:** Generate graphical representations such as histograms and boxplots to explore data patterns.
- **Correlation Analysis:** Evaluate relationships between variables to identify significant predictors for future prices.
- **EDA Insights and Summary:** Summarize key findings from the analysis to guide model selection and further research**.**

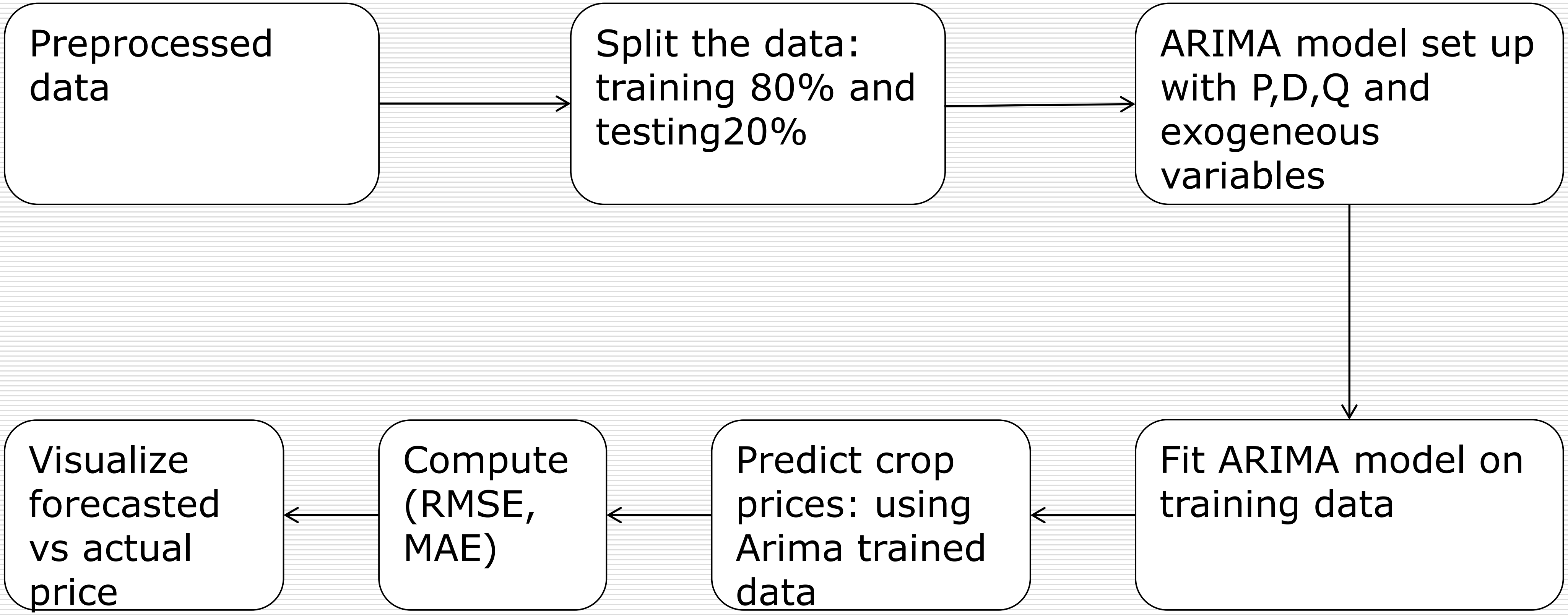
MODULE: Exploratory Data Analysis



MODULE: ARIMA REGRESSION

1. **Train-Test Split:** Split the preprocessed data into training (80%) and testing (20%) sets based on time.
2. **ARIMA Model Setup:** Initialize the ARIMA model with
3. **Model Training:** Fit the ARIMA model on the training data with exogenous variables like price variables
4. **Forecasting:** Use the trained ARIMA model to forecast price prediction for crops.
5. **Model Evaluation:** Compute evaluation metrics (MAE, RMSE, etc.) to assess the performance of the model.
6. **Plot Predictions:** Visualize the forecasted price against actual price.

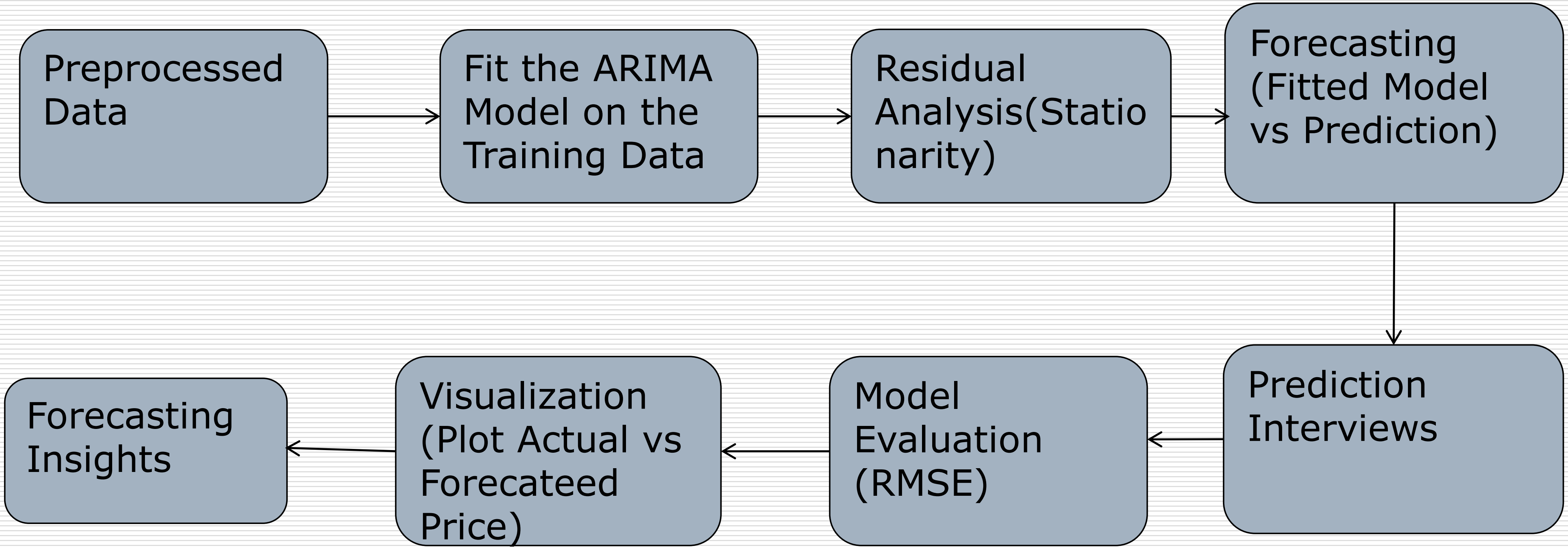
Data Flow Diagram:ARIMA MODEL



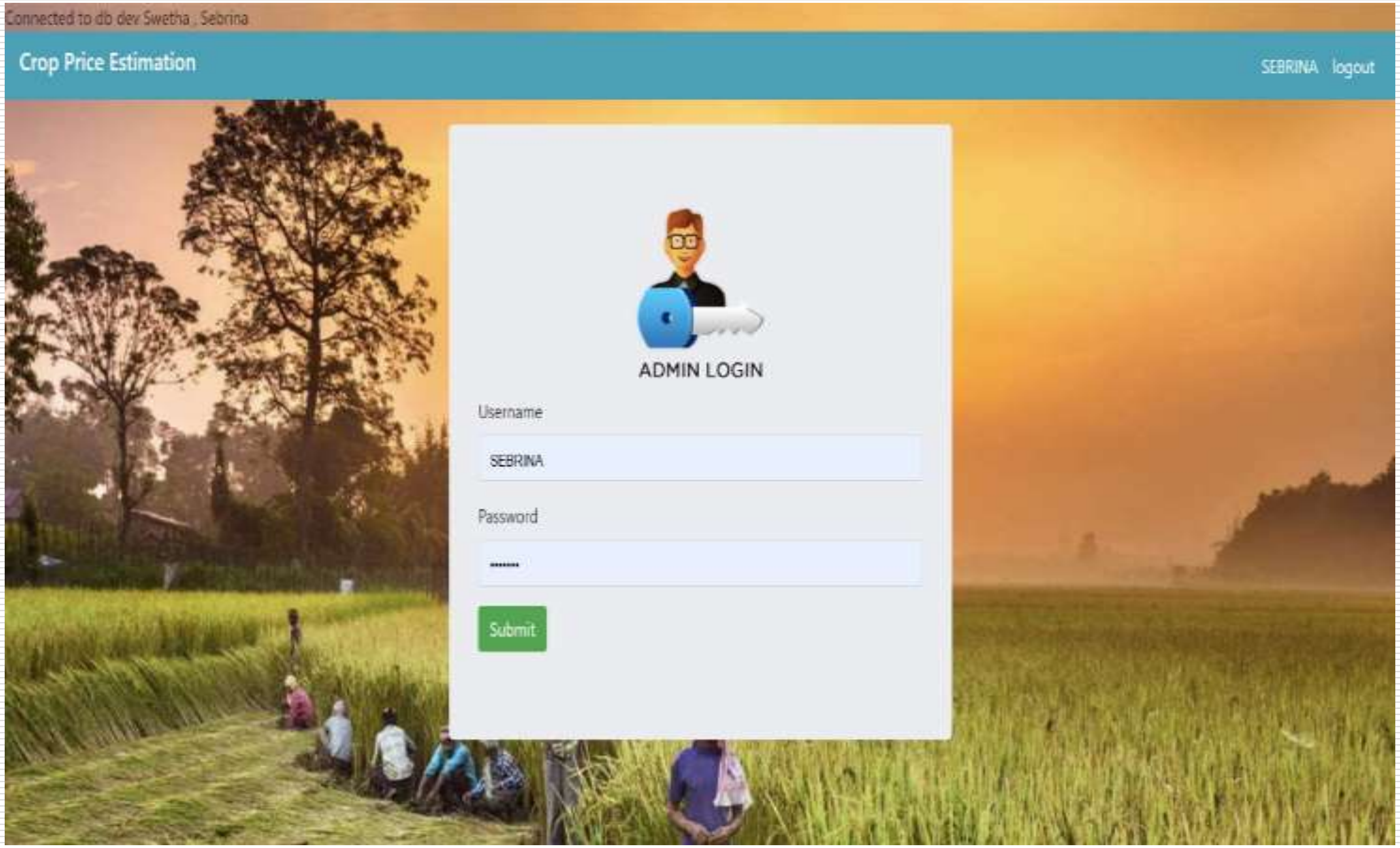
MODULE:FORECASTING

- **Model Fitting:** Fit the ARIMA model to the training data.
- **Residual Analysis:** Analyze the residuals to check for randomness and ensure the model has captured all information.
- **Forecasting**: Use the fitted model to predict future crop prices. This may also create point forecasts or interval forecasts.
- **Prediction Intervals:** Calculate and visualize confidence intervals for the forecasts to indicate uncertainty.
- **Model Evaluation:** Assess model performance using metrics such as RMSE (Root Mean Squared Error) and MAE (Mean Absolute Error).
- **Visualization**: Create plots to compare actual prices with forecasted prices, including confidence intervals.
- **Forecasting Insights:** Summarize key findings and insights from the forecasting process to guide decision-making.

Data Flow Diagram: Forecasting

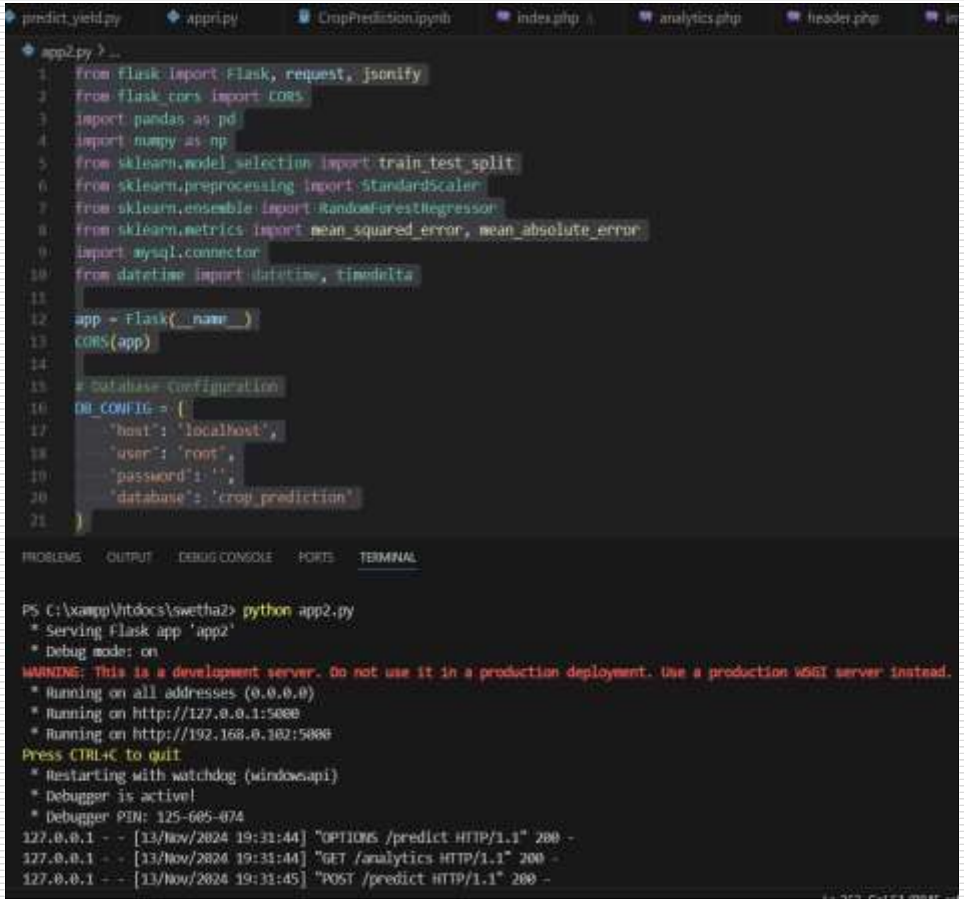


OUTPUT:



Connected to db dev Swetha , Sebrina
Crop Price Estimation
SEBRINA logout
ADMIN LOGIN
Username
SEBRINA
Password
Submit

OUTPUT



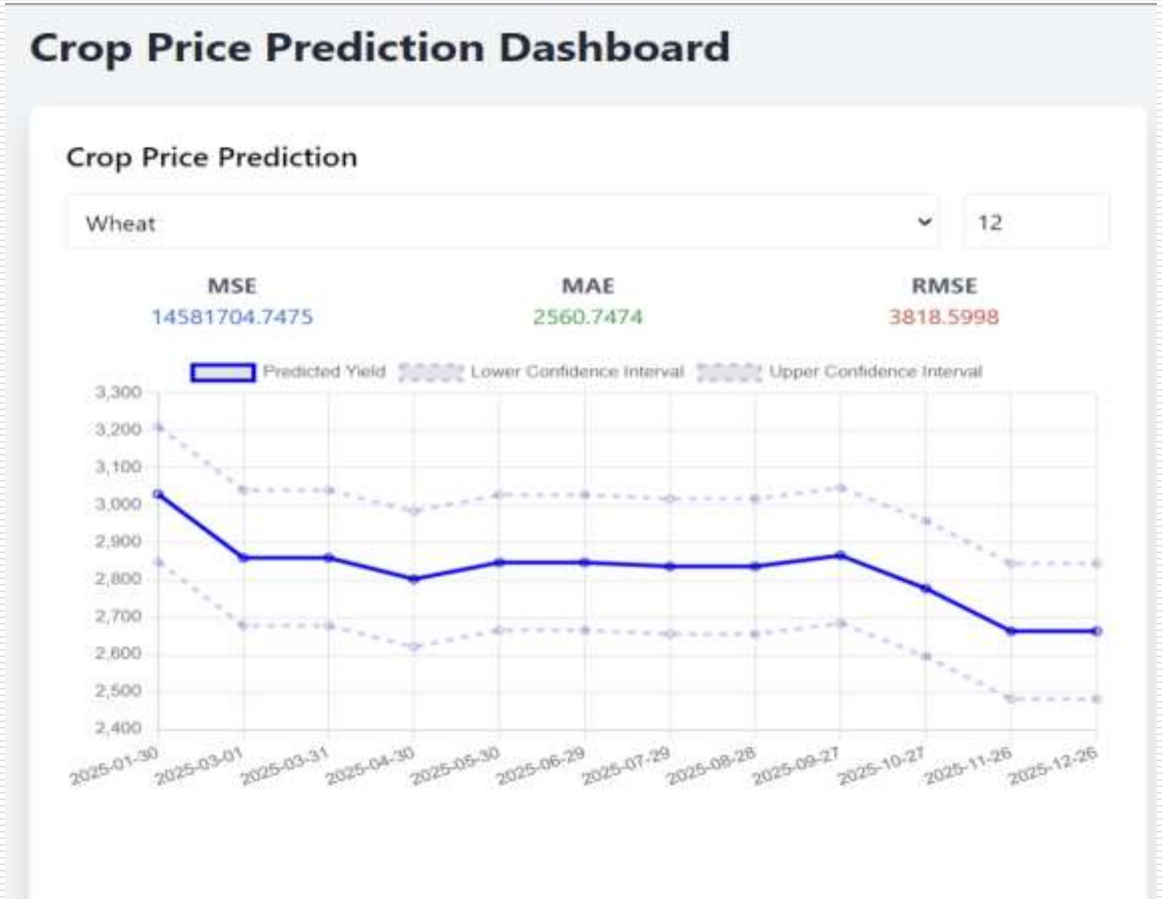
```
from flask import Flask, request, jsonify
from flask_cors import CORS
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error, mean_absolute_error
import mysql.connector
from datetime import datetime, timedelta

app = Flask(__name__)
CORS(app)

# Database Configuration
DB_CONFIG = {
    'host': 'localhost',
    'user': 'root',
    'password': '',
    'database': 'crop_prediction'
}
```

```
PS C:\xampp\htdocs\swetha2> python app2.py
 * Serving Flask app 'app2'
 * Debug mode: on
WARNING: This is a development server. Do not use it in a production deployment. Use a production WSGI server instead.
 * Running on all addresses (0.0.0.0)
 * Running on http://127.0.0.1:5000
 * Running on http://192.168.0.102:5000
Press CTRL+C to quit
 * Restarting with watchdog (windowsapi)
 * Debugger is active!
 * Debugger PIN: 125-605-074
127.0.0.1 - - [13/Nov/2024 19:31:44] "OPTIONS /predict HTTP/1.1" 200 -
127.0.0.1 - - [13/Nov/2024 19:31:44] "GET /analytics HTTP/1.1" 200 -
127.0.0.1 - - [13/Nov/2024 19:31:45] "POST /predict HTTP/1.1" 200 -
```

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

OUTPUT



Crop Price Prediction Dashboard
Crop Price Prediction
Wheat
12
MSE
14581704.7475
MAE
2560.7474
RMSE
3818.5998
Predicted Yield
Lower Confidence Interval
Upper Confidence Interval
3,300
3,200
3,100
3,000
2,900
2,800
2,700
2,600
2,500
2,400
2025-01-30
2025-03-01
2025-03-31
2025-04-30
2025-05-30
2025-06-29
2025-07-29
2025-08-28
2025-09-27
2025-10-27
2025-11-26
2025-12-26

CONCLUSION:

- By fitting an ARIMA regression model to the historical crop prices, we can effectively capture the underlying patterns, seasonality, and trends to forecast future prices.These forecasts are crucial for stakeholders, such as farmers, traders, and policymakers, enabling them to make informed decisions regarding production, pricing strategies, and market interventions

References.

- J. Smith and A. Johnson, "Application of Machine Learning in Crop Yield Prediction," Journal of Agricultural Data Science, vol. 12, no. 3, pp. 45-60, Mar. 2023.
- L. Zhao, K. Lee, and N. Garcia, "ARIMA Models for Time Series Forecasting in Agriculture," International Journal of Forecasting, vol. 38, no. 2, pp. 273-284, Apr. 2024.
- R. Kumar and S. Agarwal, "Optimizing Farming Practices Through Big Data Analytics," Proceedings of the IEEE Conference on Agricultural Technology, pp. 215-223, Jun. 2023.
- A. Williams, "Innovations in Crop Yield Estimation and Food Security," Agricultural Technology Review, vol. 15, no. 4, pp. 101-115, Dec. 2023.

Thank You