NAME- SHRADDHA INGLE
ROLL NO – 17

CROP PRICE PREDICTION

SYNOPSIS

Introduction

Crop prices fluctuate due to multiple factors such as weather conditions, production levels, and economic trends. Predicting these prices can help farmers, policymakers, and traders make informed decisions. This project builds a machine learning model to forecast crop prices based on past trends, seasonal variations, and environmental factors.

Feature Engineering & Preprocessing

Convert categorical variables (crop type, season) into numerical form using One-Hot Encoding.

Standardize numerical features using StandardScaler.

Use sine/cosine transformation for the year feature to capture time trends.

Methodologies Used

- **Data Collection**: Gathering data from various sources, including government reports, weather forecasts, and historical market prices, is crucial for building a reliable prediction model.

- **Machine Learning Techniques**: Algorithms such as Decision Trees, Support Vector Machines (SVM), and Long Short-Term Memory (LSTM) networks are commonly employed to analyze data and predict prices.

- **Feature Extraction**: Identifying relevant features that influence crop prices, such as weather conditions, soil quality, and historical price trends, is essential for improving prediction accuracy.

NAME- SHRADDHA INGLE
ROLL NO – 17

Methodology

1. Data Collection & Generation

Since real-world agricultural price data is not always readily available, we generate synthetic crop price data for this project.

The dataset includes three major crops: Rice, Wheat, and Corn, with historical data from 2020 to 2024, covering four seasons (Winter, Spring, Summer, Fall).

Key factors affecting price include:

Rainfall (in mm)

Temperature (in °C)

Production output (in thousand tons)

Seasonal variations

Yearly trends

2. Exploratory Data Analysis (EDA)

EDA helps us understand trends, correlations, and patterns in the dataset. We perform:

Statistical summaries (mean, median, standard deviation).

Missing value analysis (though synthetic data has none).

Visualizations:

Distribution of crop prices using histograms.

Box plots to compare prices by crop.

Trend analysis of prices over years.

Correlation heatmap to analyze relationships between variables.

3. Feature Engineering & Data Preprocessing

To prepare the data for machine learning models, we apply the following techniques:

Feature Engineering

Encoding categorical features (Crop type, Season) using One-Hot Encoding.

Handling time-based features using Sine & Cosine transformations to capture yearly cyclic patterns.

NAME- SHRADDHA INGLE
ROLL NO – 17

Data Preprocessing

Scaling numerical features (Rainfall, Temperature, Production) using StandardScaler.

Splitting the dataset into training (80%) and testing (20%) sets to evaluate model performance

4. Model Building & Evaluation

We experiment with multiple machine learning models:

4.1 Linear Regression

A simple baseline model.

Assumes a linear relationship between crop prices and influencing factors.

Performance evaluated using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 Score.

4.2 Random Forest Regressor

A more advanced model that handles non-linear relationships and feature interactions.

Ensemble learning technique that combines multiple decision trees.

4.3 Hyperparameter Tuning

Used GridSearchCV to optimize n_estimators, max_depth, min_samples_split for the best performance.

Final Tuned Random Forest model provides the most accurate prediction

Future Price Prediction (2025 & Beyond)

Once the best model is selected, we use it to predict crop prices for 2025 based on estimated:

Rainfall (110 mm)

Temperature (25°C)

Production levels (100k tons)

We also provide a custom prediction function, where users can input their own values for forecasting.

NAME- SHRADDHA INGLE
ROLL NO – 17

Results & Comparison

The Tuned Random Forest model performed the best, achieving the highest R^2 score and lowest error rates.

Feature importance analysis showed that temperature, production, and rainfall had the most significant impact on price