Project Title	**Predicting Crop Production Based on Agricultural Data**
Skills take away From This Project	**Data Cleaning and Preprocessing, Exploratory Data Analysis (EDA), Data Visualization ,SQL, Streamlit, Machine Learning(Regression)**
Domain	**Agriculture**

Problem Statement:

Agriculture is a key contributor to the economy, and accurately predicting crop production is essential for improving planning and decision-making. This project aims to develop a regression model that forecasts crop production (in tons) based on agricultural factors such as area harvested (in hectares), yield (in kg/ha), and the year, for various crops grown in a specific region.

Business Use Cases:

1. **Food Security and Planning**: Provide insights to governments and NGOs to plan food supply and reduce the risk of food shortages by predicting crop yields accurately.
2. **Agricultural Policy Development**: Assist policymakers in framing agricultural subsidies, crop insurance, and disaster relief programs by understanding potential production trends.
3. **Supply Chain Optimization**: Help agribusinesses and distributors plan storage, transportation, and market supply chains based on expected production volumes.
4. **Market Price Forecasting:** Enable farmers and traders to predict market trends and make informed decisions about selling crops at optimal times.

5. **Precision Farming**: Guide farmers in selecting the most suitable crops for their land and environmental conditions, optimizing resource usage like water and fertilizers.
6. **Agro-Technology Solutions**: Provide valuable data for agri-tech startups to design tools and applications that help farmers monitor and enhance crop production.

Approach:

1. Data Cleaning and Preprocessing

- Handle missing data and standardize column metrics.
- Filter relevant columns for analysis data.

2. Exploratory Data Analysis (EDA)

1. **Analyze Crop Distribution**
    - **Crop Types**: Study the distribution of the `Item` column to identify the most and least cultivated crops across regions.
    - **Geographical Distribution**: Explore the `Area` column to understand which regions focus on specific crops or have high agricultural activity.
2. **Temporal Analysis**
    - **Yearly Trends**: Analyze the `Year` column to detect trends in `Area harvested`, `Yield`, and `Production` over time.
    - **Growth Analysis**: Investigate if certain crops or regions show increasing or decreasing trends in yield or production.
3. **Environmental Relationships**
    - Although explicit environmental data is absent, infer relationships between the `Area harvested` and `Yield` to check if there's an impact of resource utilization on crop productivity.
4. **Input-Output Relationships**
    - Study correlations between `Area harvested`, `Yield`, and `Production` to understand the relationship between land usage and productivity.
5. **Comparative Analysis**
    - **Across Crops**: Compare yields (`Yield`) of different crops (`Item`) to identify high-yield vs. low-yield crops.
    - **Across Regions**: Compare production (`Production`) across different areas (`Area`) to find highly productive regions.

    - **Productivity Analysis**
        - Examine variations in `Yield` to identify efficient crops and regions.
        - Calculate productivity ratios: `Production/Area harvested` to cross-verify yields.
6. **Outliers and Anomalies**
    - Identify anomalies in `Yield` or `Production`, such as unusually high or low values, and correlate them with potential external factors like policies or environmental changes.

3. Task:

Predicting Production (`Production`, measured in tons):

- **Use Case**: Focuses on total output. It answers, *"What will the total production of a specific crop be for a given region and year?"*

Expected Results:

1. Analyze Trends: Gain insights into crop production trends by region, crop type, and year.
2. Predictive Modeling: Develop models to predict crop yield or production for informed decision-making.
3. Actionable Insights: Provide recommendations for resource allocation and agricultural planning.
4. Develop a streamlit application to showcase the predicted value based on selected features.

Project Evaluation metrics:

- Data Preparation
    - Completeness and accuracy of data cleaning.
    - Clear documentation of preprocessing steps.
- Exploratory Data Analysis
    - Depth and clarity of insights derived.
    - Use of relevant statistical methods.
- Data Visualization
    - Relevance and quality of visualizations.
    - Effective use of charts, maps, and interactivity.
- Model Performance
    - Metrics involving R2, MSE, MAE,etc.
    - Give comparison of different models.

- Business Insights
    - Actionable and stakeholder-focused findings.
    - Addressing key use cases like hotspots and trends.
- Presentation
    - Logical structure and clarity.
    - Effective communication of results.

Technical Tags:

- Data Cleaning
- Data Preprocessing
- Exploratory Data Analysis (EDA)
- Machine Learning
- Streamlit

Data Set:

Dataset: FAOSTAT_data

Data Set Explanation:

1. **Domain Code & Domain:**
    - **Domain Code:** Identifier for the data domain (e.g., QCL for crops and livestock).
    - **Domain:** The specific area of focus, such as "Crops and livestock products."
2. **Area Code (M49) & Area:**
    - **Area Code (M49):** Numerical code representing countries or regions (e.g., "4" for Afghanistan).
    - **Area:** Name of the country or region (e.g., Afghanistan).
3. **Element Code & Element:**
    - **Element Code:** Numeric code for the measured parameter (e.g., 5312 for area harvested).
    - **Element:** Description of the parameter (e.g., Area harvested, Yield, or Production).
4. **Item Code (CPC) & Item:**
    - **Item Code (CPC):** Classification code for the crop/product (e.g., 1371 for Almonds, in shell).
    - **Item:** The name of the crop/product (e.g., Almonds, in shell).
5. **Year Code & Year:**
    - **Year Code:** Numerical representation of the year.
    - **Year:** The calendar year for the recorded data.
6. **Unit & Value:**

    - **Unit:** Unit of measurement (e.g., ha for hectares, kg/ha for yield, t(tons) for production).
    - **Value:** The quantitative measure for the element and crop (e.g., 29203 hectares harvested).
7. **Flag & Flag Description:**
    - **Flag:** Coded indication of the data source or nature (e.g., "A").
    - **Flag Description:** Explanation of the flag (e.g., Official figure).

This dataset enables analysis of agricultural patterns, including area harvested, crop yield, and production by region and year.

Project Deliverables:

Cleaned Dataset:

- Final preprocessed dataset used for analysis.
- Description of cleaning steps (e.g., handling missing values, formatting).

Source Code:

- Python or other scripts used for data cleaning, analysis, model building and visualization.

Application:

- Streamlit Application showcasing predicted values

Documentation:

- A concise report explaining the approach, key findings, and actionable insights.
- Include explanations for visualizations, trends, and analyses performed.

Project Guidelines:

Follow the steps mentioned:

1) **Data Preparation**: Cleaning → Transforming → Feature Engineering → Encoding.
2) **Exploratory Data Analysis (EDA)**: Understanding distributions → Correlations → Relationships.
3) **Visualization**: Plot trends → Identify outliers → Spot patterns.

4) **Model Building**: Train → Validate → Test models.
5) **Evaluation & Insights**: Evaluate performance → Interpret results → Derive actionable insights

Code Quality:

- Write clean, well-commented, and reusable code.

Presentation:

- Keep the final presentation clear and concise.
- Emphasize key findings and actionable recommendations.

Timeline:

The project must be completed and submitted **within 7 days from the assigned date**.

References:

Project Orientation Recording (English)	Crop_Production_Prediction_Eng.mp4
Project Orientation Recording (Tamil)	Crop_Production_tamil
Project Live Evaluation	Project Live Evaluation
EDA Guide	Exploratory Data Analysis (EDA) Guide
Streamlit Reference	https://docs.streamlit.io/get-started/fundamentals/main-concepts
Capstone Explanation Guideline	Capstone Explanation Guideline

GitHub Reference	How to Use GitHub.pptx

PROJECT DOUBT CLARIFICATION SESSION (PROJECT AND CLASS DOUBTS)

About Session: The Project Doubt Clarification Session is a helpful resource for resolving questions and concerns about projects and class topics. It provides support in understanding project requirements, addressing code issues, and clarifying class concepts. The session aims to enhance comprehension and provide guidance to overcome challenges effectively.
Note: Book the slot at least before 12:00 Pm on the same day

Timing: Monday to Saturday (4:00PM to 5:00PM)

Booking link :https://forms.gle/XC553oSbMJ2Gcfug9

LIVE EVALUATION SESSION (CAPSTONE AND FINAL PROJECT)

About Session: The Live Evaluation Session for Capstone and Final Projects allows participants to showcase their projects and receive real-time feedback for improvement. It assesses project quality and provides an opportunity for discussion and evaluation.
Note: This form will Open on Saturday and Sunday Only on Every Week

Timing: Monday-Saturday (5:30PM to 6:30PM)

Booking link : https://forms.gle/1m2Gsro41fLtZurRA