INDIA AGRICULTURAL CROP PREDICTION, CLASSIFICATION AND OPTIMIZATION SYSTEM

Project Statement: Integrated Crop Yield Prediction and Optimization System

Agriculture forms the backbone of India's economy, with millions of farmers depending on it for their livelihoods. However, unpredictable weather patterns, uneven rainfall distribution, limited access to soil quality data, and suboptimal resource allocation often result in reduced crop yields and financial stress for farmers.

There is an urgent need to develop a system that empowers farmers and policymakers with actionable insights by leveraging advanced data analysis, machine learning, and deep learning techniques. By integrating diverse datasets such as crop statistics, weather forecasts, rainfall patterns, budget allocations, and soil quality images, this project aims to create a comprehensive solution for predicting crop yields and optimizing agricultural outcomes.

Such a system can support farmers in making informed decisions, ensure effective resource utilization, and ultimately contribute to food security and economic growth. Through the use of automation and model tracking tools like Airflow and MLflow, the project aspires to deliver scalable and accurate results in real-world agricultural scenarios.

Objectives:

1. **Data Engineering:**
    - Clean and integrate datasets (crop data, weather, rainfall, budget) into a unified schema.
    - Process soil sample images for feature extraction using deep learning techniques.
2. **Data Analysis:**
    - Explore and visualize the relationship between weather patterns, rainfall, and crop yield trends.
    - Analyze the impact of budget allocation on agricultural productivity across regions.

3. **Machine Learning:**
    - Develop predictive models to forecast crop yield based on weather, rainfall, and soil features.
    - Create classifiers based on your observations
    - Use MLflow to track model performance across multiple runs and configurations.
4. **Deep Learning:**
    - Train convolutional neural networks (CNNs) to classify soil quality from images.
    - Integrate soil quality predictions into the crop yield model to enhance accuracy.
5. **Model Deployment and Automation:**
    - Automate the end-to-end pipeline using Airflow for data ingestion, training, and evaluation.
    - Deploy the best model as a REST API for stakeholders to access predictions.
    - Use containerization technology for deployments

Challenges:

- Combining structured (tabular) and unstructured (image) data in a cohesive system.
- Managing and tracking experiments using MLflow.
- Designing a robust Airflow DAG for modular and automated pipeline execution.
- Deploying the final system with scalability and maintainability considerations.

Note:

- Put on your creative thinking cap and based on the above statement define your own end-goals and create a product that is useful.
- Get additional data and build additional models to improvise the system to achieve the end goal.
- You should able to defend your ideas and be able to “sell” the application (product) to your clients