Research Proposal

Crop Yield Prediction Using Open Satellite Data and Weather APIs

Submitted by: SANTHOSH KUMAR S

1. Introduction

Agriculture is a critical sector that can benefit immensely from digital transformation. With the increasing availability of open-source satellite imagery and weather data, predicting crop yield has become a realistic and impactful goal using data science. This project aims to build a machine learning-based prediction model that utilizes historical yield data, satellite imagery, and weather APIs to forecast crop yields accurately. Such a system can empower farmers, planners, and policy-makers to make data-driven decisions.

2. Problem Statement

Farmers often lack accurate tools to predict yields, relying instead on guesswork and traditional knowledge. This can result in over/underproduction, financial loss, or resource mismanagement. A scalable, data-driven solution is needed to provide timely and localized yield predictions using freely available remote sensing and climatic data.

3. Objectives

- To collect and preprocess satellite imagery and weather data.
- To analyze historical crop yield data in relation to climatic and remote sensing inputs.
- To develop a machine learning model to predict future crop yields.
- To visualize the results in a user-friendly dashboard or map interface.

4. Research Questions

- Can open-source satellite and weather data effectively predict crop yield?
- What machine learning models provide the most accurate yield predictions?
- How can these predictions be visualized and made accessible to end users?

5. Methodology

Data will be collected from NASA EarthData, Sentinel Hub, OpenWeatherMap API, and Indian government yield datasets. Preprocessing steps will include image normalization, feature engineering from weather data, and geo-tagging. Models such as Random Forest, XGBoost, and LSTM will be trained and evaluated. Apache Spark will be used for big data processing, while visualization will be done using Streamlit or Plotly dashboards.

6. Tools and Technologies

- Python (Pandas, NumPy, Scikit-learn, TensorFlow)
- Apache Spark for distributed data processing
- Google Earth Engine or OpenCV for satellite imagery
- OpenWeatherMap API
- Jupyter Notebook, Streamlit, Tableau for visualization

7. Expected Outcomes

- A working prediction model for crop yield based on real-world data
- An interactive dashboard or app showing prediction results
- Insights into how weather and satellite data correlate with yield trends

8. Novelty and Uniqueness

This research combines satellite remote sensing and weather APIs within a big data framework to predict crop yields. While each data source has been used independently, integrating them into a unified, scalable analytics platform for yield forecasting is both novel and highly practical.

9. Potential Applications

- Farm planning and advisory tools
- Government crop insurance and subsidy planning
- Food supply chain demand forecasting
- Academic and policy research in agri-tech