

WEEKLY REPORT 1

CSE623: Machine Learning Theory And Practice

Crop Yield Prediction Using Classical Machine Learning Models and Climatic Factors



The Visionary

Group Members:

1. Nishtha Mistry - AU2544001
2. Vidit Gandhi - AU2544010
3. Dev Patel - AU2544026
4. Shreyas Patel - AU2544025
5. Parth Thakkar - AU2544017

Academic Year 2025–2026

1 Introduction

This project aims to develop a crop yield prediction system using regression-based classical machine learning algorithms. The study emphasizes model comparison, interpretability, and feature importance analysis using climatic and agricultural attributes.

2 Work Done This Week

The following tasks were completed during this week:

- Thorough understanding of problem definition provided by the instructor.
- Identification of key objectives, deliverables, and expected outcomes.
- Study of two high-quality research papers related to crop yield prediction and machine learning in agriculture.
- Collection of the provided agricultural dataset from Kaggle.
- Preliminary inspection of dataset structure and variables.

3 Literature Review

A detailed review of relevant high-impact research papers was conducted to understand existing methodologies and best practices in crop yield prediction.

Paper 1: Random Forest for Global and Regional Crop Yield Prediction

Jeong et al. (2016), in their study published in *PLoS ONE*, proposed the use of Random Forest models for large-scale crop yield forecasting. The research demonstrated that ensemble tree-based methods effectively capture nonlinear interactions between climatic variables and crop production. The study concluded that Random Forest models outperform conventional regression approaches in terms of robustness and predictive accuracy.

Paper 2: Machine Learning in Agriculture – A Comprehensive Review

Kamilaris and Prenafeta-Boldú (2018), in their review published in *Computers and Electronics in Agriculture*, examined the application of machine learning techniques in agricultural systems. The study highlighted the effectiveness of regression and ensemble

models such as Support Vector Regression and Random Forest for yield prediction tasks, while emphasizing the importance of proper data preprocessing and feature selection. The findings support the adoption of classical machine learning approaches for accurate and interpretable crop yield estimation.

4 Dataset Collection

Source: Kaggle

The dataset contains agricultural records across Indian states and includes attributes such as:

- State
- Crop Type
- Season
- Area under cultivation
- Production
- Yield

5 Work to be Done Next Week

The following activities are planned for the upcoming week:

- Data cleaning and preprocessing.
- Handling missing values and duplicate records.
- Encoding categorical variables.
- Exploratory Data Analysis (EDA).
- Outlier detection and treatment.
- Feature scaling and data transformation.
- Preparation of training and testing datasets.