**CROP YIELD PREDICTION**

**PROJECT REPORT**

**1. Introduction**

This project focuses on predicting crop yield from soil and environmental features including Fertilizer quantity, Temperature, and nutrient levels (N, P, K). Accurate yield prediction can enhance agricultural planning and optimize resource allocation.

**2. Dataset**

The dataset contains records with the following features:

- Fertilizer (applied amount)

- Temperature (°C)

- Nitrogen (N) content

- Phosphorus (P) content

- Potassium (K) content

- Yield (target variable)

The original dataset was split into training (80%) and test (20%) subsets to enable model training and unbiased evaluation.

**3. Data Preparation**

* The data was checked for missing values and cleaned as needed.
* A random shuffle and an 80-20 train-test split were applied using stratified sampling to preserve data distribution.
* Features selected include the primary soil nutrients and temperature relevant to crop growth.
* No additional feature engineering or normalization was applied to keep the baseline model simple.

**4. Model Selection and Training**

A Random Forest Regressor was chosen due to its:

* Capability to model nonlinear relationships,
* Robustness to overfitting with appropriate parameters,
* Easy interpretability and fast training.

The model used 100 trees and was trained on the training data with a 10% subset reserved for validation.

**5. Results**

The model achieved a \*\*validation R² score of 0.9777\*\*, indicating it explains approximately 97.77% of the variance in the target variable on the validation set.

This is a very strong performance, signifying the model fits the validation data extremely well.

**6. Conclusion and Future Work**

This project demonstrates that Random Forest regression can effectively predict crop yield from soil and temperature features with high accuracy. Future improvements include:

* Adding weather and environmental variables such as rainfall and humidity,
* Experimenting with other algorithms like gradient boosting or neural networks,
* Applying feature engineering and scaling,
* Conducting hyperparameter tuning and cross-validation.

**7. References**

- Kaggle Crop Yield Prediction Dataset

<https://www.kaggle.com/datasets/gurudathg/crop-yield-prediction-using-soil-and-weather>

- Agricultural and machine learning literature

This report summarizes methodology, results, and interpretation clearly and concisely while highlighting the need for further validation despite the high score.