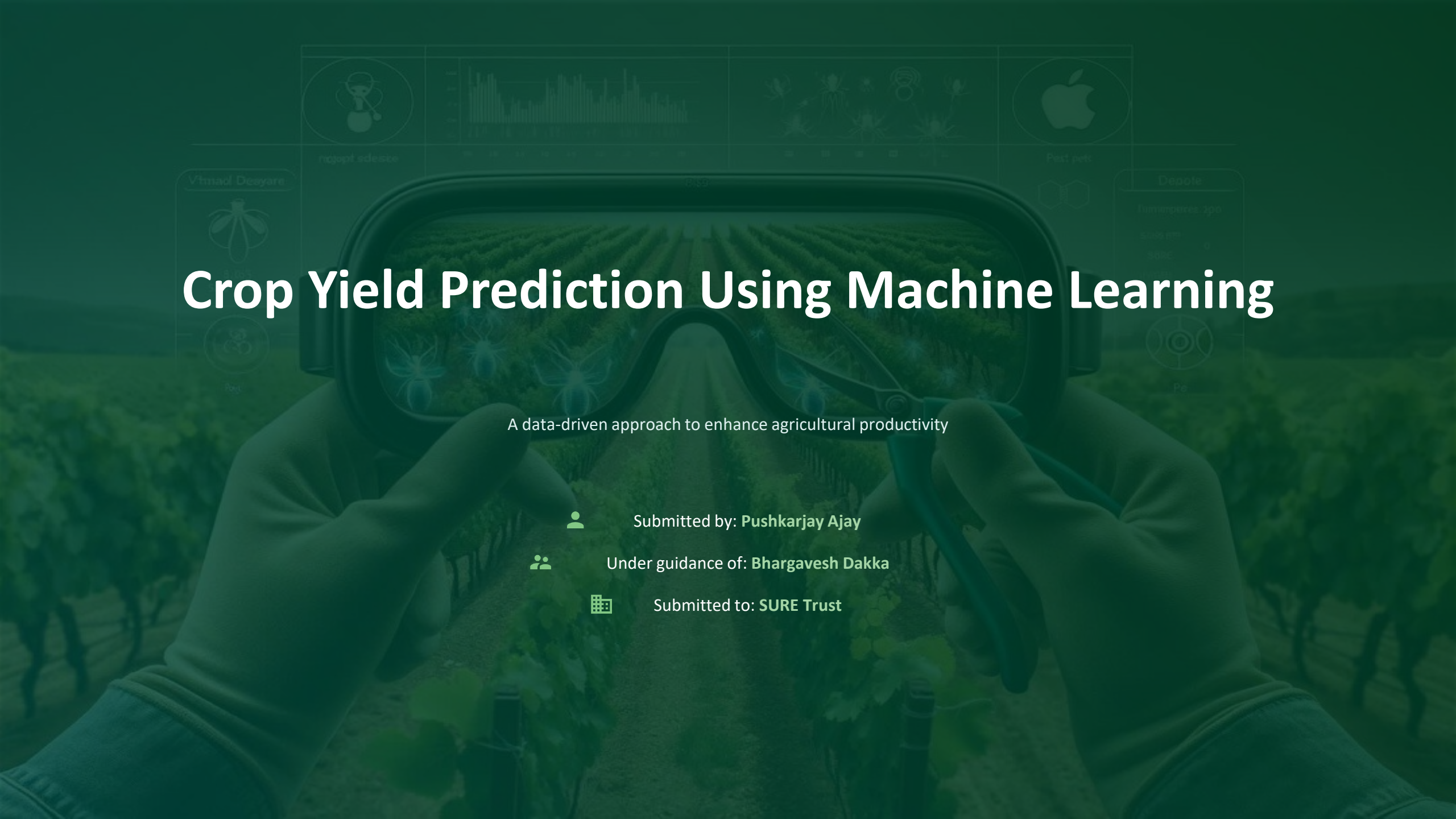


Crop Yield Prediction Using Machine Learning

A data-driven approach to enhance agricultural productivity

Submitted by: **Pushkarjay Ajay**

Under guidance of: **Bhargavesh Dakka**

Submitted to: **SURE Trust**

Agenda

01
Problem Statement

02
Objective of the Project

03
Dataset Overview

04
Data Preprocessing

05
Exploratory Data Analysis

06
ML Models & Performance

07
Feature Importance

08
Conclusion & Future Work

09
Acknowledgements

Using Machine Learning to predict crop yield and support smart agriculture

Problem Statement

Why Predict Crop Yield?

Essential for **food security**, economic planning, and resource optimization

Farmer Challenges

- **Climate variability** affecting growing conditions
- **Soil differences** across fields
- **Unknown yield** causing financial uncertainty

ML Solution

Data-driven predictions using environmental factors to forecast yield with higher accuracy

Objective of the Project

Predict Yield

Accurately forecast crop output in **Q/acre** using historical data

Analyze Conditions

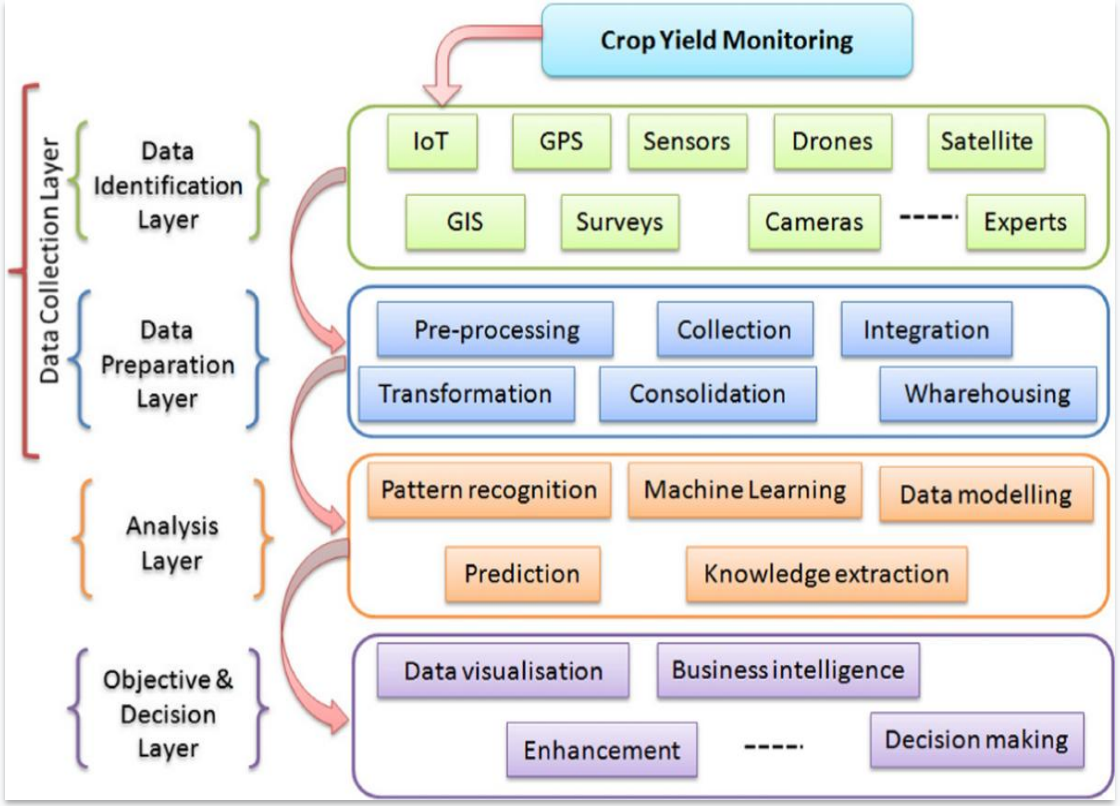
Examine **environmental** and **agricultural** factors affecting growth

Identify Key Factors

Determine which variables have the **strongest influence** on yield

Build ML Model

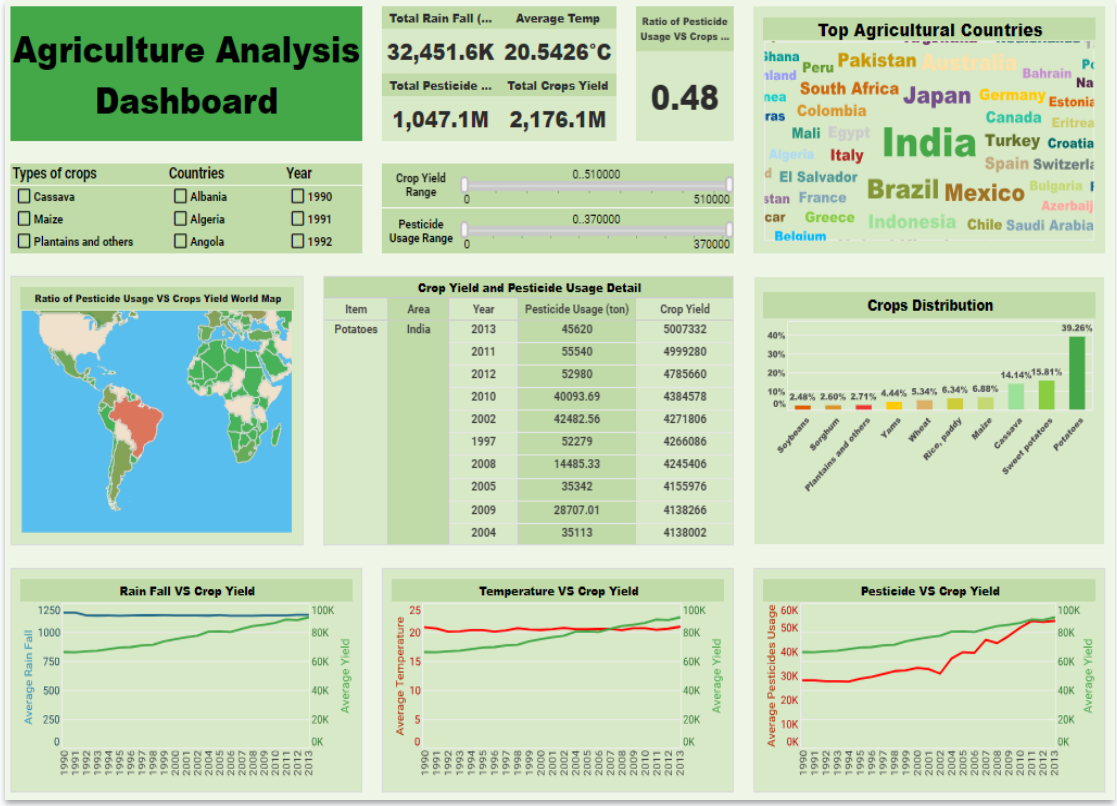
Create an **interpretable** model for practical farm use



Dataset Overview

Key Variables in Dataset

Variable	Unit	Description
Rainfall	mm	Annual precipitation
Temperature	°C	Average temperature
Fertilizer	kg	Amount applied
Nutrients	N, P, K	Nitrogen, Phosphorus, Potassium
Yield	Q/acre	Crop output per acre



Data Preprocessing Steps

Remove Invalid Values

Eliminated temperature entries with **":"** symbol

Data Type Conversion

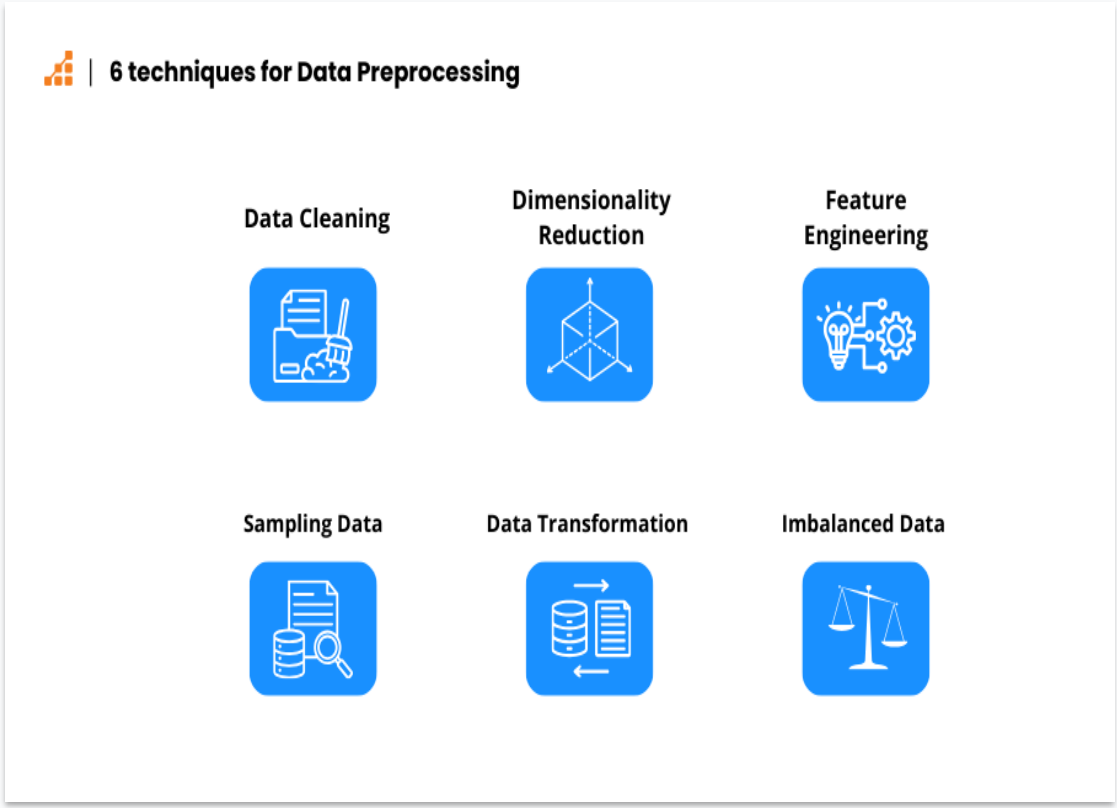
Converted temperature column to **float** for analysis

Handle Missing Values

Applied **median imputation** for missing data points

Ensure Clean Dataset

Verified all data is **numerical** and analysis-ready



Exploratory Data Analysis - Key Findings

Two Clusters

Data reveals **two distinct groups** suggesting different crop types or growing seasons

Bi-modal Patterns

Rainfall & Temperature show two distinct peaks in distribution

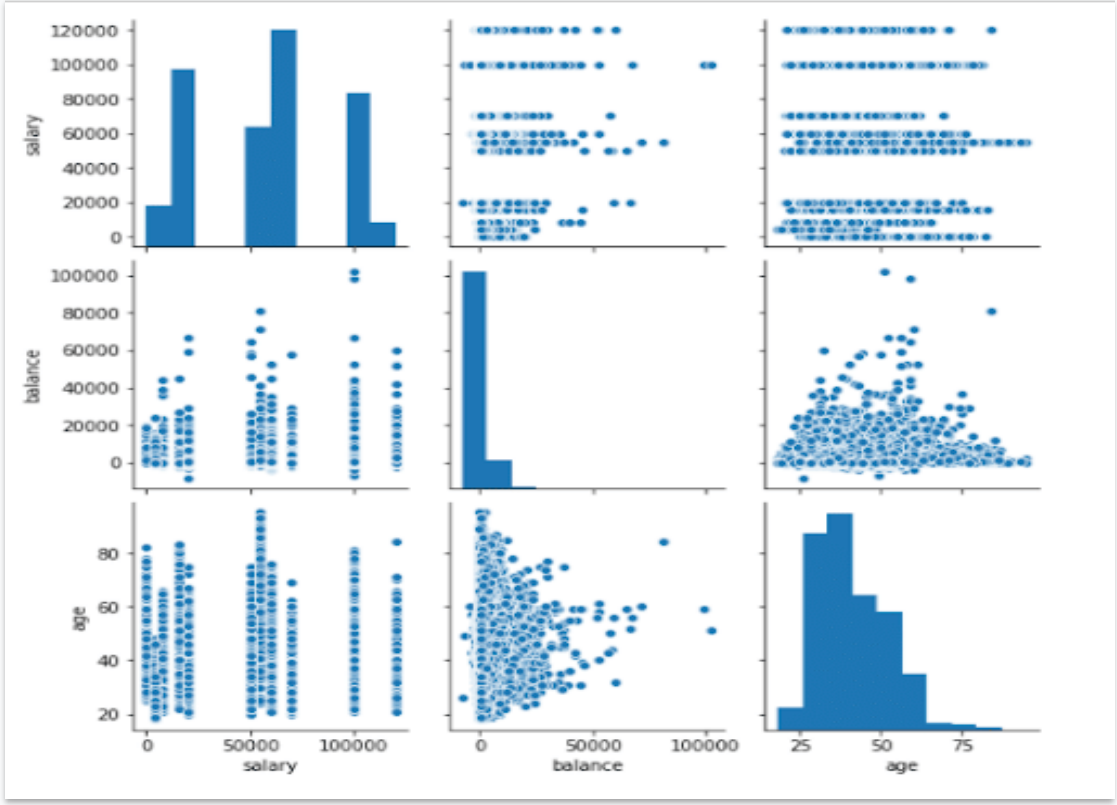
Nutrient Impact

NPK factors show **partial influence** on yield but are not dominant predictors

Key Predictor

Temperature emerges as the strongest factor affecting crop yield

Temperature is the most critical factor for yield prediction



Correlation Heatmap Summary

Temperature & Rainfall

Moderate correlation with crop yield, showing significant influence

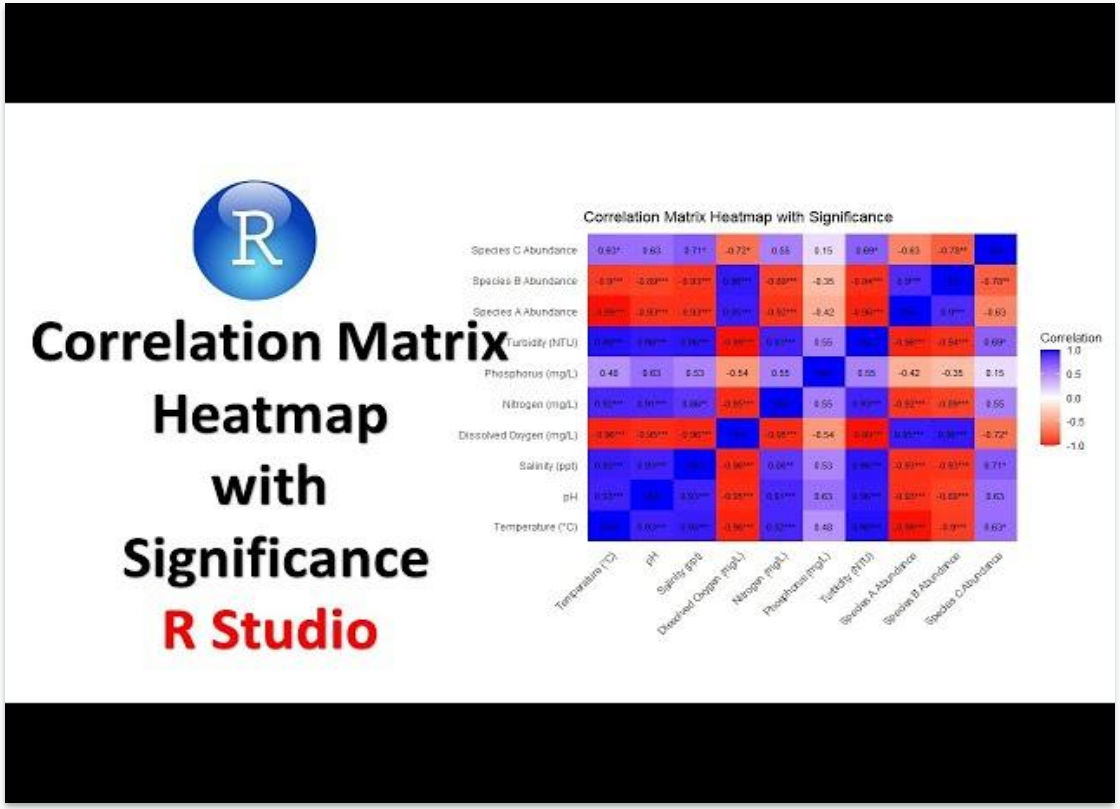
NPK Nutrients

Exhibit **smaller influence** on yield compared to climate factors

Non-linear Relationships

Dataset contains **complex patterns** requiring advanced modeling

Climate factors dominate nutrient effects on yield



ML Models Used

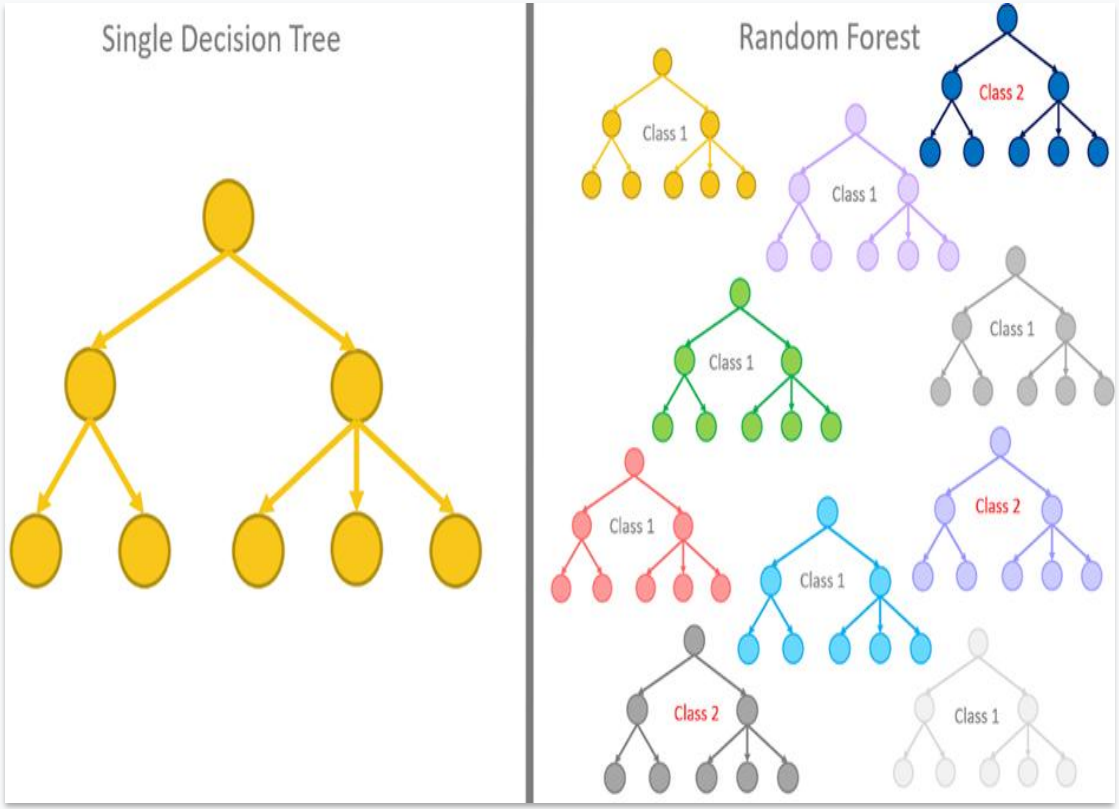
Decision Tree Regressor

Single tree model with **interpretable** structure

Random Forest Regressor

Ensemble of multiple trees with **higher accuracy**

Tree-based models chosen for their ability to handle **non-linear patterns** and **clustered data** in the dataset

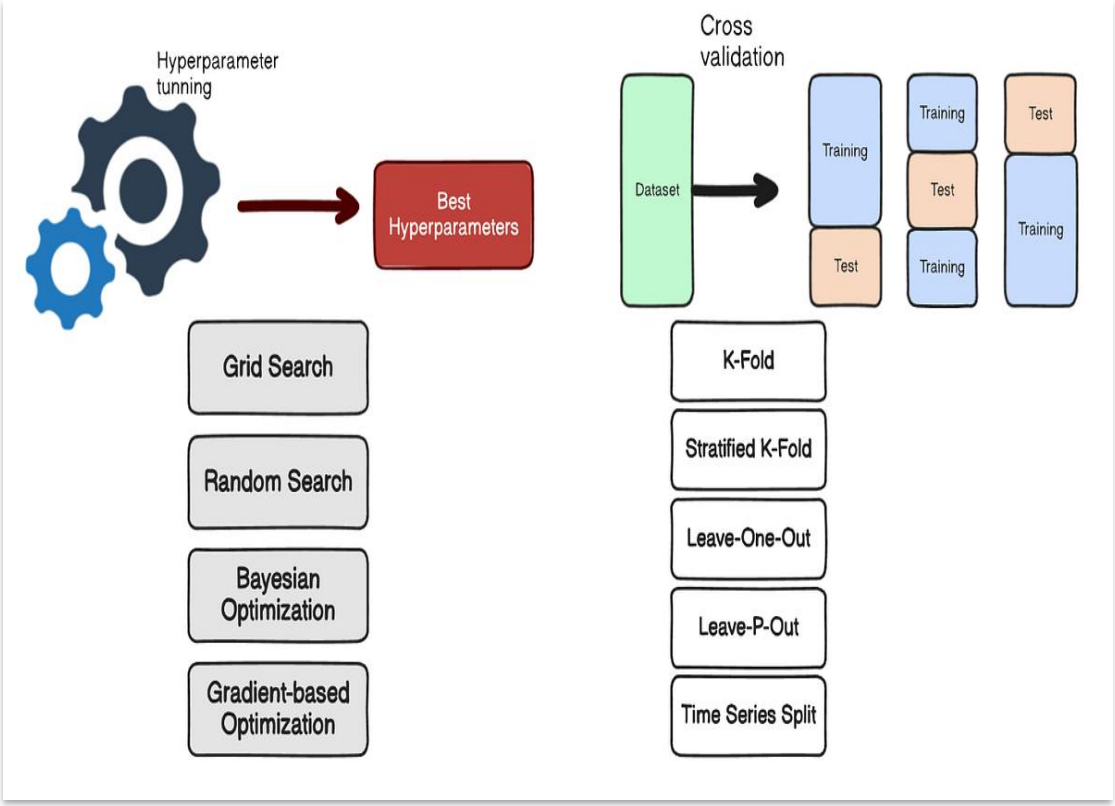


Hyperparameter Tuning (GridSearchCV)

Tuned Parameters

- **max_depth**
- **min_samples_split**
- **min_samples_leaf**
- **n_estimators**
- **random_state**

GridSearchCV systematically works through **multiple parameter combinations** to find the optimal model configuration



Model Performance Comparison

Decision Tree

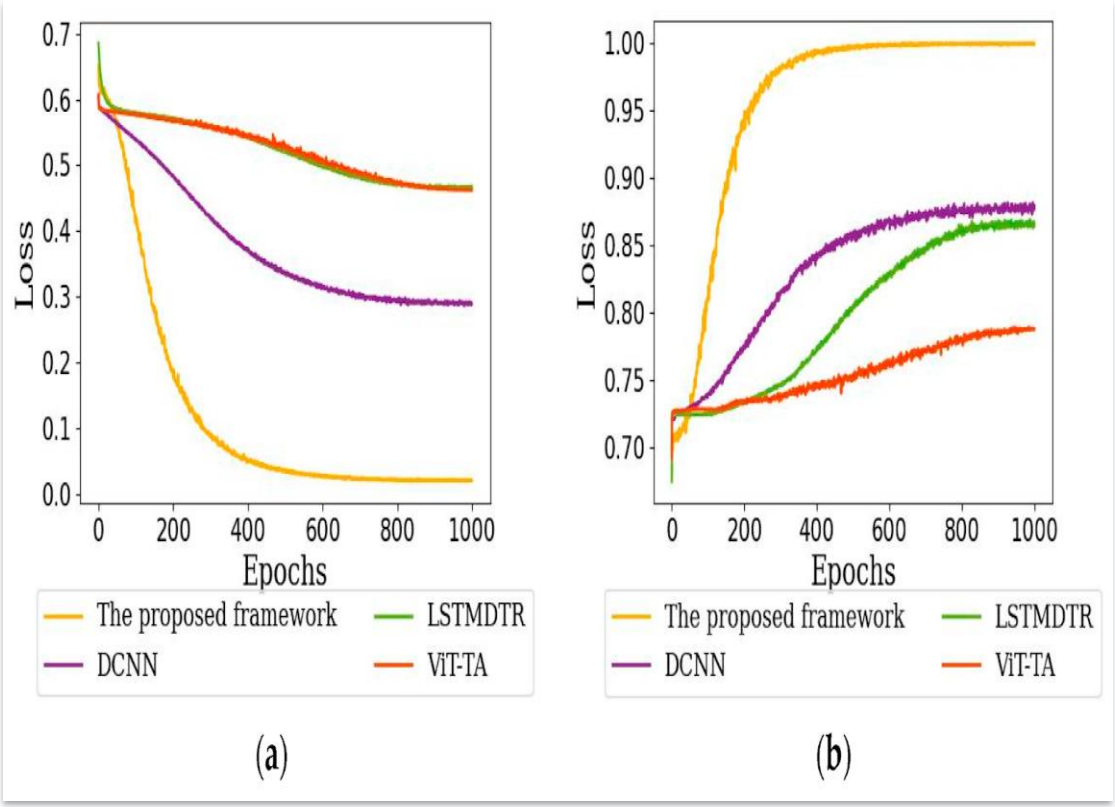
R^2	MAE
0.77	0.68

Random Forest

Better Model

R^2	MAE
0.802	0.68

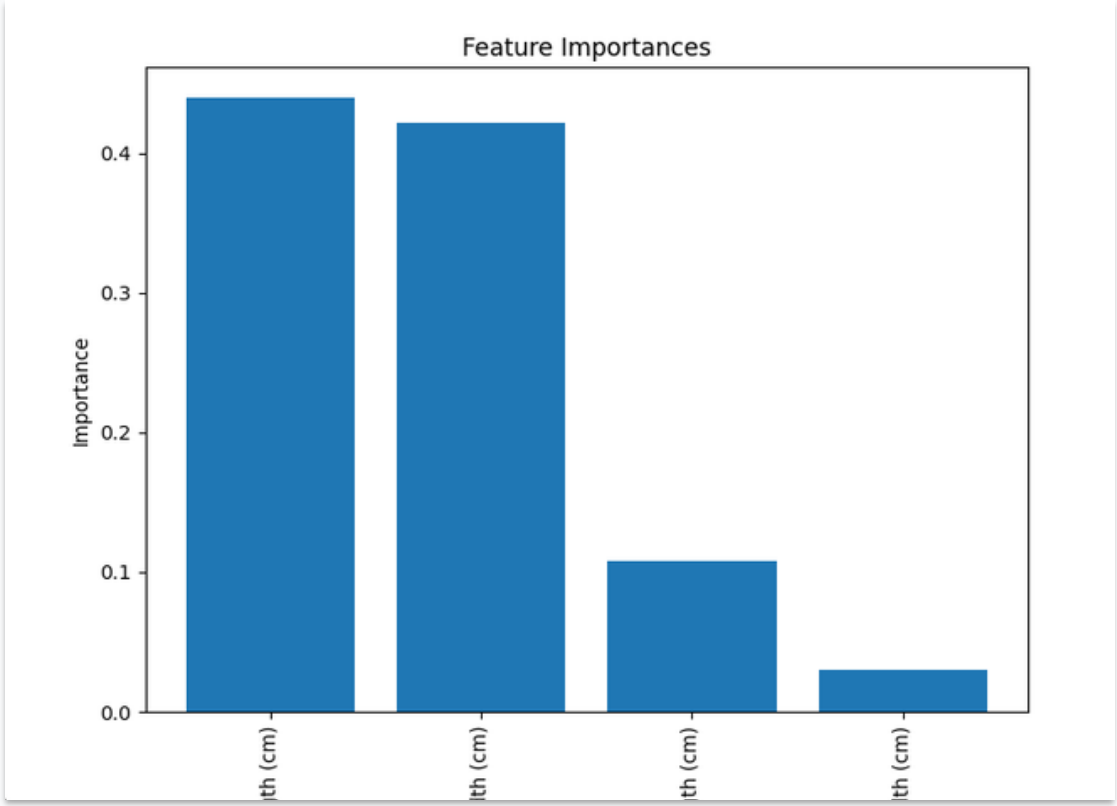
Random Forest shows **higher R^2 value** with same MAE, making it the superior model for yield prediction



Feature Importance Insights

1. Temperature
2. Rainfall
3. Fertilizer
4. Nitrogen (N)
5. Phosphorous (P)
6. Potassium (K)

Climate factors dominate nutrient effects in predicting crop yield



Conclusion

Two Crop Patterns

Dataset reveals **distinct clusters** indicating different crop types or seasons

Temperature Dominance

Highest influence on yield among all environmental factors

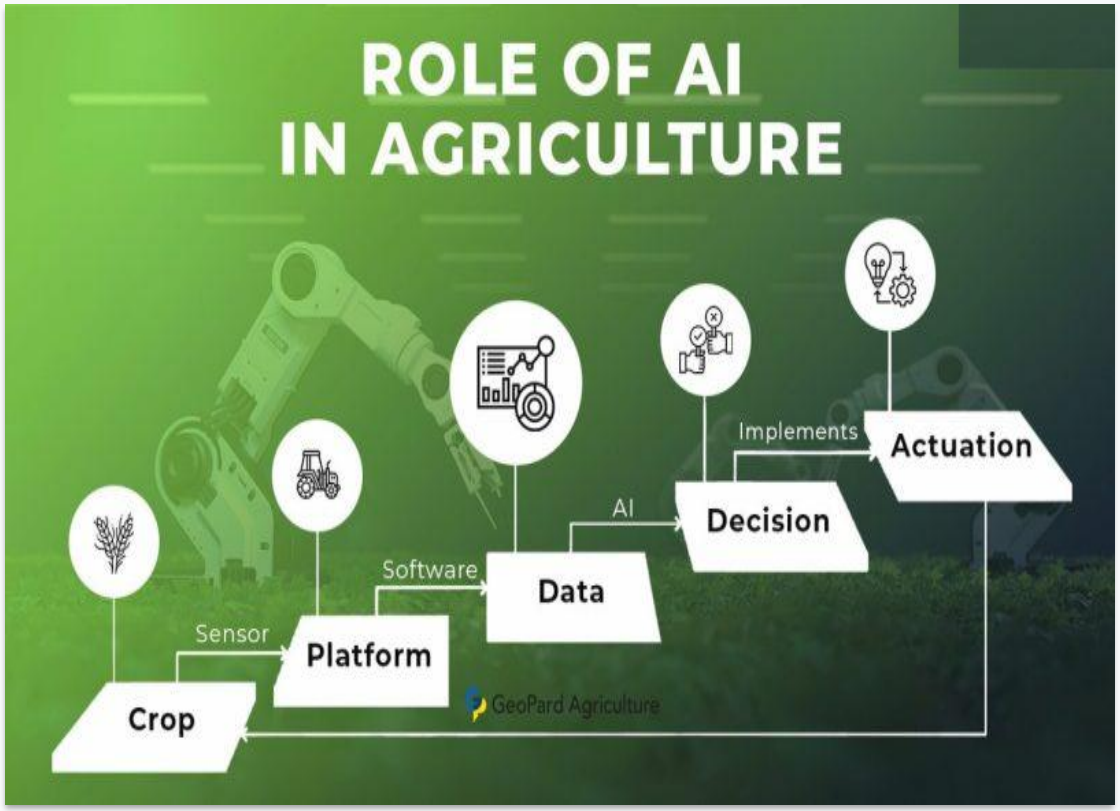
Best Performing Model

Random Forest outperforms Decision Tree with higher R^2 value

Smart Agriculture

ML models can support **data-driven decisions** for improved farming

Machine learning enables precise yield prediction for smarter agricultural planning



Limitations & Future Improvements

Missing Soil Data

Lack of comprehensive **soil parameters** affecting prediction accuracy

Environmental Factors

Missing **humidity**, **sunlight**, and **soil pH** measurements

Crop Identification

No **crop type labels** to differentiate between various crops

Advanced Models

Potential to add **boosting models** (XGBoost, CatBoost)

Build interactive web dashboard for farmers to access predictions

Acknowledgements

SURE Trust

For providing the platform and resources to pursue this project

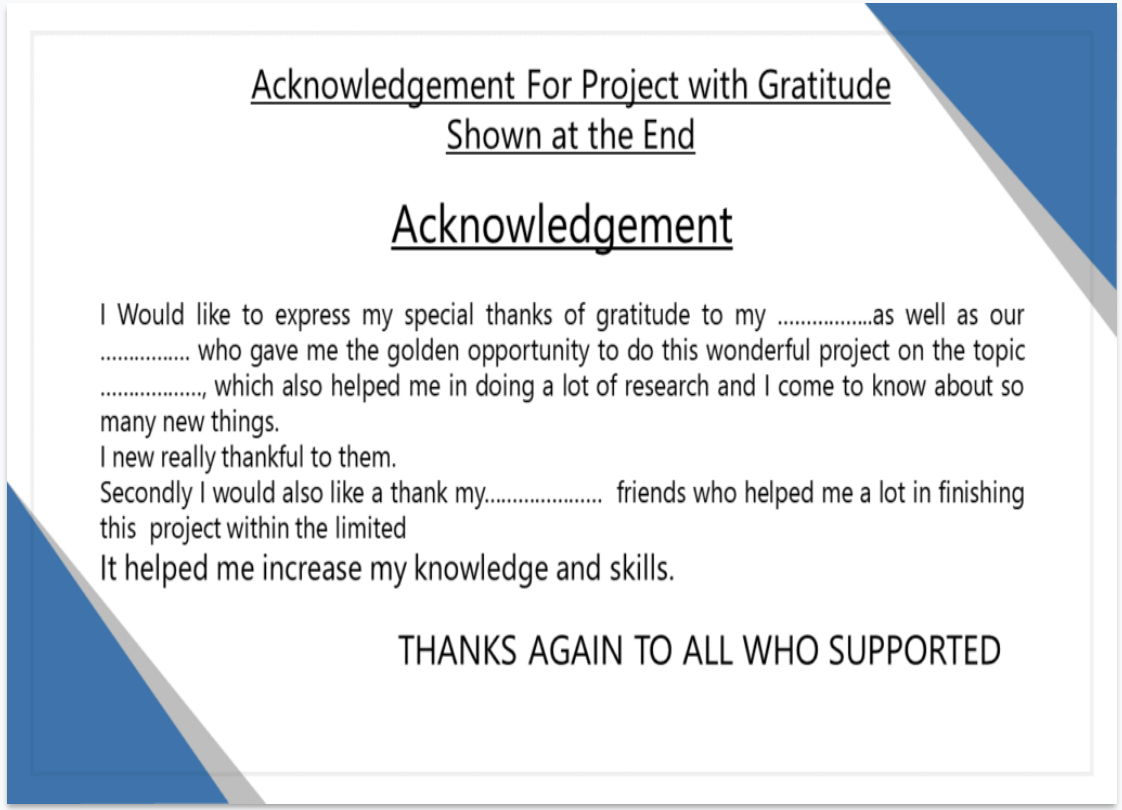
Instructor: Bhargavesh Dakka

For invaluable guidance, mentorship, and technical expertise

Support & Guidance

To all who provided feedback and encouragement throughout this journey

Sincere gratitude for making this project possible



Acknowledgement For Project with Gratitude Shown at the End

Acknowledgement

I Would like to express my special thanks of gratitude to my .....as well as our ..... who gave me the golden opportunity to do this wonderful project on the topic ....., which also helped me in doing a lot of research and I come to know about so many new things.
I new really thankful to them.
Secondly I would also like a thank my..... friends who helped me a lot in finishing this project within the limited
It helped me increase my knowledge and skills.

THANKS AGAIN TO ALL WHO SUPPORTED

Thank You

For your attention and interest in our Crop Yield Prediction project

Project Repository

Scan to access the complete code, dataset, and documentation on GitHub

Pushkarjay Ajay

Pushkarjay.ajay1@gmail.com