# Year of Millets - 2023 - A Technological Intervention

"Enhancing Millet Yield through Data Analysis and Machine Learning"

## Mini Project II

Under the guidance of -

**Dr. Ramesh Athe**

Assistant Professor at IIIT Dharwad

# Introduction

- The global food system faces complex challenges: hunger, malnutrition, population growth, climate change, and resource depletion.
- Millets have untapped potential as affordable, nutritious food, and climate-resilient crops.
- IYM 2023 is a chance to promote millets for Better production, Better Nutrition, a Better Environment, and a Better Life.
- Millets can enhance global food systems, benefiting smallholder farmers, nutrition, and the environment.
- We are developing a ML model to estimate millet yields based on historical data and weather forecasts. The model will be trained on a dataset of millet yields from different regions around the world, along with weather data from those regions.

# Hypothesis

- Enhancing millet production and yield through data-driven methods is our central hypothesis.
- Our goal is to find areas for improvement and create more effective strategies for boosting millet production.
- We believe that understanding historical patterns and correlations can lead to better and sustainable millet cultivation.
- We base our hypothesis on the idea that data-informed decision-making can increase millet production, benefiting food security and initiatives like the Year of Millets.

# Objectives

- Data Analysis and Trend Identification
- Predictive Modeling for Yield Forecasting
- Precision Agriculture Recommendations
- Data-Driven Policy Recommendations

# About Dataset

- The APY dataset on MilletStats.com is a comprehensive collection of area, production, and yield (APY) statistics for various millet crops in India.
- It covers data for four major millets (sorghum, pearl millet, finger millet, and minor millets) at both the all-India and state levels.
- The dataset spans from 1966-67 to 2022-23, providing a long-term perspective on millet production and productivity.
- Data is sourced from multiple open-source databases, including the Directorate of Economics and Statistics, Ministry of Agriculture and Farmers' Welfare, Government of India.

# SOME DATA POINTS

| | Crop | Year | State | Area | Production | Yield |
|---|---|---|---|---|---|---|
| **0** | Finger Millet (Ragi) | 1966-67 | All India | 1984.20 | 1630.60 | 821.79 |
| **1** | Finger Millet (Ragi) | 1967-68 | All India | 2291.20 | 1884.20 | 822.36 |
| **2** | Finger Millet (Ragi) | 1968-69 | All India | 2238.20 | 1648.00 | 736.31 |
| **3** | Finger Millet (Ragi) | 1969-70 | All India | 2783.40 | 2117.20 | 760.65 |
| **4** | Finger Millet (Ragi) | 1970-71 | All India | 2472.40 | 2155.00 | 871.62 |
| **...** | ... | ... | ... | ... | ... | ... |
| **4791** | Sorghum (Jowar) | 2015-16 | West Bengal | 0.04 | 0.00 | 0.00 |
| **4792** | Sorghum (Jowar) | 2016-17 | West Bengal | 0.03 | 0.00 | 0.00 |
| **4793** | Sorghum (Jowar) | 2017-18 | West Bengal | 0.04 | 0.00 | 0.00 |
| **4794** | Sorghum (Jowar) | 2018-19 | West Bengal | 0.06 | 0.03 | 469.00 |
| **4795** | Sorghum (Jowar) | 2019-20 | West Bengal | 0.18 | 0.10 | 528.00 |

4796 rows × 6 columns

# METHODOLOGY

- Data Preprocessing
- Statistical Analysis
- Machine Learning model building
- Performance Evaluation

# Data Preprocessing

- Load the dataset into a pandas DataFrame
- changing the names of the columns
- Check for missing values in each column
- Fill missing values with 0
- Convert categorical variables into numerical representations using one-hot encoding or label encoding.
- Detect and remove outliers that might affect the analysis..

# Statistical Analysis

- Descriptive Statistics:
- Mean yield: 628.518263 (higher than median yield, indicating a right-skewed distribution with high-end outliers).
- Standard deviation of yield: 1355.301635 (high variability in yield).
- Range of yield: 0 to 12898.4 (significant variation from low to high yield).

# Statistical Analysis

- Correlation Analysis:
- Weak positive correlation (r = 0.138161) between production and yield.
- Crops with higher production tend to have higher yields.
- Other factors influence yield besides production.

# Statistical Analysis

- Overall Insights:
- The yield distribution is right-skewed with high variability.
- There is a weak positive correlation between production and yield.

# Machine Learning model building

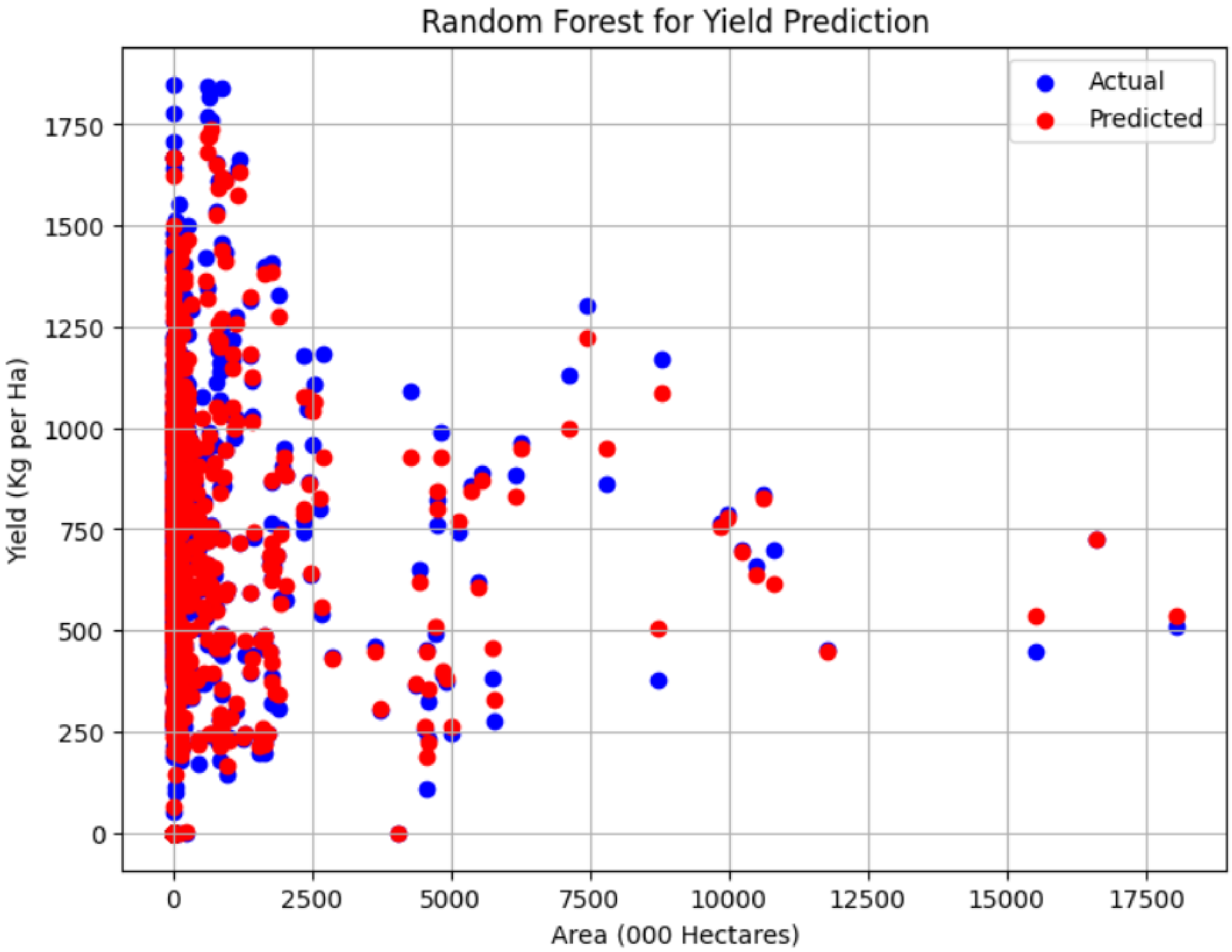
- We compared several regression models, but two models produced the best results.

1. Random Forest
2. XGBoost

# Random Forest

- **Conclusion based on Random Forest**

1. There is a positive correlation between the area and yield of finger millet.
2. The states with the highest yields tend to have a higher area under cultivation.
3. There is a significant amount of variation in yield, even for states with similar areas under cultivation.




# RANDOM FOREST

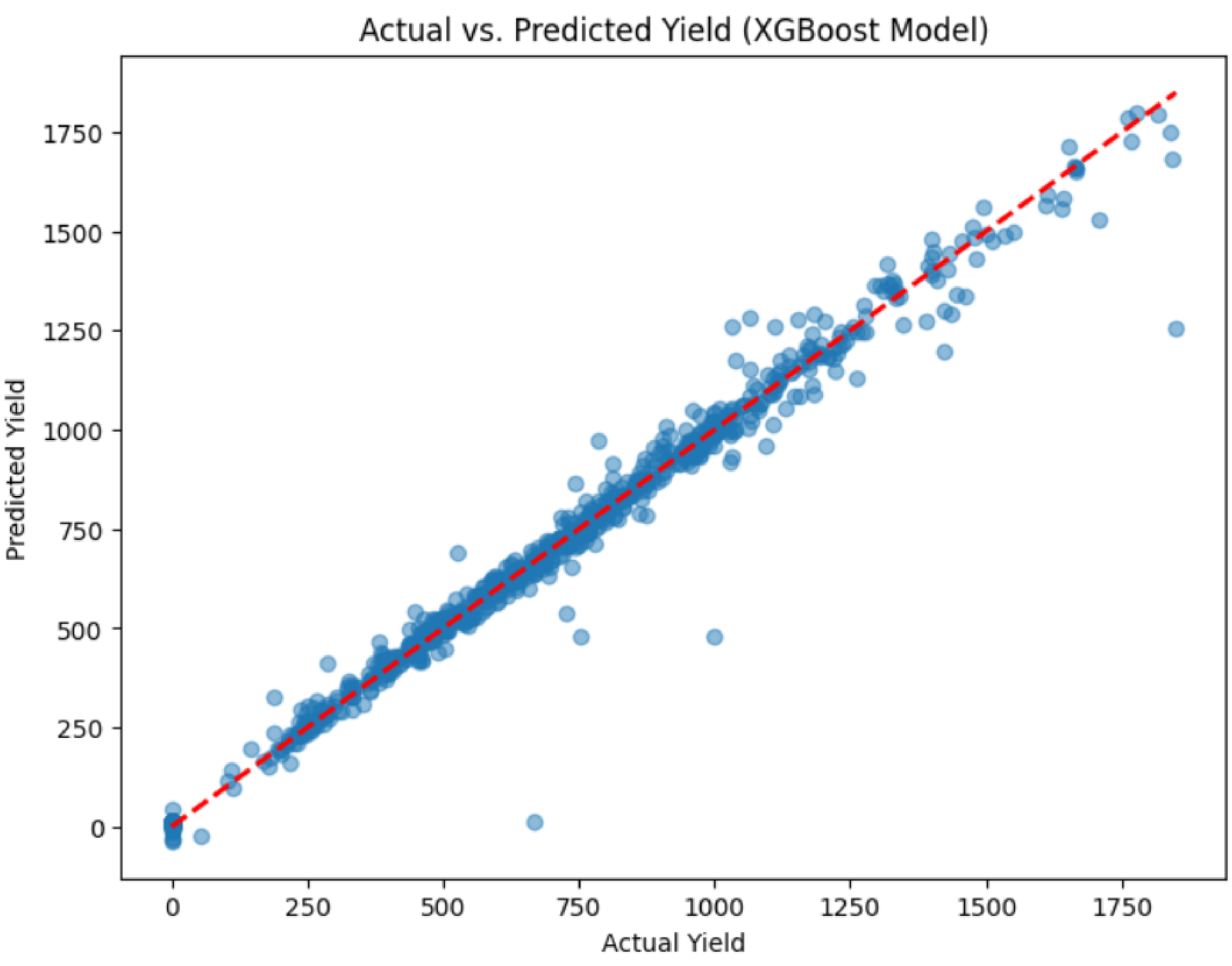
# Random Forest

- **Model Evaluation**
- Mean Squared Error : 1979.8784354025497
- R-squared : **0.9886770288925283**
- Accuracy : **0.9886770288925283**

# XGBoost

- **Conclusion based on XGBoost**

1. The model's accuracy of 0.9859 signifies its ability to predict field yield with 98.59% precision.
2. A strong positive correlation is observed between field area and yield, and the model accurately captures this relationship, making accurate predictions for fields of various sizes.
3. The model's best hyperparameters include a learning rate of 0.2, a maximum depth of 5, and 300 estimators, ensuring both accuracy and efficiency.




# XGBoost MODEL

# XGBoost

- **Model Evaluation**
- Mean Squared Error : 2467.3385272899704
- R-squared : **0.9858892332189193**
- Accuracy : **0.9858892332189193**

# Intervention: Recommendations

- Integrated Cereals Development Programmes in Coarse Cereals (ICDP-CC) under Macro Management of Agriculture (MMA).

- Initiative for Nutritional Security through Intensive Millet Promotion (INSIMP) as a part of Rashtriya Krishi Vikas Yojana (RKVY).

- Rainfed Area Development Programme (RADP) under RKVY.

# Intervention: Government Schemes

- Pradhan Mantri Fasal Bima Yojana (PMFBY): Crop insurance scheme for financial protection against crop losses.
- National Mission for Sustainable Agriculture (NMSA): Promotes sustainable agriculture practices.
- Rashtriya Krishi Vikas Yojana (RKVY): Promotes agricultural growth in India.

# Submitted by


20BDS001 Abhishek Shridhar Mane

20BDS029 Suyash Nishikant Kamble

20BDS041 Jasmeetsingh Pothiwal

20BDS044 Rishabh Singh

# THANK YOU!