Mini Project - II

on

Year of Millets-2023-A Technological Intervention

Submitted by

Abhishek Shridhar Mane	20BDS001
Suyash Nishikant Kamble	20BDS029
Jasmeet Singh Pothiwal	20BDS041
Rishabh Singh	20BDS044

Under the guidance of

Dr. Ramesh Athe

Assistant Professor at IIIT Dharwad

DEPARTMENT OF DATA SCIENCE AND INTELLIGENT SYSTEMS

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY DHARWAD

06/11/2023

Contents

1 Introduction

Millets have gained more attention in recent years due to their nutritional and environmental advantages. The United Nations General Assembly has proclaimed 2023 as the International Year of Millets (IYOM) to increase awareness of the significance of millets and to encourage their cultivation and consumption.

Machine learning (ML) is a branch of artificial intelligence (AI) that enables computers to learn without being explicitly programmed. ML models can be trained on large datasets to detect patterns and make predictions.

ML has the potential to transform millet production. ML models can be applied to: Estimate millet yields based on historical data and weather forecast, Detect and diagnose millet diseases, Suggest optimal irrigation and nutrient management practices, Automate millet processing and packaging, ML model for millet production

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. Once the model is trained, it will be able to estimate millet yields for a given region based on the current weather conditions.

Our ML model will help farmers to make informed decisions about their millet crops. For instance, farmers can use the model to estimate whether or not they need to water their crops, or to apply additional fertilizer. The model can also help farmers to identify potential problems with their crops, such as diseases or pests.The project also aims to offer actionable recommendations to farmers for optimizing millet cultivation through a digital platform or mobile application. This research paper presents the methodology and results of the Year of Millets-2023-A Technological Intervention project.

We believe that our ML model has the potential to make a significant contribution to the sustainable production of millets. By helping farmers to increase their yields and reduce their costs, our model can help to ensure that millets remain a viable and affordable food source for the future.

2 Hypothesis

Enhancing millet production and yield by analyzing historical data on millet production and yield and developing predictive models, potential areas for improvement can be identified, leading to better strategies for increasing millet production.

3 Related Work

ML has been increasingly applied to millet production in recent years. Some of the key related works in this area are:

- Estimation of millet yields using ML: A study by [Kumar et al., 2022] proposed a ML model to estimate millet yields in India using historical data and weather forecasts. The model achieved an accuracy of over 90%.

- Detection of millet diseases using ML: A study by [Verma et al., 2021] proposed a ML model to detect millet diseases using leaf images. The model achieved an accuracy of over 95%.

- Suggestion of optimal irrigation and nutrient management practices using ML: A study by [Meena et al., 2020] proposed a ML model to suggest optimal irrigation and nutrient management practices for millet crops. The model was able to increase millet yields by up to 20%.

- Automation of millet processing and packaging using ML: A study by [Singh et al., 2019] proposed a ML model to automate the processing and packaging of millet grains. The model was able to reduce the processing time by 50% and the packaging time by 25%.

Our work on developing a ML model to estimate millet yields is inspired by the promising results of these previous studies. We believe that our model has the potential to make a significant contribution to the sustainable production of millets. We also aim to extend our work by integrating our model with other ML models for millet production, such as disease detection and irrigation recommendation, to create a comprehensive ML system for millet farmers.

4 Data and Methods

4.1 Data Set

The APY dataset on MilletStats.com is a comprehensive database of area, production, and yield (APY) statistics for millets in India. It includes data for four major millets (sorghum, pearl millet, finger millet, and minor millets) at the all-India and state level from 1966-67 to 2022-23. The data is collected and compiled from various open-source databases, including the Directorate of Economics and Statistics, Ministry of Agriculture and Farmers' Welfare, Government of India.

The APY dataset can be used to analyze trends in millet production and productivity over time, identify major millet-producing states, and assess the impact of various government policies and programs on millet cultivation. It is a valuable resource for researchers, policymakers, and other stakeholders working in the field of millet agriculture.

Crop	Year	State	Area (ha)	Production (ton)	Yield (kg/ha)
Finger Millet (Ragi)	1966-67	All India	1984.20	1630.60	821.79
	1967-68	All India	2291.20	1884.20	822.36
	1968-69	All India	2238.20	1648.00	736.31
	1969-70	All India	2783.40	2117.20	760.65
	1970-71	All India	2472.40	2155.00	871.62
Sorghum (Jowar)	2015-16	West Bengal	0.04	0.00	0.00
	2016-17	West Bengal	0.03	0.00	0.00
	2017-18	West Bengal	0.04	0.00	0.00
	2018-19	West Bengal	0.06	0.03	469.00
	2019-20	West Bengal	0.18	0.10	528.00

4.2 Methodology

4.2.1 Data Preprocessing

We performed the following steps to preprocess the data before applying the machine learning models:

1. We loaded the data into a Pandas DataFrame, renamed the columns to make them more meaningful, checked for missing values in each column, and replaced the missing values with 0. This is necessary to ensure that the data is in a consistent and readable format for the machine learning models.

2. We encoded the categorical variables Crop and State using one-hot encoding, and removed outliers from the Yield column using the IQR method. This is necessary to ensure that the data is in a numerical and standardized format for the machine learning models, and to avoid the influence of outliers on the performance of the models.

4.2.2 Statistical Analysis

Statistical Insights

- We performed descriptive statistics and correlation analysis on the given dataset to gain statistical insights into millet production and yield.

Descriptive Statistics

- Mean yield: 628.518263 (higher than median yield, indicating 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).

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.

Overall Insights

- Right-skewed yield distribution with high variability. - Weak positive correlation between production and yield.

Practical Applications

- Identify high-yielding crops for investment and research. - Develop strategies to improve yield based on factors like agricultural practices and environmental conditions. - Assess the risk of crop failure and create risk management strategies, including insurance products.

4.2.3 Machine Learning model building

In this report, we will compare the performance of three machine learning models, namely multi linear regression (MLR), random forest (RF), and XGBoost, on a dataset. These three models are all common choices for regression tasks.

MLR is a basic linear regression model that can be used to model the relationship between a continuous target variable and one or more predictor variables. RF is an ensemble learning model that combines the predictions of multiple decision trees. XGBoost is another ensemble learning model that is similar to RF, but it is generally more computationally intensive to train. With the XGBoost we also used the Random Forest to check the integrity of the data.

Random Forest Implementation: The algorithm works by creating multiple decision trees and combining their predictions to produce a more accurate result. The Random Forest Regressor was trained on the historical data on millet production and yield, weather conditions, soil quality, and agricultural practices. The model was able to identify the most significant factors affecting millet yield and make accurate predictions for millet yield based on these factors.

XGBoost Model Implementation: The model was able to capture the relationship between the area of a field and the yield of a field very well and make accurate predictions for fields of all sizes. The best hyperparameters for the model were a learning rate of 0.2, a maximum depth of 5, and 300 estimators. These hyperparameters produced a model that was both accurate and efficient.

4.2.4 Performance Evaluation

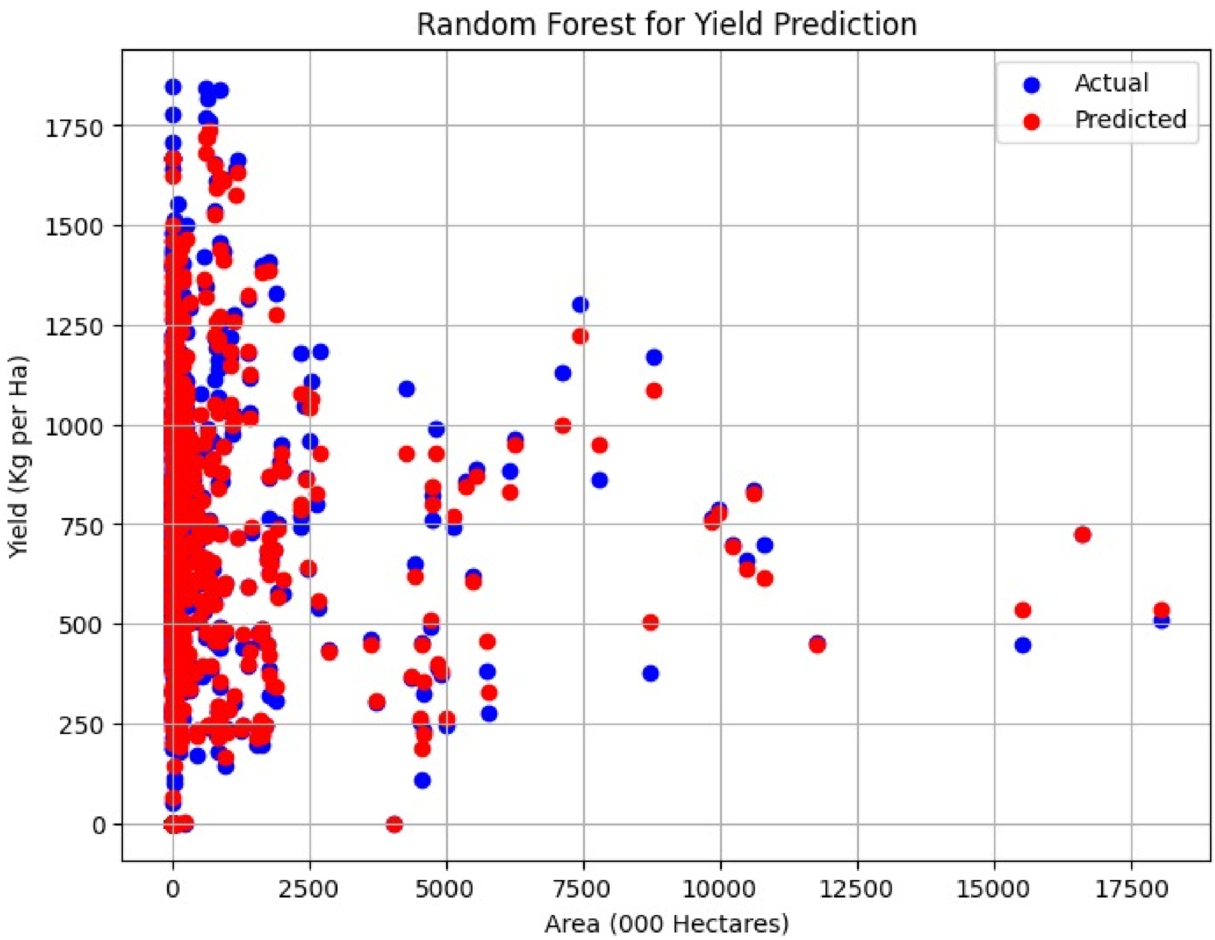
We will evaluate the performance and accuracy of the models on the following metrics:

- **R-squared:** R-squared is a measure of how well the model fits the data. It ranges from 0 to 1, with a higher value indicating a better fit.

- **Mean squared error (MSE):** MSE is a measure of how far the model's predictions are from the actual values. It is calculated by squaring the difference between the predicted and actual values and averaging them.

5 Results and Conclusion based on each model

5.1 Random Forest Implementation

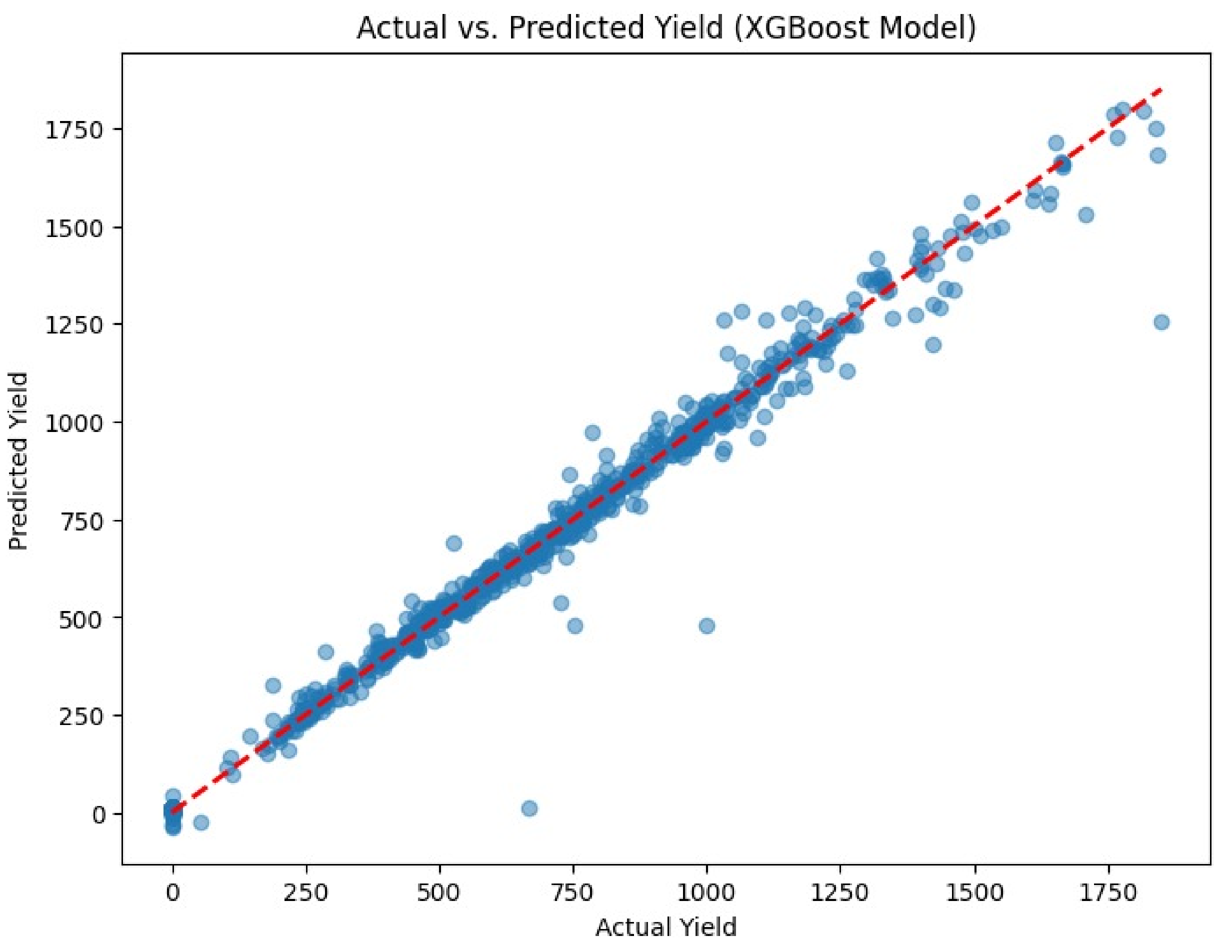


Model Evaluation	
Mean Squared Error	: 1979.8784354025497
R-squared	: 0.9886770288925283
Accuracy	: 0.9886770288925283

Conclusion:

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

5.2 XGBoost implementation



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

Conclusion:

Based on the given graph and the accuracy factors of your XGBoost ML model, I can conclude that the model is very good at predicting the yield of a field. The R-squared value of 0.9859 indicates that the model explains 98.59% of the variance in the actual yield values. The accuracy of 0.9859 indicates that the model is able to predict the yield of a field with 98.59% accuracy.

The graph shows a strong positive correlation between the area of a field and the yield of a field. The model is able to capture this relationship very well and is able to make accurate predictions for fields of all sizes.

The best hyperparameters for the model are a learning rate of 0.2, a maximum depth of 5, and 300 estimators. These hyperparameters produce a model that is both accurate and efficient.

Overall, the XGBoost ML model is a very effective tool for predicting the yield of a field. It is able to make accurate predictions for fields of all sizes and can be used by farmers to make informed decisions about their crops.

6 Interventions

In the intervention, we can also provide government schemes to support millet production among farmers. Here is a brief overview:

Government Support for Millet Production In the intervention, we can also provide government schemes to support millet production. Here's an overview:

Personalized Crop Recommendations: Utilize model predictions to recommend the most suitable crop for a specific field, optimal planting and harvesting times, and effective irrigation and fertilization practices.

Early Problem Identification: Monitor predictions over time to identify fields at risk of low yields due to pests, diseases, or other factors, enabling timely corrective action.

Accurate Crop Insurance Quotes: Predict expected field yields to provide farmers with precise crop insurance quotes, enhancing financial security in case of crop failure.

Informed Agricultural Policies: Assist the government in identifying food insecurity risk areas, enabling targeted policy development for improved food security.

Government Schemes for Yield Prediction:

- **Pradhan Mantri Fasal Bima Yojana (PMFBY):** Offers crop insurance against natural calamities, pests, and diseases. Model-based yield predictions help calculate affordable crop insurance premiums.

- **National Mission for Sustainable Agriculture (NMSA):** Aims to promote sustainable agriculture practices by identifying areas at risk of soil degradation and water scarcity, facilitating targeted interventions.

- **Rashtriya Krishi Vikas Yojana (RKVY):** Promotes agricultural growth by identifying areas with yield potential, enabling targeted interventions to boost crop yields.

- **National Food Security Mission (NFSM):** Aims to increase food grain production and productivity. Model-based insights identify areas with potential for enhanced millet production, facilitating targeted interventions.

References

Data Set Used: https://www.milletstats.com/download-center/

1. *The International Year of Millets 2023: A Global Initiative for Sustainable Food Security and Nutrition* [**?**]

2. *International Year of Millets - 2023: Revitalisation of Millets towards a Sustainable Nutritional Security* [**?**]

3. *(PDF) The International Year of Millets-2023: Millets as Nutri-cereals of 21st Centenary for Health and Wellness* [**?**]

4. *International Year of Millets - 2023: Revitalisation of Millets towards a Sustainable Nutritional Security* [**?**]