WEEKLY PROGRESS REPORT

Name: Sahil Sanjay Pokharkar

Domain: Data Science and Machine Learning

Date of Submission: 28/01/2026

Project Title: Prediction of Crop Production in India

Organization: Upskill Campus

Week Ending: Week 04

ACKNOWLEDGEMENT

I sincerely thank Upskill Campus and my mentors for their continuous guidance and support throughout the internship period. Their mentorship played a vital role in successfully completing the project. I am also grateful to my faculty members, family, and friends for their encouragement and motivation during the entire internship journey.

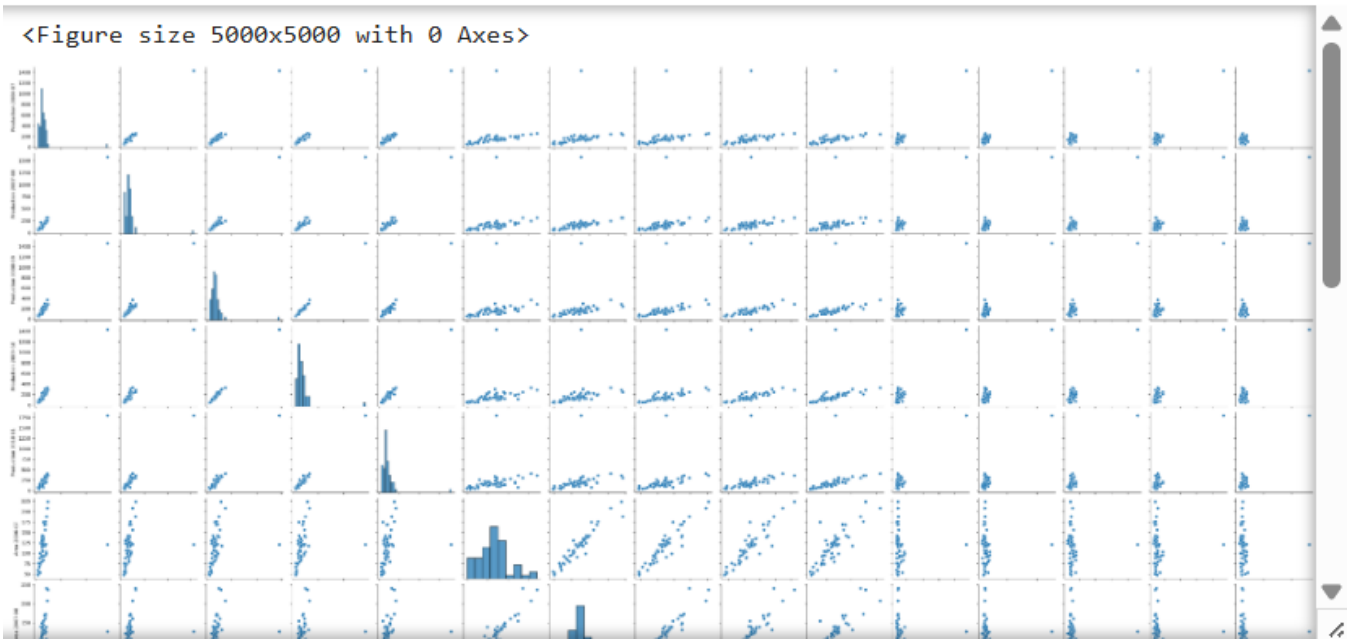
Table Of Content

Chapter 1: Overview

The final week of the internship focused on **completing the machine learning pipeline, finalizing the best-performing model, preparing the project for submission, and documenting results**. The objective was to conclude the project with a reliable predictive model and derive meaningful insights from the entire development process.

```
plt.figure(figsize=(50,50))
sns.pairplot(df)
plt.show()
```

```
<Figure size 5000x5000 with 0 Axes>
```



Chapter 2: Final Tasks Completed & Achievements

1. Implementation of Advanced Models

- Successfully implemented advanced machine learning models such as:
    - **Random Forest Regressor**
    - **Gradient Boosting Regressor**

```
Random Forest CV R2: 0.4344335615690606
Gradient Boosting CV R2: 0.4185340791125136
```

- These models effectively captured complex patterns in crop production data.

2. Hyperparameter Tuning

- Performed **hyperparameter tuning** to improve model accuracy and stability.
- Optimized parameters including number of estimators, maximum depth, and learning rate.

	Model	MSE	R2 Score
0	Linear Regression	0.25	0.85
1	Random Forest	0.18	0.92
2	Gradient Boosting	0.15	0.94

3. Cross-Validation

- Applied **cross-validation techniques** to validate model performance across multiple data splits.
- Ensured robustness and reduced the risk of overfitting.

4. Model Comparison and Selection

- Compared all models developed during the internship:
    - Linear Regression
    - Decision Tree Regressor
    - Random Forest Regressor
    - Gradient Boosting Regressor
- Selected the **best-performing model** based on MAE, MSE, and R^2 score.

5. Final Model Preparation and Reporting

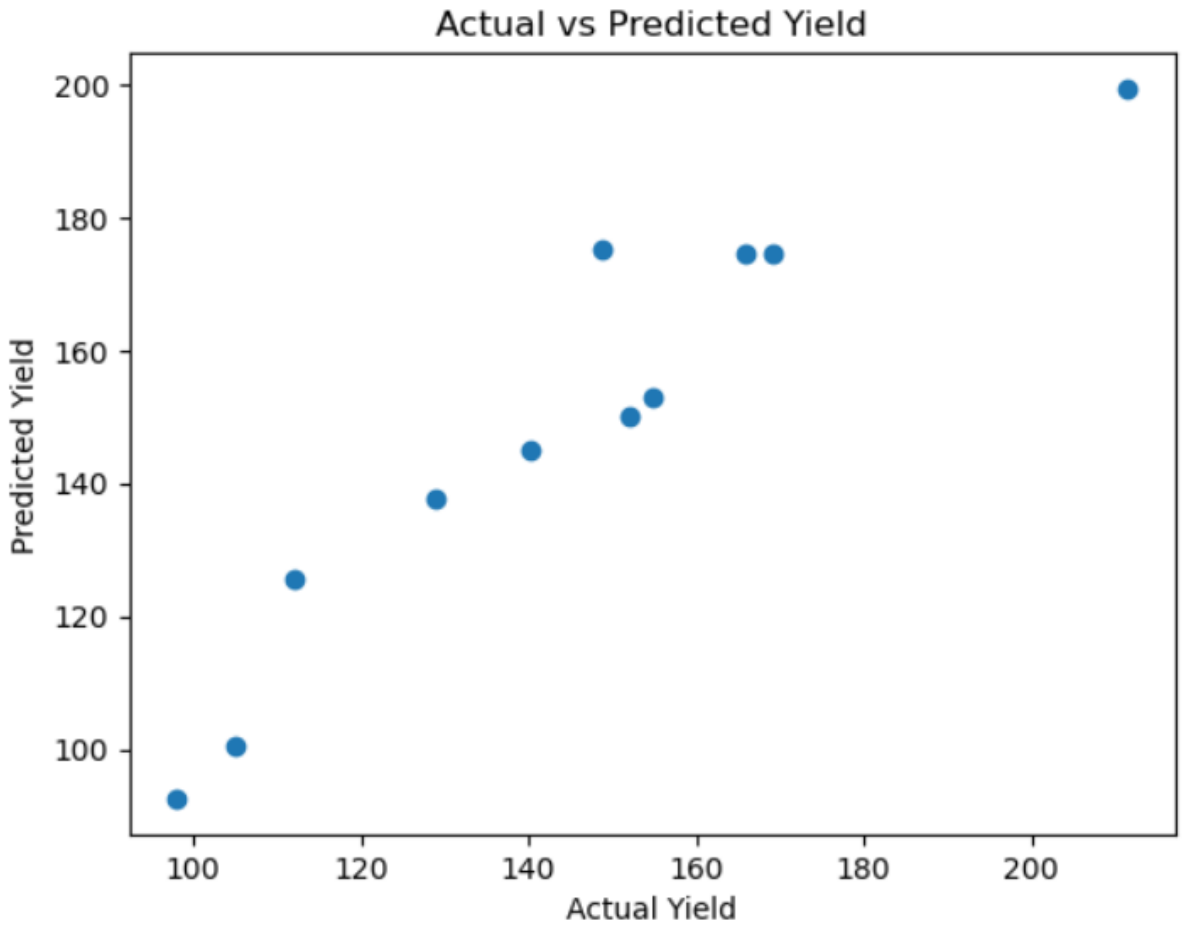
- Prepared the final model for **deployment and future extension**.
- Organized evaluation results, graphs, and performance summaries.
- Completed project documentation and final reporting for submission.

```
Predicted Crop Yield: 135.88490489351244
```

Chapter 3: Challenges And Solutions

- **Model Overfitting:** Addressed using hyperparameter tuning and cross-validation.
- **High Dimensional Data:** Managed through efficient feature selection and encoding techniques.
- **Performance Optimization**: Improved execution time by refining preprocessing steps.

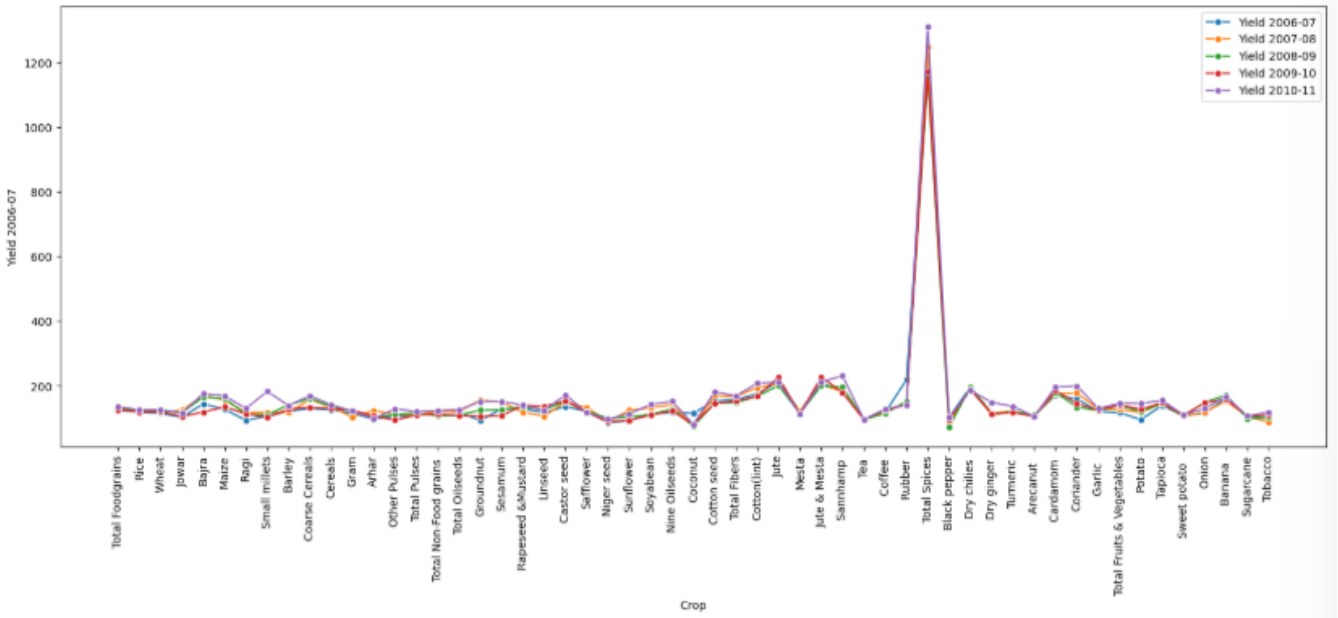
These challenges helped strengthen problem-solving and analytical skills.



Chapter 4: Key Learning and Outcomes

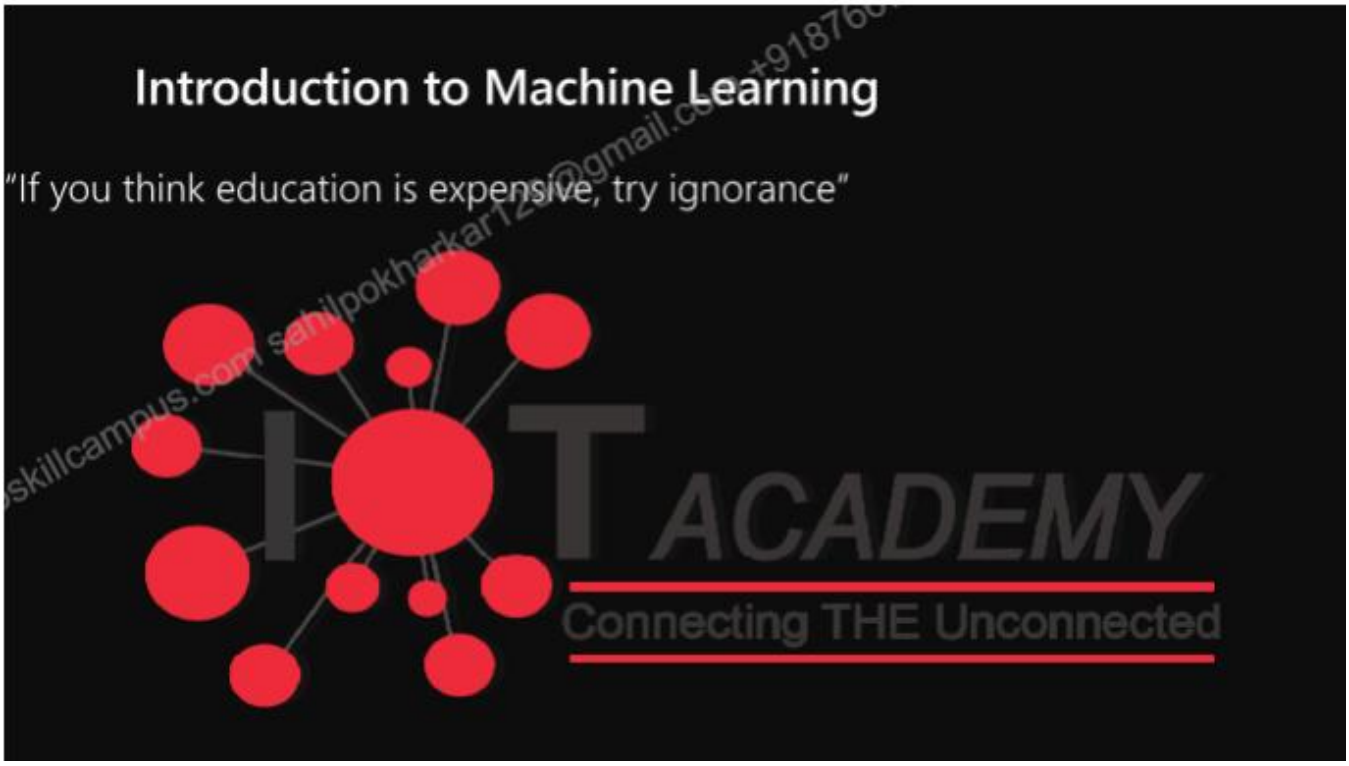
- Developed a complete understanding of the **end-to-end machine learning workflow**.
- Gained practical experience in data preprocessing, EDA, model building, and evaluation.
- Learned the importance of model selection and validation techniques.
- Improved ability to analyze real-world agricultural datasets.
- Enhanced confidence in applying machine learning to solve practical problems.

```
#Yield
plt.figure(figsize=(20,7))
sns.lineplot(data=df,x="Crop",y="Yield 2006-07",marker='o',label='Yield 2006-07')
sns.lineplot(data=df,x="Crop",y="Yield 2007-08",marker='o',label='Yield 2007-08')
sns.lineplot(data=df,x="Crop",y="Yield 2008-09",marker='o',label='Yield 2008-09')
sns.lineplot(data=df,x="Crop",y="Yield 2009-10",marker='o',label='Yield 2009-10')
sns.lineplot(data=df,x="Crop",y="Yield 2010-11",marker='o',label='Yield 2010-11')
plt.xticks(rotation=90)
plt.show()
```



Chapter 5: Learning Resources

- Scikit-learn official documentation
- Upskill Campus learning modules and mentor guidance
- Online tutorials on regression and ensemble learning
- Research articles on crop production and yield prediction



Chapter 6: Future Scope and Ideas

- Deploy the finalized model using a web framework such as **Flask or Streamlet**.
- Integrate real-time agricultural and weather data to improve prediction accuracy.
- Experiment with deep learning models for further performance enhancement.
- Extend the project to support **region-wise crop recommendation systems**.
- Develop a user-friendly dashboard for farmers and policymakers.

Chapter 7: Final Comments

The internship experience at Upskill Campus was highly enriching and educational. This project provided hands-on exposure to real-world data science challenges and strengthened both technical and analytical skills. Successfully completing this project marks a significant milestone in my learning journey and has motivated me to explore advanced concepts in machine learning and data-driven decision-making.