WEEKLY PROGRESS REPORT

Name: Sahil Sanjay Pokharkar

Domain: Data Science and Machine Learning

Date of Submission: 04/01/2026

Project Title: Prediction of Crop Production in India

Organization: Upskill Campus

Week Ending: Week 01

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to **Upskill Campus** for providing me with the opportunity to undertake this internship in the domain of **Data Science and Machine Learning**. This internship has given me valuable exposure to real-world applications of data science, particularly in the area of agricultural analytics.

I am deeply thankful to my mentors and coordinators at Upskill Campus for their continuous guidance, encouragement, and support throughout the internship period. Their valuable suggestions and feedback have played a crucial role in enhancing my technical knowledge and practical understanding of machine learning concepts.

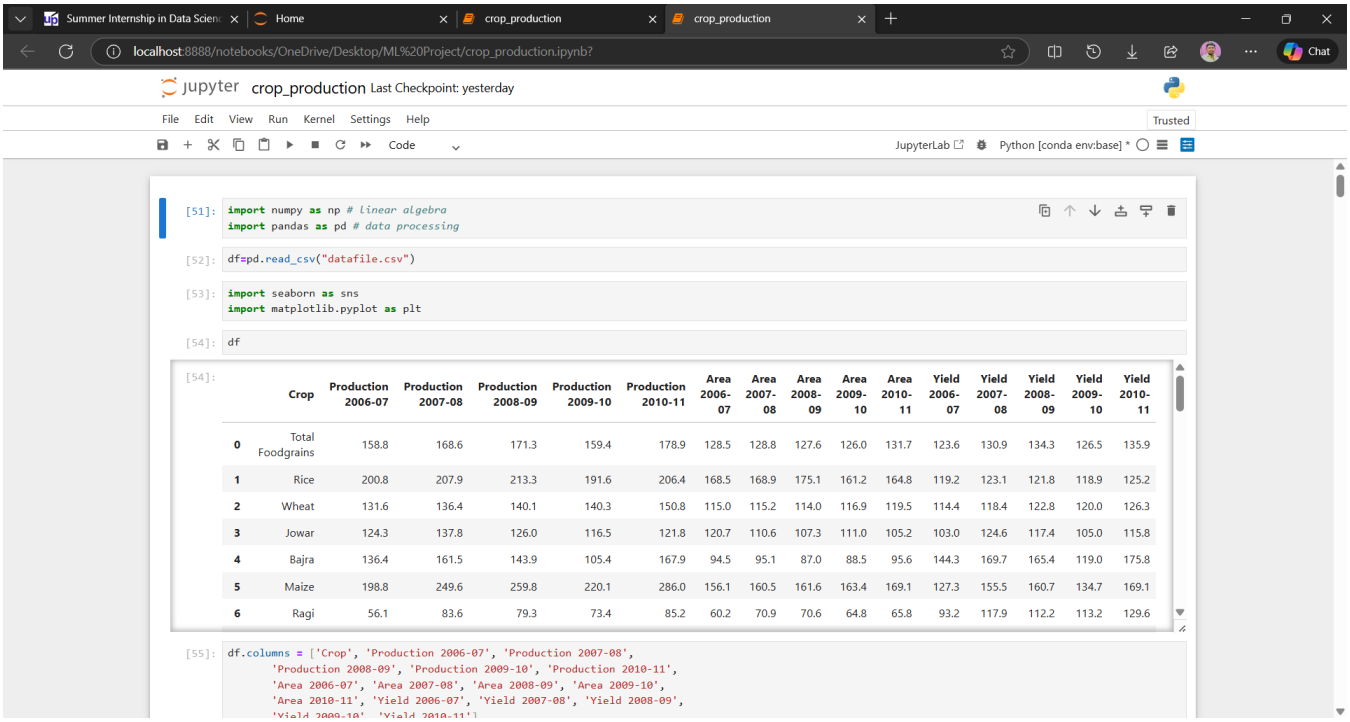
I would also like to extend my heartfelt thanks to my faculty members for their constant motivation and academic support, which helped me successfully apply theoretical concepts to practical project work.

Lastly, I am grateful to my family and friends for their encouragement and support, which motivated me to stay focused and committed to completing this internship successfully.

Table Of Content

Chapter 1: Overview

During this week, the primary focus was on gaining a comprehensive understanding of the project problem statement related to crop production prediction in India. The week involved studying the agricultural domain, identifying key factors that influence crop yield, and exploring how data science and machine learning techniques can be effectively applied to agricultural datasets. Initial exposure to the dataset helped in understanding its structure, attributes, and overall scope for predictive modelling.



Summer Internship in Data Scien | Home | crop_production | crop_production

localhost:8888/notebooks/OneDrive/Desktop/ML%20Project/crop_production.ipynb?

jupyter crop_production Last Checkpoint: yesterday

File Edit View Run Kernel Settings Help

Trusted

Code

JupyterLab

Python [conda env:base]

[51]:
```
import numpy as np # linear algebra
import pandas as pd # data processing
```

[52]:
```
df=pd.read_csv("datafile.csv")
```

[53]:
```
import seaborn as sns
import matplotlib.pyplot as plt
```

[54]:
```
df
```

[54]:

	Crop	Production 2006-07	Production 2007-08	Production 2008-09	Production 2009-10	Production 2010-11	Area 2006-07	Area 2007-08	Area 2008-09	Area 2009-10	Area 2010-11	Yield 2006-07	Yield 2007-08	Yield 2008-09	Yield 2009-10	Yield 2010-11
0	Total Foodgrains	158.8	168.6	171.3	159.4	178.9	128.5	128.8	127.6	126.0	131.7	123.6	130.9	134.3	126.5	135.9
1	Rice	200.8	207.9	213.3	191.6	206.4	168.5	168.9	175.1	161.2	164.8	119.2	123.1	121.8	118.9	125.2
2	Wheat	131.6	136.4	140.1	140.3	150.8	115.0	115.2	114.0	116.9	119.5	114.4	118.4	122.8	120.0	126.3
3	Jowar	124.3	137.8	126.0	116.5	121.8	120.7	110.6	107.3	111.0	105.2	103.0	124.6	117.4	105.0	115.8
4	Bajra	136.4	161.5	143.9	105.4	167.9	94.5	95.1	87.0	88.5	95.6	144.3	169.7	165.4	119.0	175.8
5	Maize	198.8	249.6	259.8	220.1	286.0	156.1	160.5	161.6	163.4	169.1	127.3	155.5	160.7	134.7	169.1
6	Ragi	56.1	83.6	79.3	73.4	85.2	60.2	70.9	70.6	64.8	65.8	93.2	117.9	112.2	113.2	129.6

[55]:
```
df.columns = ['Crop', 'Production 2006-07', 'Production 2007-08',
       'Production 2008-09', 'Production 2009-10', 'Production 2010-11',
       'Area 2006-07', 'Area 2007-08', 'Area 2008-09', 'Area 2009-10',
       'Area 2010-11', 'Yield 2006-07', 'Yield 2007-08', 'Yield 2008-09',
       'Yield 2009-10', 'Yield 2010-11']
```

Chapter 2: Achievements

3. Problem Understanding and Domain Research

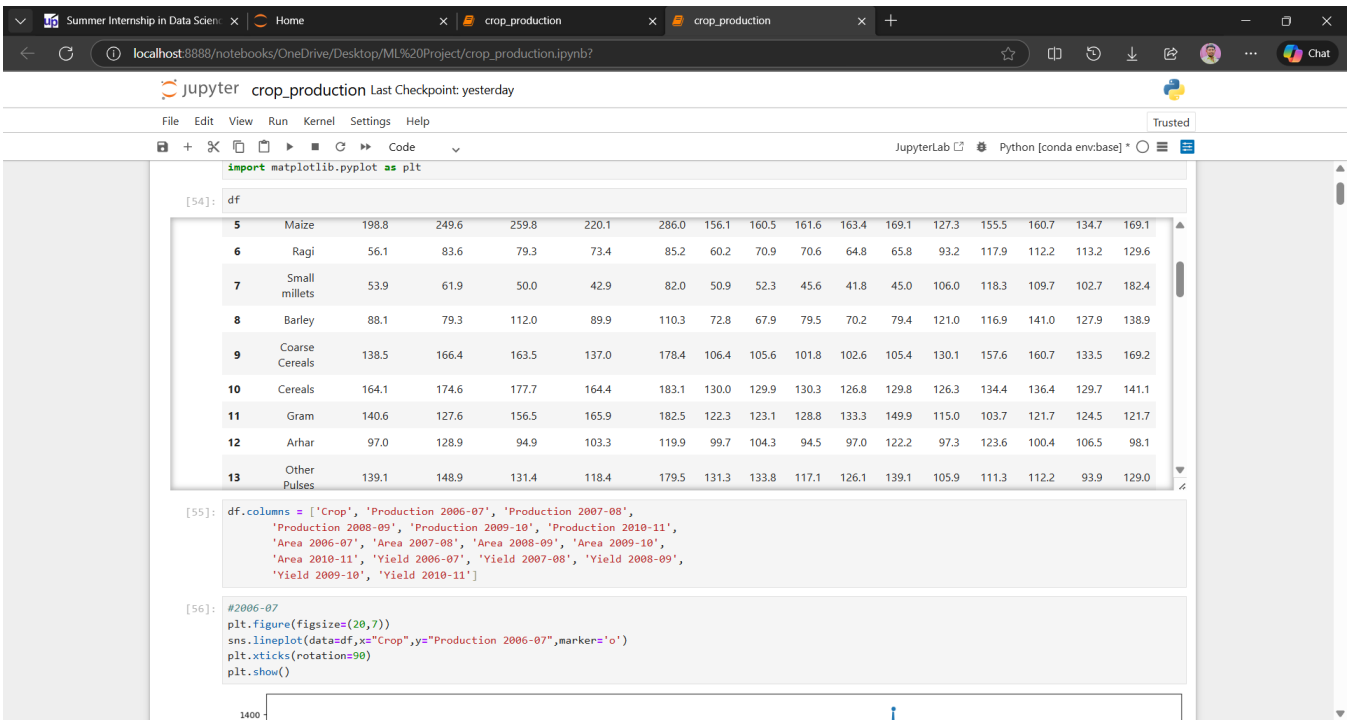
- Conducted detailed research on Indian agriculture, crop seasons (Kharif, Rabi, Zaid), and production cycles.
- Identified major factors affecting crop production such as geographical location, seasonal variations, rainfall, and cultivated area.
- Understood the importance of crop production prediction for farmers, government bodies, and policy planners.

2. Dataset Collection and Initial Exploration

- Collected the 'Crop Production in India' dataset from a reliable and structured data source.
- Reviewed dataset attributes including state, district, crop, season, area, and production values.
- Performed initial inspection to analyse dataset size, column types, and basic statistics.

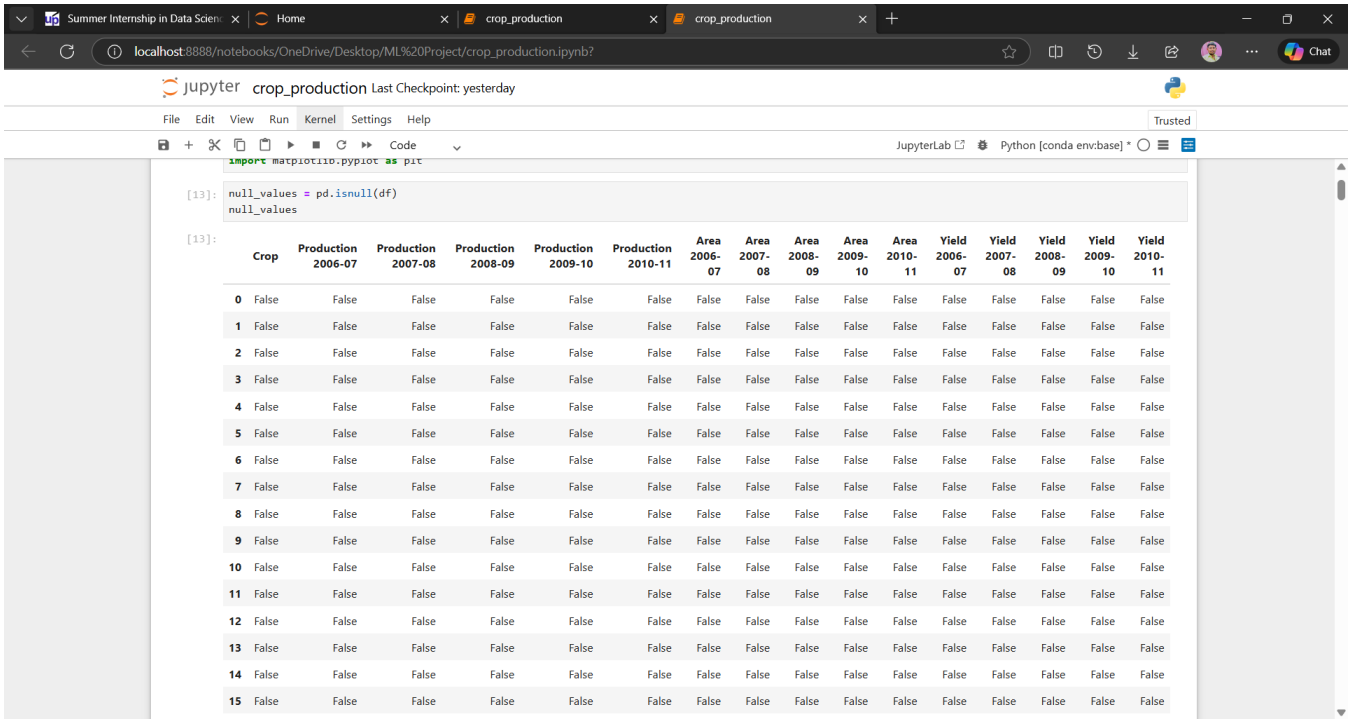
3. Python and Data Science Skill Development

- Revised Python fundamentals essential for data analysis and machine learning.
- Practiced using libraries such as NumPy and Pandas for data handling and manipulation.
- Explored Matplotlib and Seaborn for basic data visualization purposes.



Chapter 3: Challenges

- Presence of missing, null, and inconsistent values within the dataset.
- Difficulty in interpreting certain categorical agricultural values related to crops and seasons.
- Limited initial understanding of how agricultural factors translate into numerical data features.



	Crop	Production 2006-07	Production 2007-08	Production 2008-09	Production 2009-10	Production 2010-11	Area 2006-07	Area 2007-08	Area 2008-09	Area 2009-10	Area 2010-11	Yield 2006-07	Yield 2007-08	Yield 2008-09	Yield 2009-10	Yield 2010-11
0	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
5	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
6	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
7	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
8	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
9	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
10	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
11	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
12	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
13	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
14	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
15	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False

Chapter 4: Learning Resources

- Online tutorials and documentation on Exploratory Data Analysis (EDA) using Python.
- Case studies and sample projects related to regression-based crop yield prediction.
- Video tutorials and practice exercises focused on Pandas, NumPy, and data visualization.
- UpSkill Campus YouTube video tutorials and internship program lectures.

Chapter 5: Next Week's Goals

- Perform data cleaning by handling missing values and correcting inconsistencies.
- Conduct detailed Exploratory Data Analysis (EDA) to identify patterns and trends.
- Visualize crop production with respect to state, crop type, and season.
- Select suitable machine learning algorithms such as Linear Regression and Random Forest.
- Begin implementation of initial machine learning models using Jupyter Notebook.

Chapter 6: Additional Comments

This week played a crucial role in building a strong foundation for the project. The understanding gained regarding the dataset structure, agricultural domain, and data science workflow will significantly contribute to the successful execution of the project in the upcoming weeks. Continuous learning and hands-on practice will remain a key focus moving forward.