WEEKLY PROGRESS REPORT

Name: Purva Jagdish Salgaonkar

Domain: Data Science & Machine Learning Internship

Date of submission: 10 Dec 2025

WEEK ENDING: 01

I. OVERVIEW:

During Week-1 of the Data Science &Machine Learning Internship, I focused on building strong foundational knowledge in Data Science, Machine Learning, and real-world ML project development. I explored two major ML projects from the repository (Crop Production Prediction & Smart City Traffic Forecasting) and understood the complete workflow including data preprocessing, model building, evaluation, and visualization. Along with this, I studied learning materials, videos, and an e-book to strengthen theoretical understanding.

II. ACHIEVEMENTS:

a. Introduction to Data Science – Video Learning

I completed the video *"What is Data Science | Introduction to Data Science"* which provided a clear understanding of:

- What big data is and how huge amounts of information are created every second
- Types of data: text, images, videos, personal information
- How industries like agriculture, medicine, business, and technology use data
- Importance of Data Science skills for career growth
- Programming tools such as Python, C, algorithms, and visualization techniques
- How data helps in making predictions, solving problems, and understanding trends

Video:

The speaker explained how massive data is generated every moment and how data science helps analyze this information to solve real-world problems. They highlighted career opportunities, tools like Python, and real applications like

prediction models and classification systems. The video emphasizes that learning data science enables better decisions, career improvement, and understanding global processes.

b. Introduction to Machine Learning – Video Learning

The second video provided a simple explanation of how machine learning works and how machines learn from data.ML models learn patterns from examples. They improve automatically as more data is given Real-world examples:

- Search engines predicting queries
- YouTube & websites recommending videos/products
- Camera features & language translation
- Apps adapting based on user data

Video:

The video explained that machine learning allows computers to learn from data and produce smart outputs without explicit programming. It used examples like YouTube recommendations, search predictions, and translation apps. It also showed how ML takes data as input, builds a model, and generates better results over time.

c. Machine Learning Internship Projects

I explored the ML projects provided by Upskill Cloud Technologies, including:

- Crop Production Prediction in India (2001–2014)
- Smart City Traffic Forecasting

Both projects helped me understand:

- How to clean datasets
- How to handle missing data
- How to use Linear Regression for prediction
- How to extract features and prepare data for ML
- How to evaluate accuracy
- How visualizations help compare actual vs predicted values

These projects helped me become familiar with Google Colab, Pandas, NumPy, Scikit-learn, and Matplotlib.

GitHub Repository: https://github.com/Purva0210/UpSkill

d. Practical Understanding

I gained a practical understanding of how ML models learn, the steps of data collection, preprocessing, and prediction, and the importance of exploring and analyzing real-world datasets, while also learning how an end-to-end ML pipeline is structured, implemented, and applied in real-world projects to solve meaningful problems.

III. CHALLENGES:

a. Dataset Handling

Government agricultural datasets had inconsistent column names, missing values in key fields, and varying numerical ranges. I addressed these by standardizing columns and applying appropriate missing-value strategies.

b. ZIP File Structure

The dataset had nested ZIP files, requiring multi-stage extraction. I implemented Python logic to systematically detect and extract all files, gaining experience in complex filesystem handling.

c. Feature Selection

Dataset structures varied, making it challenging to identify the target variable and relevant features. Manual dropping and encoding ensured the model received only useful, correctly formatted data.

IV. LEARNING RESOURCES:

a. Videos Watched

- Explored Big Data, Data Science & Career Opportunities to understand the field, types of data, and career roles.
- Watched Introduction to Machine Learning to learn how ML models work, the learning process, and the end-to-end ML pipeline.
- These videos provided both theoretical knowledge and practical insights for real-world applications.

b. Technical Learning

- Used Pandas, NumPy, and Scikit-learn documentation to understand coding tasks and implement them effectively.
- Practiced model-building on Google Colab, including data preprocessing, feature selection, and model evaluation.
- Referred to project file guidelines for task clarity and proper workflow structuring.
- These resources strengthened both conceptual understanding and hands-on ML skills.

c. UCT Documentation

- Studied UCT documentation to understand standardized approaches for ML projects.

- Applied the guidelines to create and upload two ML project reports to my repository:
    - Crop Production Prediction in India (2001–2014)
    - Smart City Traffic Forecasting

- This enhanced my ability to follow structured workflows and produce well-documented ML projects.

V. NEXT WEEK'S GOALS

a. e-Book Study

During this week, I will focus on studying the e-book "Introducing Data Science & Machine Learning" to gain a comprehensive understanding of the fundamentals. I aim to grasp the basics of data science, learn how machine learning pipelines are structured, and explore different types of ML models along with their practical use cases. I will also understand the importance of working with both structured and unstructured data and examine real-life applications of data science and machine learning across various industries to connect theory with practical scenarios.

b. Quiz Preparation

I will prepare for "Test Your Knowledge DS: Quiz-1" by revising key foundational concepts. This includes understanding the definitions of data science, learning about different types of data, and reviewing the benefits of machine learning. I will also focus on familiarizing myself with essential Python libraries commonly used in data science and explore various machine learning applications to ensure a strong conceptual understanding before progressing to more advanced topics.

c. Weekly Report

This week, I will document and submit the Week-2 progress report, summarizing all the activities I undertook, the concepts I learned, and the tools I explored. The report will provide a clear overview of my progress, consolidate my understanding of the topics covered, and serve as a reference for upcoming tasks and learning objectives.

VI. ADDITIONAL COMMENTS

This week helped me build strong foundations in Data Science and Machine Learning. I now feel confident in understanding ML pipelines and ready to take on more challenging tasks, including writing and improving ML code. The internship structure, weekly reports, and projects are helping me grow steadily toward becoming industry-ready.