WEEKLY PROGRESS REPORT

Name: Purva Jagdish Salgaonkar

Domain: Data Science & Machine Learning Internship

Date of submission: 17 Dec 2025

WEEK ENDING: 02

I. OVERVIEW:

During Week-2 of the Data Science & Machine Learning Internship, I focused on strengthening my theoretical foundation in Data Science and Machine Learning. The primary activity for this week was studying the prescribed e-book "Introducing Data Science: Big Data, Machine Learning, and More, Using Python Tools" by Manning Publications. Along with the reading, I attempted the weekly quiz to assess my understanding of the concepts covered. This week helped me clearly understand the data science workflow, types of data, machine learning basics, and the role of big data technologies.

II. ACHIEVEMENTS:

a. E-Book Study: Introducing Data Science & Machine Learning

I thoroughly studied the assigned e-book chapters related to the fundamentals of data science and machine learning. Through this reading, I gained clear insights into:

- What data science is and how it differs from traditional data analysis
- The relationship between data science and big data
- The characteristics of big data - Volume, Variety, Velocity, and Veracity
- Different types of data such as structured, unstructured, machine-generated, graph-based, streaming, text, image, and video data
- Real-world applications of data science in industries like finance, healthcare, government, business, and education
- An overview of the data science ecosystem including Hadoop, distributed systems, machine learning frameworks, and NoSQL databases

The book also explained how data science projects are carried out in practice and how theory connects to real-world implementation using Python tools.
Understanding the Data Science Process
From the e-book, I learned the six-step data science process in detail:

- Setting the research goal
- Retrieving data
- Data preparation
- Data exploration
- Data modeling
- Presentation and automation

This structured approach helped me understand how data science projects move from problem definition to final results and deployment. The explanation of iterative learning and real-world case studies made the process easy to understand and practical.

b. Quiz Attempt and Performance

After completing the reading, I attempted the Week-2 quiz titled "Test Your Knowledge – Data Science". The quiz consisted of multiple-choice questions based on machine learning concepts, statistical basics, feature engineering, PCA, model evaluation, and Python fundamentals.

- Score Achieved: 38 / 40
- Percentage: 95%

This result reflects my strong understanding of the theoretical concepts covered in the e-book and confirms that I was able to apply my learning effectively. Key Concepts Learned Through the Quiz. By preparing for and attempting the quiz, I reinforced my understanding of:

- Evaluation metrics used to compare machine learning models
- Importance of p-values in statistical hypothesis testing
- Feature selection and feature engineering concepts
- Dimensionality reduction techniques such as PCA
- Overfitting prevention techniques
- Differences between fit() and fit_transform() in Scikit-learn
- Ensemble learning basics such as bagging
- Python list operations and Pandas DataFrame functions
- Visualization techniques for detecting outliers

c. Practical Understanding

Although this week focused mainly on theory, it significantly improved my conceptual clarity. I now have a better understanding of why specific techniques are used in machine learning, how models are evaluated, and how data preparation impacts model performance. This theoretical base will strongly

support upcoming practical tasks related to probability, statistics, and model building.

III. CHALLENGES:

a. Conceptual Understanding
Some theoretical concepts such as p-values, Principal Component Analysis (PCA), and model evaluation metrics required careful reading and revision. Understanding their purpose and significance in machine learning models took additional effort.

b. Theory to Practical Mapping
Connecting the theoretical explanations from the e-book with real-world machine learning scenarios was initially challenging. Extra revision was needed to understand how these concepts influence model performance and decision-making.

c. Statistical Interpretation
Interpreting statistical results and understanding how statistical concepts support machine learning algorithms required focused study. Revising quiz-based questions helped strengthen clarity and confidence.

IV. LEARNING RESOURCES:

a. E-Book Study
- Studied the e-book "*Introducing Data Science: Big Data, Machine Learning, and More*" to understand the fundamentals of data science and machine learning.

- Learned about data types, data science workflows, big data concepts, feature engineering, dimensionality reduction, and model evaluation.

- The e-book provided strong theoretical knowledge supported by real-world examples.

b. Quiz-Based Learning
- Attempted the weekly quiz to assess understanding of data science and machine learning concepts.

- Reviewed quiz questions to identify weak areas and revise important topics such as PCA, feature selection, overfitting, and statistical testing.

- This self-assessment approach helped reinforce conceptual clarity.

c. Supporting Learning Materials

- This self-assessment approach helped reinforce conceptual clarity.

- Used previous project notes to relate theoretical concepts to practical machine learning workflows.

- These resources helped strengthen both conceptual understanding and readiness for upcoming topics.

V. NEXT WEEK'S GOALS

a. Introduction to Probability and Statistics

In the upcoming week, I will focus on learning probability and statistics, which form the mathematical foundation of data science and machine learning. This includes understanding probability concepts, probability distributions, measures of central tendency such as mean and median, measures of spread like variance and standard deviation, correlation analysis, and the basics of hypothesis testing.

b. Practical Application

Along with theoretical study, I aim to apply statistical concepts to real dataset**s** using Python libraries such as NumPy, Pandas, and Matplotlib. This will help me visualize data distributions, analyze relationships between variables, and strengthen my analytical and problem-solving skills.

c. Weekly Report

I will continue maintaining detailed documentation of my learning activities and submit the Week-3 progress report, clearly summarizing my understanding of probability and statistics concepts and how they relate to machine learning models.

VI. ADDITIONAL COMMENTS

Week-2 played a crucial role in building my theoretical foundation in data science and machine learning. Studying the e-book and performing well in the quiz boosted my confidence and prepared me for more advanced topics such as probability, statistics, and model optimization in the coming weeks.