

# TR-102 Daily Diary (Machine Learning)

## WEEK 1 (19 June – 20 June 2025)

### Day 1 – Thursday, 19 June 2025

**Topic:** Introduction to Artificial Intelligence and Machine Learning

**Objective:** Understand the fundamentals of AI, its applications, and training goals.

**Summary:**

The training began with an overview of Artificial Intelligence and its major domains — Machine Learning, Deep Learning, and Data Science. The instructor explained real-world applications such as recommendation systems, autonomous cars, and voice assistants. We discussed the difference between supervised, unsupervised, and reinforcement learning. The trainer also gave an introduction to the company's development environment and the tools we'll be using in this program.

---

### Day 2 – Friday, 20 June 2025

**Topic:** Setting Up Python Environment and Basic Syntax

**Objective:** Install and configure Python, Jupyter Notebook, and practice syntax basics.

**Summary:**

Today's session covered the installation of Python, Anaconda, and Jupyter Notebook for hands-on work. I learned about Python's print statements, indentation rules, and variable declarations. The instructor explained how Python's dynamic typing makes it ideal for ML tasks. We also executed simple programs like addition, string manipulation, and basic input/output operations.

```
name = input("Enter your name: ")
```

```
print("Welcome to AI Training", name)
```

---

## WEEK 2 (23 June – 27 June 2025)

### Day 3 – Monday, 23 June 2025

**Topic:** Python Data Types, Operators, and Expressions

**Objective:** Understand Python's data types and perform basic operations.

**Summary:**

We explored Python's built-in data types: integers, floats, strings, lists, tuples, and dictionaries. The trainer explained mutable vs immutable data structures and how to apply arithmetic, logical, and comparison operators. This helped me understand how to structure and process data in ML workflows.

---

## Day 4 – Tuesday, 24 June 2025

**Topic:** Conditional Statements and Loops

**Objective:** Learn control flow mechanisms in Python.

**Summary:**

I learned how to implement decision-making statements like if, elif, and else, and looping constructs like for and while. The trainer explained how loops are used in data iteration and preprocessing tasks. I practiced writing programs using nested loops and conditionals.

```
for i in range(5):
```

```
    if i % 2 == 0:
```

```
        print(i, "is even")
```

---

## Day 5 – Wednesday, 25 June 2025

**Topic:** Functions and Modular Programming

**Objective:** Create reusable code structures using user-defined functions.

**Summary:**

Today's session focused on writing functions in Python for code reusability. We discussed parameter passing, scope of variables, and the use of return statements. The concept of modular programming was introduced, allowing functions to be reused across different files.

---

## Day 6 – Thursday, 26 June 2025

**Topic:** Lists, Tuples, and Dictionaries in Depth

**Objective:** Understand complex data structures and their practical uses.

**Summary:**

I studied list indexing, slicing, and dictionary key-value manipulation. The trainer explained how these structures are used to handle datasets in ML pipelines. I practiced sorting, appending, and iterating through lists and dictionaries.

---

## Day 7 – Friday, 27 June 2025

**Topic:** Introduction to Python Libraries (NumPy, Pandas, Matplotlib)

**Objective:** Learn the purpose and usage of essential Python libraries.

**Summary:**

Today, we explored the three most important Python libraries: NumPy for numerical operations, Pandas for data analysis, and Matplotlib for data visualization. I practiced importing libraries and performing basic computations using NumPy arrays and Pandas dataframes.

---

## **WEEK 3 (30 June – 4 July 2025)**

### **Day 8 – Monday, 30 June 2025**

**Topic:** Data Collection and Loading with Pandas

**Objective:** Learn to import and preview datasets for ML tasks.

**Summary:**

We learned how to load datasets from CSV and Excel files using Pandas functions like `read_csv()` and `read_excel()`. The trainer explained how to check data shape, column names, and data types using Pandas methods. I imported an email dataset to use in my project later.

---

### **Day 9 – Tuesday, 1 July 2025**

**Topic:** Data Cleaning and Handling Missing Values

**Objective:** Clean raw data by handling nulls and duplicates.

**Summary:**

Today's session focused on data preprocessing steps. We removed duplicate records, handled missing values using `fillna()`, and replaced unwanted data. I learned that data cleaning is one of the most time-consuming yet critical stages in any ML project.

---

### **Day 10 – Wednesday, 2 July 2025**

**Topic:** Feature Engineering and Data Transformation

**Objective:** Prepare data for better model performance.

**Summary:**

We studied feature engineering techniques such as encoding categorical variables and normalization. The trainer explained how scaling affects model accuracy. I used Scikit-learn's preprocessing tools to normalize numerical features.

---

### **Day 11 – Thursday, 3 July 2025**

**Topic:** Data Visualization with Matplotlib

**Objective:** Visualize patterns in data.

**Summary:**

I learned how to use Matplotlib to plot bar charts, line graphs, and histograms. Visualizations helped me understand how different attributes are distributed. We also created simple spam-vs-ham graphs using sample email data.

---

### **Day 12 – Friday, 4 July 2025**

**Topic:** Correlation and Feature Selection

**Objective:** Analyze relationships between features using correlation matrices.

**Summary:**

The session explained how correlation helps in identifying strong predictors. I learned to use Seaborn heatmaps for visual representation. This knowledge is vital for selecting features in the spam detection model.

---

**WEEK 4 (7 July – 11 July 2025)****Day 13 – Monday, 7 July 2025**

**Topic:** Introduction to Machine Learning and Its Workflow

**Objective:** Learn ML pipeline – data, model, training, and testing.

**Summary:**

We discussed supervised and unsupervised learning concepts, dataset splitting, and model evaluation. The instructor provided an overview of Scikit-learn and how it simplifies ML implementation.

---

**Day 14 – Tuesday, 8 July 2025**

**Topic:** Linear Regression Model

**Objective:** Understand regression techniques.

**Summary:**

I learned how to apply Linear Regression to predict continuous outcomes. Using sample data, I trained a model, fitted a regression line, and visualized predictions.

---

**Day 15 – Wednesday, 9 July 2025**

**Topic:** Logistic Regression for Binary Classification

**Objective:** Apply classification algorithm for spam detection.

**Summary:**

We explored Logistic Regression to classify data into two classes. I applied the model to email text features to predict spam or ham. It was my first step in building the Email Spam Checker.

---

**Day 16 – Thursday, 10 July 2025**

**Topic:** Model Evaluation Metrics

**Objective:** Evaluate ML model performance.

**Summary:**

I studied accuracy, precision, recall, and F1-score. The trainer explained confusion matrix interpretation and how to evaluate classifier reliability.

---

## **Day 17 – Friday, 11 July 2025**

**Topic:** K-Nearest Neighbors (k-NN) Algorithm

**Objective:** Learn and implement KNN classifier.

**Summary:**

I learned how KNN works on the principle of proximity and similarity. We implemented the algorithm using the Scikit-learn library and compared its accuracy with Logistic Regression.

---

## **WEEK 5 (14 July – 18 July 2025)**

### **Day 18 – Monday, 14 July 2025**

**Topic:** Introduction to Natural Language Processing (NLP)

**Objective:** Understand text processing fundamentals.

**Summary:**

The trainer introduced NLP concepts and explained tokenization, stopword removal, and word frequency analysis. These are essential for the email spam checker dataset.

---

### **Day 19 – Tuesday, 15 July 2025**

**Topic:** Text Preprocessing Using NLTK

**Objective:** Apply tokenization, stemming, and lemmatization.

**Summary:**

I learned how to process raw text using the NLTK library. We applied stemming and stopword removal to clean email content.

---

### **Day 20 – Wednesday, 16 July 2025**

**Topic:** TF-IDF Vectorization

**Objective:** Convert text data to numerical features.

**Summary:**

We studied TF-IDF (Term Frequency–Inverse Document Frequency) for feature extraction. It helped in assigning importance to unique words in spam emails.

---

### **Day 21 – Thursday, 17 July 2025**

**Topic:** Naïve Bayes Algorithm

**Objective:** Implement and test Naïve Bayes classifier.

**Summary:**

I implemented the Multinomial Naïve Bayes model to classify emails. It achieved high accuracy and became the core of my final project.

---

## **Day 22 – Friday, 18 July 2025**

**Topic:** Model Testing and Accuracy Evaluation

**Objective:** Validate model predictions.

**Summary:**

We tested our model using test datasets and evaluated accuracy using confusion matrices. The trainer guided us on fine-tuning hyperparameters to reduce false positives.

---

## **WEEK 6 (21 July – 25 July 2025)**

### **Day 23 – Monday, 21 July 2025**

**Topic:** Model Training and Implementation

**Objective:** Train machine learning models using student performance dataset.

**Summary:**

We trained multiple machine learning models including Linear Regression, Random Forest, and Decision Tree on the student dataset. The goal was to predict academic performance based on study hours, attendance, sleep patterns, and tutoring sessions. I compared model accuracies using  $R^2$  score and selected Random Forest for its better results.

---

### **Day 24 – Tuesday, 22 July 2025**

**Topic:** Model Evaluation and Error Analysis

**Objective:** Assess model accuracy using performance metrics.

**Summary:**

I evaluated the trained models using metrics like Mean Absolute Error (MAE) and  $R^2$  score. Visualization of prediction errors using Matplotlib helped identify outliers. The Random Forest model showed consistent predictions and minimal error rate, confirming its robustness for student performance prediction.

---

### **Day 25 – Wednesday, 23 July 2025**

**Topic:** Streamlit Interface Development

**Objective:** Build a simple web interface for the prediction system.

**Summary:**

Today, I developed an interactive interface using Streamlit that allows users to input student data and view predicted performance instantly. The interface displays prediction results along with confidence levels, making the system easy to use and visually informative.

---

**Day 26 – Thursday, 24 July 2025**

**Topic:** Model Optimization and Feature Tuning

**Objective:** Enhance model performance through parameter optimization.

**Summary:**

We tuned hyperparameters for Random Forest and Linear Regression models using GridSearchCV. Feature scaling and normalization further improved accuracy. I also experimented with removing less relevant attributes to increase efficiency.

---

**Day 27 – Friday, 25 July 2025**

**Topic:** Final Model Testing and Validation

**Objective:** Validate final model before deployment.

**Summary:**

The final model was tested on unseen data to ensure generalization. The system successfully predicted student scores with high accuracy. All modules — preprocessing, model training, and prediction — were integrated into a single functional workflow.

---

**WEEK 7 (28 July – 30 July 2025)****Day 28 – Monday, 28 July 2025**

**Topic:** Report Documentation and Visualization

**Objective:** Prepare report and add visual representations.

**Summary:**

I started preparing project documentation, including workflow diagrams, data analysis charts, and screenshots of the Streamlit interface. The visual elements improved clarity and professional presentation of the report

---

**Day 29 – Tuesday, 29 July 2025**

**Topic:** Project Review and Final Testing

**Objective:** Test the system end-to-end and finalize improvements.

**Summary:**

A complete project review was conducted with the trainer. I tested multiple input combinations to ensure consistent performance and corrected minor formatting issues in the interface. The trainer provided feedback for report refinement.

---

### **Day 30 – Wednesday, 30 July 2025**

#### **Topic: Final Presentation and Submission**

**Objective:** Present and conclude the Student Performance Prediction project.

#### **Summary:**

On the final day, I presented my Student Performance Prediction system, demonstrating its data flow, ML algorithms, and interface. The project was well-received for its practical utility in educational analytics and its high accuracy in predicting academic outcomes.