Summer Internship 2025

Machine Learning Internship – Week 1 Tasks

Objective:

Understand the basics of **classification** and **regression** using supervised learning algorithms, and learn how to use machine learning models to make predictions using Python

Tool Required

- Python
- Jupyter Notebook
- Libraries: pandas, numpy, matplotlib, scikit-learn

Key Concepts (Read Before Starting)

- ✓ Features and labels
- ✓ Training vs testing data
- ✓ Accuracy, error, and overfitting
- ✓ Simple ML models: Linear Regression and Decision Tree

Task1: Regression – Predict House Prices

In this task, you'll use the **California Housing dataset** from "sklearn.datasets" to predict **median house prices** based on features like house age, rooms, and location. You'll train a **Linear Regression** model using scikit-learn, then evaluate it using **Mean Squared Error (MSE)** and **R^2 Score** to measure prediction accuracy.

Tip: Use .fit(), .predict(), and train_test_split() from sklearn.

Task 2: Classification – Predict Survival on the Titanic

In this task, you'll work with the Titanic dataset from Kaggle or GitHub to predict passenger survival. You'll start by cleaning the data—handling missing values using .fillna() or .dropna(), and encoding text columns into numbers. Then, you'll train a Decision Tree Classifier using scikit-learn and evaluate its performance using accuracy and a confusion matrix to understand how well the model predicts survival outcomes**.**

Submission Instructions:

A single Jupyter Notebook (.ipynb file) with:

- ✓ Comments explaining each step
- ✓ Visualizations (bar charts, confusion matrix, etc.)
- ✓ Final accuracy/error of both models

1 paragraph summary at the end: What did you learn this week?

Send your Tasks via Gmail (nextgenlearners.official@gmail.com)