Summer Internship 2025

Machine Learning Internship – Week 2 Tasks

Objective:

This week, you will explore **Unsupervised Learning** using clustering algorithms like **K-Means**. You'll analyze patterns in unlabeled data and visualize the results to understand how algorithms group similar data points.

Task1: Customer Segmentation using K-Means Clustering

Instructions:

1. **Dataset:**
   - → Use the **Mall Customers Dataset**
   - → You can download it from Kaggle or use any version containing features like:
     - CustomerID, Age, Annual Income (k$), Spending Score (1–100)
2. **Steps to Follow:**
   - → Load and understand the dataset (check for missing values, data types)
   - → Perform **EDA (Exploratory Data Analysis)**:
     - Visualize data distributions using seaborn/matplotlib
   - → Apply **K-Means Clustering** on relevant features (e.g., Income and Spending Score)
   - → Use **Elbow Method** to determine optimal number of clusters (k)
   - → Visualize the final clusters using a 2D scatter plot

Goal of Task 1: Understand how clustering can be used to group similar data points without labels and practice real-world segmentation tasks commonly used in marketing and analytics.

Summer Internship 2025

Machine Learning Internship – Week 2 Tasks

Task 2: Movie Recommendation system

In this task, you will use **unsupervised learning** to perform **movie clustering** based on user ratings and genres. Your goal is to group similar movies together so they can be recommended to users with matching preferences. Download a movie dataset such as the MovieLens 100k dataset (or a similar one from Kaggle). Use features like average rating, number of ratings, and one-hot encoded genres. After preparing the data, apply **K-Means clustering** or **Hierarchical Clustering** to group movies into meaningful clusters. Visualize the clusters using PCA or a 2D plot and try to interpret what makes each group similar (e.g., all romantic comedies, high-rated thrillers, etc.). This task will help you understand how unsupervised learning can be applied in building **movie recommendation systems**, a widely used application in platforms like Netflix or IMDb

Submission Instructions:

Submit a **Jupyter Notebook (.ipynb)** file with:

- Clear code
- Proper comments and headings
- Graphs and cluster visualizations
- Your brief analysis or conclusion

Naming format: `Week2_Name_Task1_Unsupervised.ipynb`
`Week2_Name_Task2_Unsupervised.ipynb`

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