Machine Learning Internship (4 Weeks)

Week 1: Data Understanding & Visualization

Project Title: *Sales Trends Analysis of a Superstore*

Goal: Learn to clean, explore, and visualize a real-world dataset.
Dataset Suggestion: Superstore Sales Dataset

Tasks:

- Load and clean the dataset (handle missing values, date parsing)
- Perform exploratory analysis (e.g., Which category sells best? Which region earns the most?)
- Visualize insights using bar charts, heatmaps, line graphs
- Present 3–5 key insights with visual evidence

Learning Focus:

✓ Data types, null handling
✓ Grouping, filtering, aggregation
✓ Basic plots & visual storytelling

Week 2: Classification Model (Supervised Learning)

Project Title: *Student Performance Predictor*

Goal: Predict if a student will pass or fail based on their features.
Dataset Suggestion: Student Performance Dataset

Tasks:

- Select target (e.g., Pass/Fail using threshold on scores)
- Train simple models: Logistic Regression, Decision Tree
- Evaluate using accuracy, precision, confusion matrix
- Explain how certain features affect performance

Learning Focus:

✓ Supervised learning
✓ Binary classification
✓ Model evaluation and comparison

Week 3: Clustering Model (Unsupervised Learning)

Project Title: *Customer Segmentation for a Retail Business*

Goal: Segment customers based on behavior and demographics.
Dataset Suggestion: Mall Customers Dataset

Tasks:

- Scale the features (Annual Income, Spending Score, Age)
- Use K-Means to create 3–5 customer groups
- Visualize clusters in 2D using PCA or scatter plots
- Describe types of customers in each group

Learning Focus:

✓ Unsupervised learning
✓ Clustering + evaluation (Elbow method)
✓ Pattern recognition

Week 4: End-to-End Mini ML Project

Project Title: *Loan Approval Prediction System*

Goal: Build a complete ML solution from data cleaning to prediction.
Dataset Suggestion: Loan Prediction Dataset

Tasks:

- Preprocess data (handle nulls, encode categories)
- Train and test classification models
- Save best model with joblib or pickle
- Create a basic interface using Streamlit or CLI to input new data and show predictions

Learning Focus:

✓ End-to-end pipeline
✓ Real-world prediction flow
✓ Model saving and reuse

Optional Bonus Tasks:

- Try GridSearchCV for tuning
- Compare multiple models using a leaderboard
- Try submitting a simple Kaggle notebook
- Reflect weekly in a short report or blog