One-Month Machine Learning Internship Plan

Week 1: Introduction and Fundamentals

Day 1-2: Orientation and Overview

- Introduction to the company, team members, and resources.
- Overview of the internship plan and objectives.
- Setting up the development environment (Python, Jupyter Notebook, essential libraries).
  Link: Python Setup

Day 3-5: Basics of Machine Learning

- Introduction to Machine Learning concepts and types (supervised, unsupervised, reinforcement learning).
  Links: Supervised Machine Learning Advanced Algorithms
  Unsupervised Learning "1st Module"
- Basics of Python for ML (NumPy, Pandas, Matplotlib).
  Links: Matplotlib Pandas NumPy
- Hands-on Project: Project 1

Week 2: Data Preprocessing and Exploration

Day 8-9: Data Collection and Cleaning

- Understanding different data sources and collection methods.
- Techniques for data cleaning and preprocessing (handling missing values, encoding categorical variables).
- Hands-on Project: Project 2

Day 10-12: Exploratory Data Analysis (EDA)

- Introduction to EDA and its importance in ML.
  Link: EDA Course
- Techniques for EDA: Descriptive statistics, data visualization, correlation analysis.
- Hands-on Project: Project 3

Week 3: Model Building and Evaluation

Day 15-17: Supervised Learning

- Introduction to supervised learning algorithms (Linear Regression, Logistic Regression, Decision Trees, etc.).
- Understanding model training, validation, and evaluation metrics.
- Hands-on Project: Project 4

Day 18-19: Advanced Supervised Learning Techniques

- Introduction to advanced algorithms (Random Forests, Gradient Boosting, SVMs).
- Hyperparameter tuning and cross-validation.
- Hands-on Project: Project 5

Week 4: Advanced Topics

Day 22-24: Unsupervised Learning

- Introduction to unsupervised learning algorithms (K-means, Hierarchical Clustering, PCA).
- Applications and use cases of unsupervised learning.
- Hands-on Project: Project 6

Day 25-26: Deep Learning and Neural Networks

- Introduction to neural networks and deep learning concepts.
- Overview of deep learning frameworks (TensorFlow, Keras, PyTorch).
- Hands-on Project: Project 7

Project 1: Data Manipulation and Visualization

- **Objective:** To get hands-on experience with data manipulation and basic visualization techniques.
- **Dataset:** Titanic Dataset
- **Tasks:**
  - Load and inspect the dataset using pandas.
  - Perform basic data manipulation: filtering, sorting, grouping, and aggregating.
  - Create visualizations using matplotlib and seaborn to understand data distributions and relationships.
- **Deliverables:** A Jupyter notebook with data manipulation steps and visualizations, along with brief explanations.

Project 2: Data Preprocessing

- **Objective: To learn data cleaning and preprocessing techniques.**
- **Dataset:** House Prices
- **Tasks:**
  - Handle missing values using various techniques (mean/mode imputation, forward fill, etc.).
  - Encode categorical variables (one-hot encoding, label encoding).
  - Normalize/standardize numerical features.
  - Identify and remove outliers if necessary.

- **Deliverables:** A Jupyter notebook with preprocessing steps, rationale for chosen methods, and a cleaned dataset.

Project 3: Exploratory Data Analysis (EDA)

- **Objective: To perform a thorough exploratory data analysis.**
- **Dataset:** House Prices
- **Tasks:**
  - Calculate and interpret summary statistics.
  - Create visualizations to explore relationships between variables (scatter plots, box plots, histograms, heatmaps).
  - Identify and analyze key patterns and insights from the data.

- **Deliverables:** A Jupyter notebook with EDA steps, visualizations, and a summary of findings.

Project 4: Supervised Learning (Regression)

- **Objective: To build and evaluate a regression model.**
- **Dataset:** House Prices
- **Tasks:**
  - Split the dataset into training and testing sets.
  - Train a linear regression model and evaluate its performance using metrics like RMSE, MAE, and R^2.
  - Experiment with other regression algorithms (Ridge, Lasso, Decision Tree Regressor).
  - Perform hyperparameter tuning to improve model performance.

- **Deliverables:** A Jupyter notebook with model training steps, evaluation metrics, and comparison of different models.

Project 5: Supervised Learning (Classification)

- **Objective: To build and evaluate a classification model.**
- **Dataset:** Titanic Dataset
- **Tasks:**
  - Split the dataset into training and testing sets.
  - Train a logistic regression model and evaluate its performance using metrics like accuracy, precision, recall, and F1-score.
  - Experiment with other classification algorithms (Random Forest, SVM, KNN).
  - Perform hyperparameter tuning to improve model performance.

- **Deliverables:** A Jupyter notebook with model training steps, evaluation metrics, and comparison of different models.

Project 6: Unsupervised Learning (Clustering)

- **Objective: To implement and analyze an unsupervised learning algorithm.**
- **Dataset:** Customer Segmentation
- **Tasks:**
    - Preprocess the dataset (normalize features if necessary).
    - Implement K-means clustering and determine the optimal number of clusters using the elbow method.
    - Visualize clusters and interpret the results.
    - Experiment with hierarchical clustering and compare results.

- **Deliverables:** A Jupyter notebook with clustering steps, visualizations, and interpretations.

Project 7: Deep Learning (Neural Network)

- **Objective: To build and evaluate a simple neural network for a classification task.**
- **Dataset:** MNIST
- **Tasks:**
    - Preprocess the dataset (normalize pixel values, one-hot encode labels).
    - Build a neural network using TensorFlow/Keras.
    - Train the network and evaluate its performance on the test set.
    - Experiment with different network architectures and hyperparameters.

- **Deliverables:** A Jupyter notebook with model-building steps, training process, evaluation metrics, and comparison of different architectures.