



AI/ML Internship Report

Submitted To :

DevelopersHub Corporation

Submitted By :

Kashaf Zahra

Due Date :

26th June, 2025

Task 1: Exploring and Visualizing the Iris Dataset

Objective:

To explore and visualize the Iris dataset using basic data analysis and plotting techniques. The goal was to understand the structure, features, and separability of flower species.

Dataset:

- Iris dataset (built-in with Seaborn library)
- Contains 150 samples of 3 flower species with 4 features: sepal_length, sepal_width, petal_length, petal_width

Tools and Methods:

- Python (pandas, matplotlib, seaborn)
- Jupyter Notebook (VS Code)
- Scatter plots, histograms, box plots, summary statistics

Key Findings:

- Petal measurements are more effective in separating flower species than sepal features.
 - No missing data was found.
 - Data is balanced across all species.
 - A basic Random Forest model achieved over **90% accuracy**, showing strong feature separability.
-

Task 2: Predicting Future Stock Prices (Short-Term)

Objective:

To predict the next day's closing stock price for Apple Inc. (AAPL) using historical data fetched from Yahoo Finance.

Dataset:

- Apple stock data from 2020 to 2024
- Collected using the yfinance library
- Features used: Open, High, Low, Volume
- Target: Next day's Close price (using .shift(-1))

Models Applied:

- Linear Regression (main)
- Random Forest Regressor (tested for comparison)

Results:

- **Linear Regression** performed very well.
- The actual vs predicted price plot showed a **close match**, with predictions tracking real trends.
- Mean Absolute Error and other metrics indicated **strong accuracy** for a basic model.

Observations:

- Using time-lagged features improved results.
 - The model successfully captured the short-term trend of stock prices.
-

Task 3: Heart Disease Prediction

Objective:

To predict whether a patient is at risk of heart disease based on health attributes using a classification model.

Dataset:

- UCI Heart Disease Dataset from Kaggle
- Features: age, sex, cholesterol, max heart rate, chest pain type, etc.
- Target: 1 (disease) or 0 (no disease)

Model Used:

- Logistic Regression

Results:

- **Accuracy:** 79.5%
- **ROC AUC Score:** 0.88
- Confusion matrix showed balanced predictions

Important Features:

- Chest pain type (cp)
- Max heart rate (thalach)
- ST depression (oldpeak)
- Slope of the ST segment (slope)

Observations:

- The model is effective and interpretable.
 - It can help in early prediction of heart disease risk.
 - The ROC curve confirmed the model's strong classification performance.
-