**DevelopersHub Corporation**

# AI/ML Engineering Internship Tasks

**Due Date:** 2nd August, 2025

# Overview

As part of your AI/ML Engineering Internship at DevelopersHub Corporation, you are required to complete **at least 3 out of the following 6 tasks**. These tasks are designed to help you build a strong foundation in artificial intelligence and machine learning through real-world datasets and use cases.

You will practice skills like data preprocessing, model training, evaluation, prompt engineering, and even chatbot development — all of which are essential in AI/ML careers.

# Task 1: Exploring and Visualizing a Simple Dataset

**Objective:**

Learn how to load, inspect, and visualize a dataset to understand data trends and distributions.

**Dataset:**

Iris Dataset (CSV format, can be loaded via seaborn or downloaded) **Instructions:**

* Load the dataset using pandas.

* Print the shape, column names, and the first few rows using .head().

* Use .info() and .describe() for summary statistics.

* Visualize the dataset:

○ Create a scatter plot to show relationships between features.

○ Use histograms to show value distributions.

○ Use box plots to identify outliers.

* Use matplotlib and seaborn for plotting.

**Skills:**

* Data loading and inspection using pandas

* Descriptive statistics and data exploration

* Basic plotting and visualization with seaborn and matplotlib

# Task 2: Predict Future Stock Prices (Short-Term)

**Objective:**

Use historical stock data to predict the next day's closing price.

**Dataset:**

Stock market data from Yahoo Finance (retrieved using the yfinance Python library)

**Instructions:**

* Select a stock (e.g., Apple, Tesla).

* Load historical data using the yfinance library.

* Use features like Open, High, Low, and Volume to predict the next Close price.

* Train a Linear Regression or Random Forest model.

* Plot actual vs predicted closing prices for comparison.

**Skills:**

* Time series data handling

* Regression modeling

* Data fetching using APIs (yfinance)

* Plotting predictions vs real data

# Task 3: Heart Disease Prediction

**Objective:**

Build a model to predict whether a person is at risk of heart disease based on their health data.

**Dataset:**

Heart Disease UCI Dataset (available on Kaggle)

**Instructions:**

* Clean the dataset (handle missing values if any).

* Perform Exploratory Data Analysis (EDA) to understand trends.

* Train a classification model (Logistic Regression or Decision Tree).

* Evaluate using metrics: accuracy, ROC curve, and confusion matrix.

* Highlight important features affecting prediction.

**Skills:**

* Binary classification

* Medical data understanding and interpretation

* Model evaluation using ROC-AUC and confusion matrix

* Feature importance analysis

# Task 4: General Health Query Chatbot (Prompt Engineering Based)

**Objective:**

Create a chatbot that can answer general health-related questions using an LLM (Large Language Model).

**Tools:**

OpenAI GPT-3.5 via API (or use a free open-source model like Mistral-7B-Instruct on Hugging Face)

**Instructions:**

* Build a simple Python script or notebook that sends user queries to an LLM.

* Use prompt engineering to make the responses friendly and clear (e.g., "Act like a helpful medical assistant").

* Add safety filters to avoid giving medical advice that could be harmful.

* Example queries:

○ "What causes a sore throat?"

○ "Is paracetamol safe for children?"

**Skills:**

* Prompt design and testing

* Using APIs for LLMs (e.g., OpenAI, Hugging Face)

* Safety handling in chatbot responses

* Building simple conversational agents

# Task 5: Mental Health Support Chatbot (Fine-Tuned)

**Objective:**

Build a basic chatbot that provides supportive and empathetic responses for stress, anxiety, and emotional wellness.

**Model Base:**

DistilGPT2, GPT-Neo, or Mistral (7B)

**Dataset for Fine-Tuning:**

EmpatheticDialogues (Facebook AI)

**Instructions:**

* Fine-tune a small LLM using Hugging Face’s Trainer API.

* Train it to respond empathetically using real human dialogues.

* Ensure the tone is gentle and emotionally supportive.

* Build a command-line or Streamlit-based interface to test it.

**Skills:**

* Model fine-tuning with Hugging Face Transformers

* Emotional tone design for safe chatbots

* Conversation modeling

* Deployment using CLI or simple web apps

# Task 6: House Price Prediction

**Objective:**

Predict house prices using property features such as size, bedrooms, and location.

**Dataset:**

House Price Prediction Dataset (available on Kaggle)

**Instructions:**

* Perform preprocessing on features like square footage, number of bedrooms, and location.

* Train a regression model (Linear Regression or Gradient Boosting).

* Visualize predicted prices compared to actual prices.

* Evaluate with Mean Absolute Error (MAE) and RMSE.

**Skills:**

* Regression modeling

* Feature scaling and selection

* Model evaluation (MAE, RMSE)

* Real estate data understanding

# Submission Requirements (Checklist for Each Task)

To receive credit for a task, ensure the following:

## 1. Jupyter Notebook

* Clear problem statement and goal

* Dataset loading and preprocessing

* Data visualization and exploration

* Model training and evaluation

* Explanation of results and final insights

## 2. Code Quality

* Code should be clean, modular, and commented

* Include explanations of each major step

## 3. GitHub Repository

* Create a GitHub repo for your AI/ML internship tasks

* Add a README.md file summarizing:

○ Task objective

○ Dataset used

○ Models applied

○ Key results and findings

## 4. Submission on Google Classroom

● Share your GitHub repository link for each completed task.

# Important Note

* You are required to complete **at least 3 out of the 6 tasks** before the due date: **2nd August 2025**.

* You may complete more tasks for additional learning and project portfolio strength.

* If you need help, feel free to ask questions in the group or consult with your mentors.