ML PROBLEM STATEMENTS | HUBBLEMIND

Project Description:

The goal of this project is to predict stock prices using linear regression. Interns will work with a historical dataset that includes various commodity prices, volumes, and stock prices. This project will help interns understand and apply key machine learning concepts, including Exploratory Data Analysis (EDA), feature engineering, model development, model evaluation, and prediction system creation.

Dataset Link : [Download](#)

Problem Statements:

Develop a linear regression model to predict Amazon stock prices based on historical commodity prices and volumes. The project will cover the entire ML process, from EDA and feature engineering to model development and evaluation.

Week 1: Data Exploration and Preprocessing

Task 1: Import and Understand Data
- ➤ Load the dataset.
- ➤ Display the first few rows to understand its structure.
- ➤ Check data types, missing values, and basic statistics.

Task 2: Exploratory Data Analysis (EDA)
- ➤ Visualize the distribution of the target variable (Amazon_Price).
- ➤ Analyze relationships between Amazon_Price and other features using scatter plots.
- ➤ Compute and visualize the correlation matrix.

Task 3: Feature Engineering
- ➤ Handle missing values by filling them with the mean of the respective columns.
- ➤ Standardize the features using StandardScaler.

Week 2: Model Development

Task 1: Split the Data
- ➢ Define the features (X) and the target variable (y).
- ➢ Split the data into training and testing sets using an 80-20 split.

Task 2: Train the Model
- ➢ Initialize a Linear Regression model.
- ➢ Train the model on the training data.
- ➢ Predict on the testing data.
- ➢ Evaluate the model using Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE).

Week 3: Model Validation and Testing

Task 1: Cross-Validation
- ➢ Perform 5-fold cross-validation to assess the model's performance.
- ➢ Calculate the mean and standard deviation of the cross-validation scores.

Task 2: Test Model on Unseen Data
- ➢ Use a portion of the testing data as recent data for validation.
- ➢ Predict on recent data and evaluate the model using MAE and RMSE.

Task 3: Feature Importance Analysis
- ➢ Analyze the coefficients of the linear regression model to determine feature importance.
- ➢ Visualize the feature importances using a bar plot.

Week 4 : Documentation and Submission

Focus: Summarizing the project, documenting code, and submitting your work.

- ➢ **Write a Project Summary:** Cover the introduction, challenges, solutions, roadblocks, and conclusion.
- ➢ **Compile Code:** Gather all code from Weeks 1 to 3 into a single Jupyter Notebook, ensuring explanations are provided for each code block.
- ➢ **Upload to GitHub:** Upload the Jupyter Notebook to a GitHub repository and include a `README.md` file that summarizes the project.
- ➢ **Create a Google Doc:** Include the project summary and the GitHub repository link.
- ➢ **Submit the Google Doc:** Submit the document via the provided **Google Form link.**