Introduction

This code implements the prediction model for stock market data using Multiple Linear Regression and Long Short-term Memory. It is designed to predict the future value of stock (‘Open’, ‘High’, ‘Low’, ‘Close’) for a given company.

Prerequisites

1. Ensure you have Python installed, along with the required libraries: csv, pandas, sklearn, numpy, and matplotlib.
2. Place the data files, 'sp500wiki.csv' and 'data.csv', in the same directory as the code.

Code Overview

1. The code begins with importing necessary libraries.
2. It reads the stock market data from 'sp500wiki.csv' and 'data.csv' files and stores data in dictionary format.
3. It excludes the insignificant variables and removes rows with missing values (zero entries).
4. A 5-day moving average calculation is applied to the ‘Open’, ‘High’, ‘Low’, ‘Close’ features.
5. The ‘createData’ function prepares x and y datasets for a specific company symbol and feature, where the x dataset contains the stock price and news volume of the previous day along with the 5-day moving averages, and the y dataset contains the stock price of the current day.
6. The ‘linear_regression’ and ‘lstm’ function will be implemented for the training of models and will be explained in detail below.
7. The ‘predict_using_linear_regression’ and ‘predict_using_lstm’ are implemented for prediction of prices.

1. **Multiple Linear Regression**

   The ‘linear_regression’ function performs the following stages:

   - Calls 'createData' to create 'X' and 'y' datasets.
   - Splits the data into training and testing sets (8:2 ratio).
   - Normalizes the training and testing datasets.
   - Trains a Linear Regression model based on the training data.
   - Predicts values using the testing data.
   - Calculates the Root Mean Squared Error (RMSE) for model evaluation.

   Input

   - Call the ‘linear_regression’ function with your preferred company symbol and feature (‘Open’, ‘High’, ‘Low’, ‘Close’) for prediction.

   Output

   - The ‘linear_regression’ function will output the RMSE value, providing insight into the accuracy of the prediction for the specified company and feature.

2. **Long Short-Term Memory**

   The ‘lstm’ function performs the following stages:

   - Calls 'createData' to create 'X' and 'y' datasets.
   - Splits the data into training and testing sets (8:2 ratio).
   - Normalizes the training and testing datasets.
   - Trains a LSTM model based on the training data.
   - Predicts values using the testing data.
   - Calculates the Root Mean Squared Error (RMSE) for model evaluation.

   Input

   - Call the ‘lstm’ function with your preferred company symbol and feature (‘Open’, ‘High’, ‘Low’, ‘Close’) for prediction.

   Output

   - The ‘lstm’ function will output the RMSE value, providing insight into the accuracy of the prediction for the specified company and feature.

3. **Prediction**

There are two prediction functions: ‘predict_using_mlr’ and ‘predict_using_lstm’. The functions perform the following stages:

- Calls ‘linear_regression’ or ‘lstm’ to create the prediction model.
- Calls ‘mlr_predict’ or ‘lstm_predict’ to find the predicted value of the target feature using the trained model.
- Returns the predicted outcome.

Input

- Call the ‘predict_using_mlr’ or ‘predict_using_lstm’ function with the symbol of the company to predict (i.e. ‘MMM’), the feature to predict (i.e. ‘Close’), and the date to predict (i.e. ‘2022-06-07’).

Output

- The ‘predict_using_mlr’ and ‘predict_using_lstm’ functions will output the predicted value of the target feature of stock price on a certain day for the chosen company.