Joseph Baruch

CS 212

Project 1

10/20/2023

Project Description

This program will retrieve data from an API by making a HTTP GET request. After the data is retrieved, it will be converted into a Pandas data frame for easier manipulation and cleaning. With the now cleaned data, matplotlib will be used to plot the data, and prepare it for simple machine learning algorithms with scikit-learn. Through fetching, manipulating and presenting this data, I hope to compare it with external factors to give reason to changing trends.

Goals

I’ve always been interested in learning more about the stock market and utilizing the power of computers to help me along that path. With so much readily available data, stock data would be a great way to develop my skills as a python programmer with the goals of:

* Learning stock market trends (comparing this data to external factors).
* Becoming proficient in data cleaning and analysis.
* Gaining the ability to perform http requests to API’s.
* Learning the basics of machine learning using scikit-learn.

Test Data

Building upon my goals on understanding the stock market, daily values over the span of 5-20 years of the IBM stock would be a perfect example of test data. IBM is a huge name in the tech and programming industry, so it is well fit for this project. I will be using the Alpha Vantage API services as the root of my data that I will fetch.

Method

The below information describes different aspects on how I will complete this project including, version control, storing source code, API/Data sources, python libraries utilized and main Integrated Development Environment (IDE).

Version Control Tool: Git

Source Code Location: GitHub repository (Username: JosephPBaruch)

API/Data Source: Alpha Vantage Stock Market API (IBM Daily Values)

Python Libraries:

* Requests.
* Matplotlib.
* Pandas.
* Scikit-learn.

IDE: Jupyter (Jupyter Notebook)

Objectives

To keep this project on track and balanced with the rest of my school and work schedules, the following are objective checkpoint to be met by October 27th, 2023.

1. API Handling Complete:
   1. Gain Public access token.
   2. Complete get request to API and returned correct JSON format.
2. Pandas Data Frame:
   1. Convert entire returned JSON from API call to a Pandas Data Frame.
   2. Complete all data cleaning and only have data that will be graphed in a data frame.
3. Data Frame to Matplotlib for graphing and future machine learning handling.
   1. Convert data frame into matplotlib arrays (or lists).
   2. Plot data onto graph and ensure data is presented well.

After completing these main objectives, the only remaining steps will be to analyze the data using a machine learning library such as Scikit-learn and report on the data I have presented.