ENGR - E 516: Engineering Cloud Computing

Financial Portfolio Management

Authors: Athulya Anand, Atharva Pandit, Sricharraan Ramaswamy

1. Project Overview:

1.1. Need of Project:

In today's era, with everyone's schedule being so busy it can be difficult to keep up with regular updates and speed up with everything going on around us. One important factor out of which could include managing and getting timely updates on the ever-changing financial affairs, specifically because investment products are unlikely to remain the same over a timeline. Considering such a scenario, even meaningful financial advice can sometimes not serve the purpose, becoming invaluable. Thus, a customized financial portfolio holding stocks can exclusively cater to the investment needs of individuals, considering their financial interest, financial goals, requirement, willingness, and ability to deal with risk. Additionally, in most cases, investment portfolios must be designed in such a way where it can minimize an individual's tax burden.

1.2. Objective:

Through this project, we predominantly focus on building a customized financial portfolio to hold best performing stocks for investors and individuals who are interested in keeping up with financial affairs given their need and interests.

2. Project Introduction:

In alignment with our objective, a more detailed version of the proposed system is as explained below:

- Our proposed system will aim on building a **cloud-based** financial portfolio management system using **Machine Learning** algorithms such as ***Random Forest Algorithm*** to help users maximize their profit for a given time for which he/she chooses to invest.

- Secondly, this system will also focus on the total amount of money which the individual is willing to invest in, based on which our system will distribute the invested amount optimally on basis of the performances of stocks in the given time.

- Lastly, through this project, we mainly aim to bring out an answer to our ***research question*** as follows:
    - What is the performance of this financial portfolio system on a local machine Vs that on cloud.
    - Which has a better performance and why?

3. Related work and Gap analysis:

After conducting in depth research and analysis on the topic, we collectively accumulated abundant work found on two strands of the market timing through the expected returns and volatilities. Although, there is some work combining both the factors and its integration with machine learning, we have not come across any approach including the integration of cloud resources in it. Some of the studies conducted in this field are as mentioned:

- Kandel and Stambaugh (1996) study the predictability of equity returns and conclude that a predictor variable like dividend yield may have a significant impact on the best stock-to-cash allocation.

- Goyal and Welch (2008) thoroughly assess the performance of variables that have been recommended by the academic literature as being effective predictors of the equity premium and come to conflicting conclusions.

- Johannes et al. (2014) provides compelling evidence that, if investors include estimation risk and time-varying volatility in their optimum portfolio issues, they can leverage predictability to enhance out-of-sample performance.

- Gu et al. (2020)2 demonstrated the advantage of applying machine learning for empirical asset pricing and linked the improved predictive performance to the acceptance of non-linear predictor interactions. The most accurate methods for predicting returns were trees and neural networks.

- Nystrup et al. (2016)16 offer a dynamic asset allocation method based on hidden Markov models that is based on change point detection rather than fitting a model with a given number of regimes to the data, estimating any parameters, or assuming a certain distribution of the data.

  Referring to the important and significant aspects from each of the studies listed above, we have decided to implement machine learning models like Random Forest Classifier to predict the best stock that the user can invest in as per his constraints. Additionally, we will be focusing on incorporating the usage of cloud resources into this system to understand the overall improvements and impacts of cloud on the system.

4. Proposed Tasks and Detailed Methodology:

Our proposed system can be divided into parts:

- The first part will effectively yields to the need of users, enabling user interaction. We aim to collect the following details from each user:
  - The total amount the user is willing to invest.
  - The duration (in years) for which the user wants to invest.
  - The total number of stocks in which the user wishes to invest in.

  Let us take an example:
  - Suppose we have a user who is wanting to invest a sum of $500 in a single stock over the duration of 2 years, our proposed system will give the user a suggestion of the top two stocks with high performance and a good return value, given the user invests for a shorter time period.

- The second part of the proposed system is intended to be implemented on a desired cloud platform using necessary resources, thereby improving the efficiency of the system. We would then compare these results with the efficiency observed on the local system.

4.1. Team Members & Workload Allocation per member:

Our team consists of three members and each of their workload allocation is specified below:

- Athulya Anand – Integrating the desired proposed system on the cloud.
- Atharva Pandit – Full stack development
- Sricharraan Ramaswamy – Build the data pipeline and model by implementing the machine learning algorithm.

4.2. Planned Project Timeline:

We have planned to complete the overall building of the proposed system over the given timeline:

- ***End of October*** – Data collection, data cleaning & preprocessing, feature extraction, exploratory data analysis & visualizations along with relevant machine learning pipelines and models.
- ***Mid November –*** Complete first stage of cloud integration and testing of ML model.
- ***End of November*** – Complete final stage of cloud integration along with UI (User Interface) optimization.

5. Progress on the proposed tasks and Preliminary Observations

As per the above-mentioned proposed tasks, we have completed the following tasks:

- **Data collection**: Live data was collected from *yahoo finance*. This data mainly comprises of opening rate and closing rate per day for the period January 1st, 2012 till today. Additionally, it also contains the highest rate and lowest rate of a stock per day along with the adjacent closing rate and volume of each stock.
- **ML Model testing:** We decided to move forward with the Long Short-Term Memory (LSTM) ML model for this project. Using the live data collected, we predicted the daily closing stock price for the year 2022 and compared it with the current value and computed the root mean squared error value as 5%. Currently, we have worked on 5 companies namely, Apple (AAPL), Microsoft (MSFT), Disney (DIS), JP Morgan (JPM) and Micron Technology (MU).

  The table below depicts the time taken for each epoch to run and the root mean squared error achieved with respect to each company.

Stock Name	Epoch run time (in sec)	Root Mean Squared error (in %)
AAPL	3660	7.814
MSFT	3095	10.997
DIS	3762	5.000
JPM	3441	4.926
MU	3602	3.231

Table 1: Performance of LSTM model on each company

The below graph depicts the prediction of stock price for AAPL.

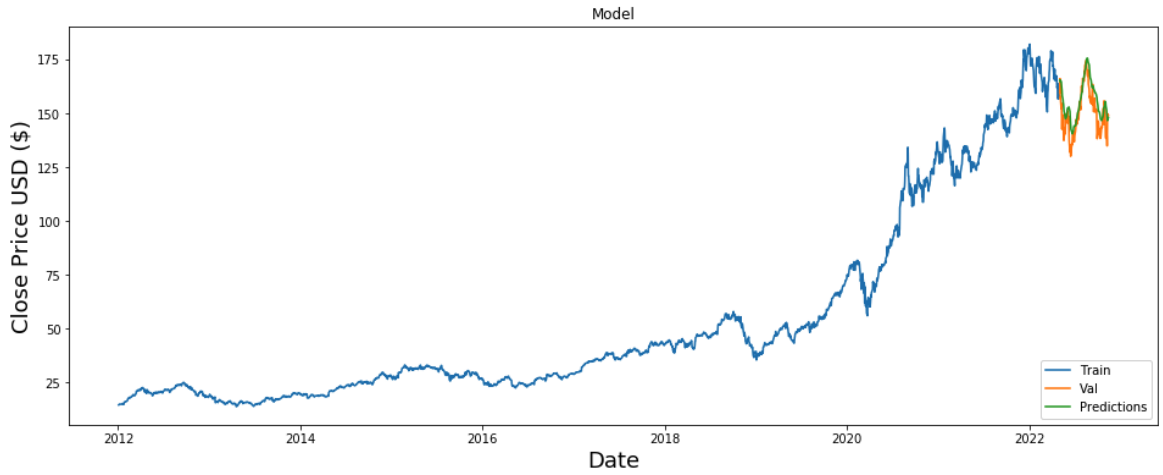

Fig 1: Stock Price prediction for AAPL

As per Fig 1, for the year 2022, the blue line represents the test data, the orange line represents the validation i.e. the actual closing rate of each stock and the green line represents the predictions i.e. the predictions made by our model for the stock. We infer that since the orange and green lines are approximately coinciding each other, the actual closing rate and the model predicted rate is nearly the same as shown in Fig 2. Hence, our model is successfully furnishing accurate results.

Date	Close	Predictions
2022-05-03	159.479996	165.074448
2022-05-04	166.020004	164.702194
2022-05-05	156.770004	165.340469
2022-05-06	157.279999	165.127457
2022-05-09	152.059998	164.671280
...	...	...
2022-11-08	139.500000	149.637512
2022-11-09	134.869995	148.254761
2022-11-10	146.869995	146.654755
2022-11-11	149.699997	146.750076
2022-11-14	148.985001	147.929581

136 rows × 2 columns

Fig 2: Closing rate Vs LSTM Model predicted rate for AAPL

- **User-Interface:** We have created a web page for users to access and manage their financial portfolio. This web page allows existing users to login with their credentials and new users to create a new account and login in a secure manner. Upon Signing in, the user can select the company for which they want to view the stock performance and predictions.

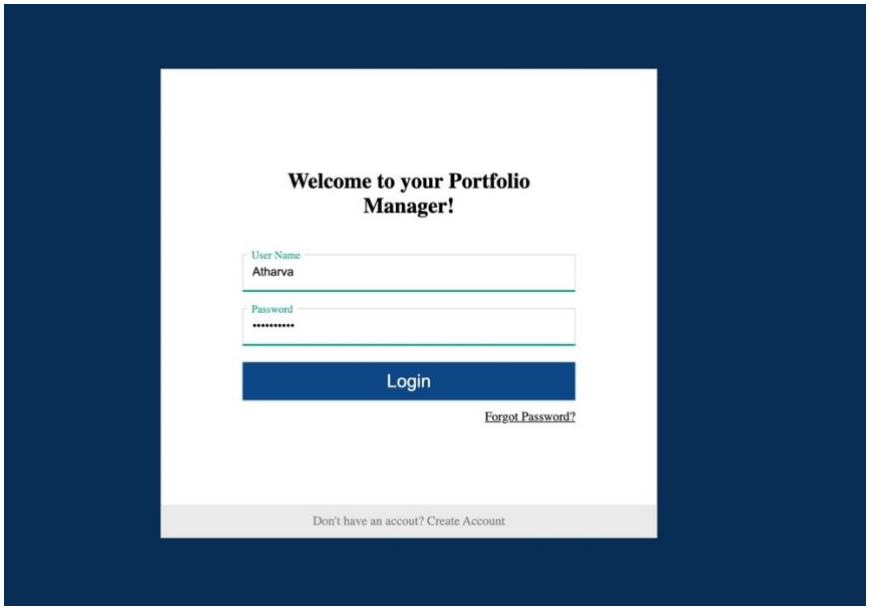

Fig 3: Log in page of the portfolio management system

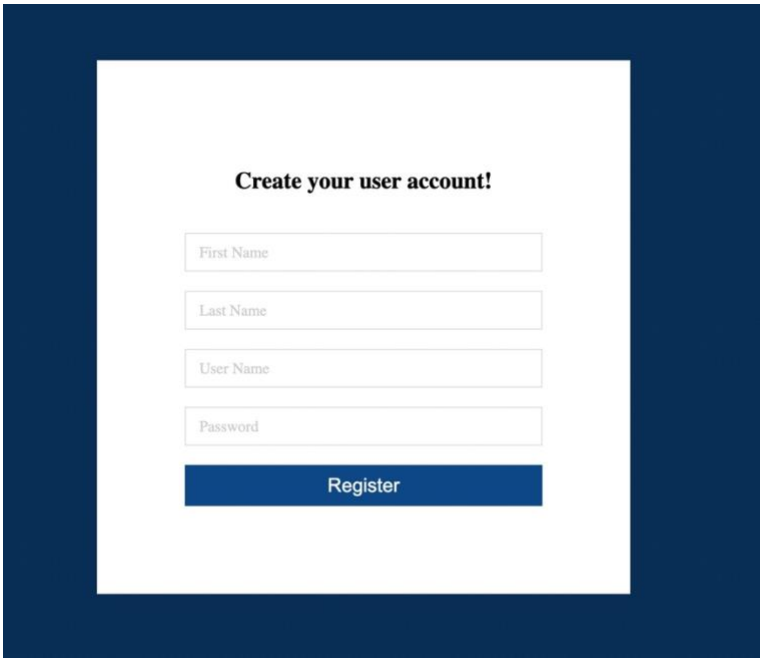

Fig 4: Registration page for new users

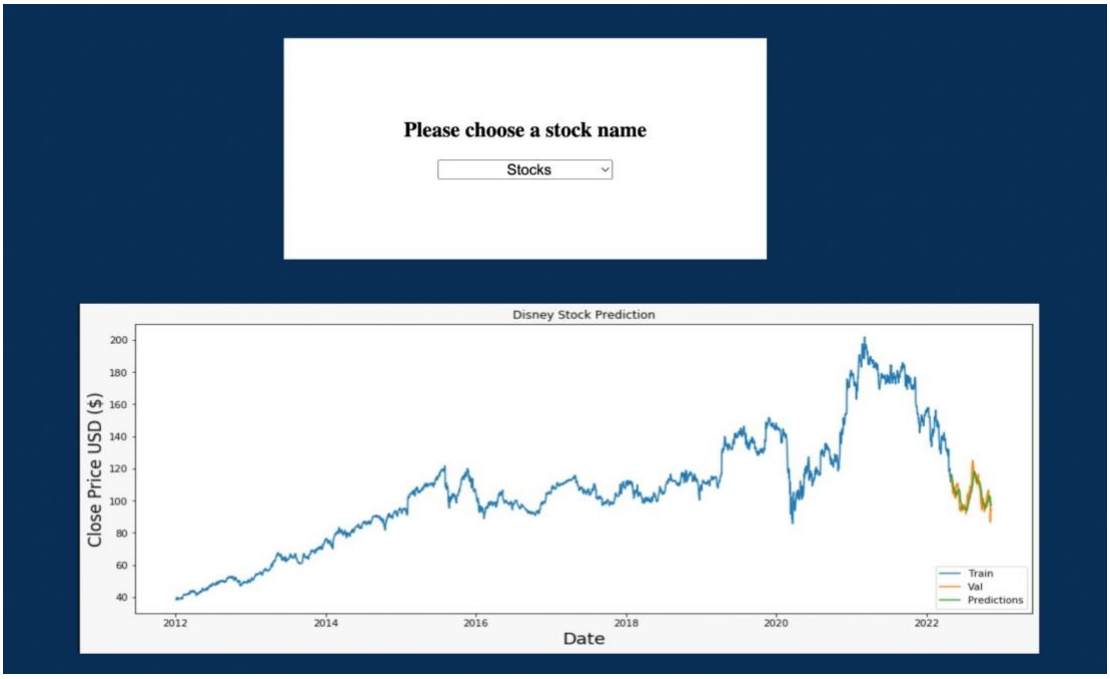

Fig 5: Page displaying stock performance and prediction

6. Next planned steps

- The major focus of this project is to integrate our system on Amazon Cloud Services.
- We plan on storing the generated database on Amazon S3 services.
- All the modules and packages required will be loaded on S3 services since AWS does not contain its own python packages.
- Linking S3 services with AWS lambda on which our machine learning model will be deployed.
- Allow users to choose multiple stocks for which he wants to view the stock performance and predictions.
- Call an FastAPI to get the user input and pass it to the machine learning model.
- Run the prediction model for each stock on each instance (similar to a mapper function).
- Collect the results of prediction model from each instance (similar to a reducer function) and display it to the user using the FastAPI.