Title – Portfolio Optimization Using Deep Learning Techniques

Group Number - 5

Group Members:

1. **Sukhad Joshi**
2. **Mihir Holmukhe**
3. **Vaibhav Gaikwad**
4. **Biswadip Bhattacharyya**

Objective :-

What is Portfolio Optimization?

1. Portfolio optimization involves selecting the best asset mix to maximize returns while minimizing risk, based on historical data and future predictions. Traditionally, methods like Markowitz's Mean-Variance Optimization are used. However, with financial markets becoming more complex, deep learning offers new ways to handle non-linear patterns and large datasets.

Why Deep Learning in Portfolio Optimization?

- Deep learning can model non-linear relationships, process large amounts of data, and capture complex dependencies over time. It provides advantages such as:

1. Handling high-dimensional financial data
2. Extracting features automatically
3. Making robust predictions based on time series data
4. Challenges in traditional methods include difficulty in handling non-linearities and market volatility, which deep learning techniques address more effective.

Problem: Traditional portfolio optimization methods can't fully grasp the complex, changing nature of financial markets, which limits their effectiveness.

Goal: To build a better way of predicting asset prices and market trends, helping investors make smarter decisions and optimize their portfolios.

Why Use CNN-LSTM:

- **Captures Asset Interactions**: CNN helps us understand how different assets relate to each other, which is key when you're managing a diverse portfolio.
- **Handles Time-Series Data**: LSTM is great at analyzing how things change over time, like stock prices, which gives us more accurate forecasts.
- **Combines Strengths**: Together, CNN and LSTM allow us to spot both short-term patterns and long-term trends, improving our ability to optimize portfolios.

- We're trying to help businesses, investors to make better choices when it comes to investing their money. Many traditional methods struggle to keep up with how quickly things change in the market and don't fully understand how different investments affect each other. By using a new approach called CNN-LSTM, we can more accurately predict future market trends.

DATASET DESCRIPTION:-

Overview –

For this Project, we are using Alpha Vantage API which is a popular service that provides real-time and historical financial data. It offers a wide range of data related to stocks, forex, cryptocurrencies, and technical indicators via an easy-to-use API. We are using Stock Data (Equities) which are provided as intraday, daily, weekly, and monthly stock prices. This includes open, high, low, close prices, and trading volume. The function which we have used is TIME_SERIES_DAILY which provides the daily stock price and volume.

The dataset consists of columns such as:

Timestamp - Represents the date of each data entry.

Open - The opening price of the stock for that specific day or period.

High - The highest price reached by the stock during the trading day or period.

Low - The lowest price reached by the stock during the trading day or period.

Close - The closing price of the stock at the end of the trading day or period.

Volume - The number of shares or contracts traded during the specific day or period.

NUMBER OF ROWS AND COLUMNS –

As we are collecting real time data from the API, the number of rows totally depend on the number of stocks in user's portfolio. During our analysis, we have observed that the API returns the previous 100 working days' entries which means

The number of rows = number of stocks in user's portfolio * 100

We are using 6 columns as mentioned previously. But after merging the dataframes, we will have an additional column "Symbol" which shows the name of the stock.

SAMPLE PREDICTORS –

- Open Price (1. open)
- High Price (2. high):
- Low Price (3. low):
- Close Price (4. close):

Preview of dataset (first 5 rows) for the sample portfolio

	timestamp	open	high	low	close	volume	Symbol
0	2024-10-21	91.19	91.260	90.360	90.87	2920760	XLC
1	2024-10-18	91.14	91.480	91.050	91.23	2665486	XLC
2	2024-10-17	91.39	91.410	90.400	90.56	2237560	XLC
3	2024-10-16	90.64	91.110	90.550	90.99	3198673	XLC
4	2024-10-15	91.05	91.445	90.765	91.05	3436527	XLC

DATASET URL –

This is a demo URL which shows the prices of IBM stock:

https://www.alphavantage.co/query?function=TIME_SERIES_WEEKLY_ADJUSTED&symbol=IBM&apikey=UGG1T1CD1N8A30I8

INTERESTING FACT ABOUT DATA –

There are significant price fluctuations within a single day, particularly between the high and low prices. These fluctuations can be influenced by news, earnings reports, or macroeconomic events. Unusually large price swings on specific dates, it could indicate unexpected market reactions or external factors impacting the stocks.

PRELIMINARY DATA EXPLORATION –

In the sample portfolio, the user has 2 stocks (XLC, XLV)

```
combined_data.shape
```

```
(200, 7)
```

```
combined_data.dtypes
```

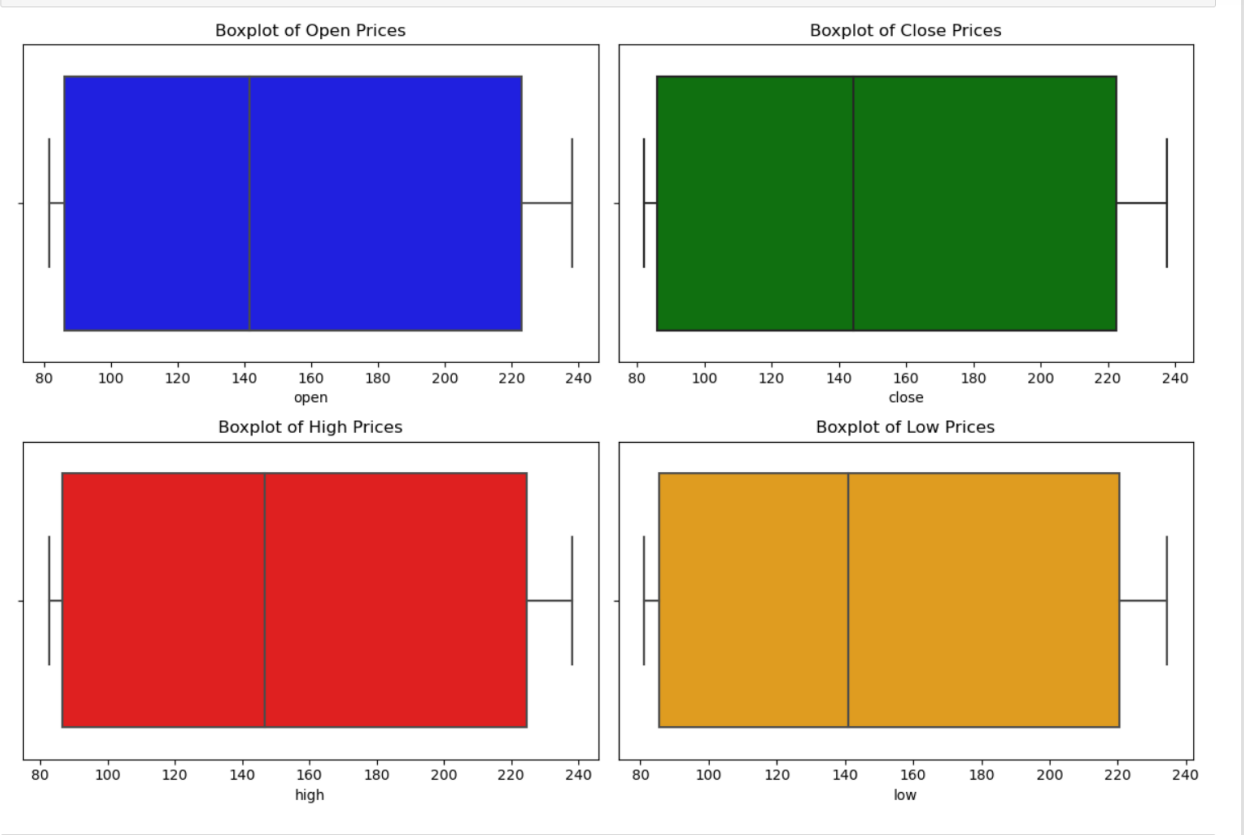
```
timestamp     object
open         float64
high         float64
low          float64
close        float64
volume         int64
Symbol        object
dtype: object
```

Therefore, we have 200 rows (100 entries per stock). The timestamp column is an object here so before performing any analysis we converted it into date time and used it as an index.

timestamp	open	high	low	close	volume	Symbol
2024-10-21	91.19	91.260	90.360	90.87	2920760	XLC
2024-10-18	91.14	91.480	91.050	91.23	2665486	XLC
2024-10-17	91.39	91.410	90.400	90.56	2237560	XLC
2024-10-16	90.64	91.110	90.550	90.99	3198673	XLC
2024-10-15	91.05	91.445	90.765	91.05	3436527	XLC

	open	high	low	close	volume
count	200.000000	200.000000	200.000000	200.000000	2.000000e+02
mean	153.938775	155.110746	152.538980	153.862950	4.573672e+06
std	67.907007	68.539856	67.035644	67.824161	1.967747e+06
min	81.460000	82.550000	81.150000	82.160000	1.861445e+06
25%	86.012500	86.507500	85.535000	86.055000	3.222389e+06
50%	141.647500	146.600000	140.897500	144.285000	4.030755e+06
75%	222.810000	224.685625	220.560000	222.500000	5.342176e+06
max	238.040000	238.140000	234.570000	237.680000	1.571495e+07

The summary statistics indicate the major components like mean for all the columns.

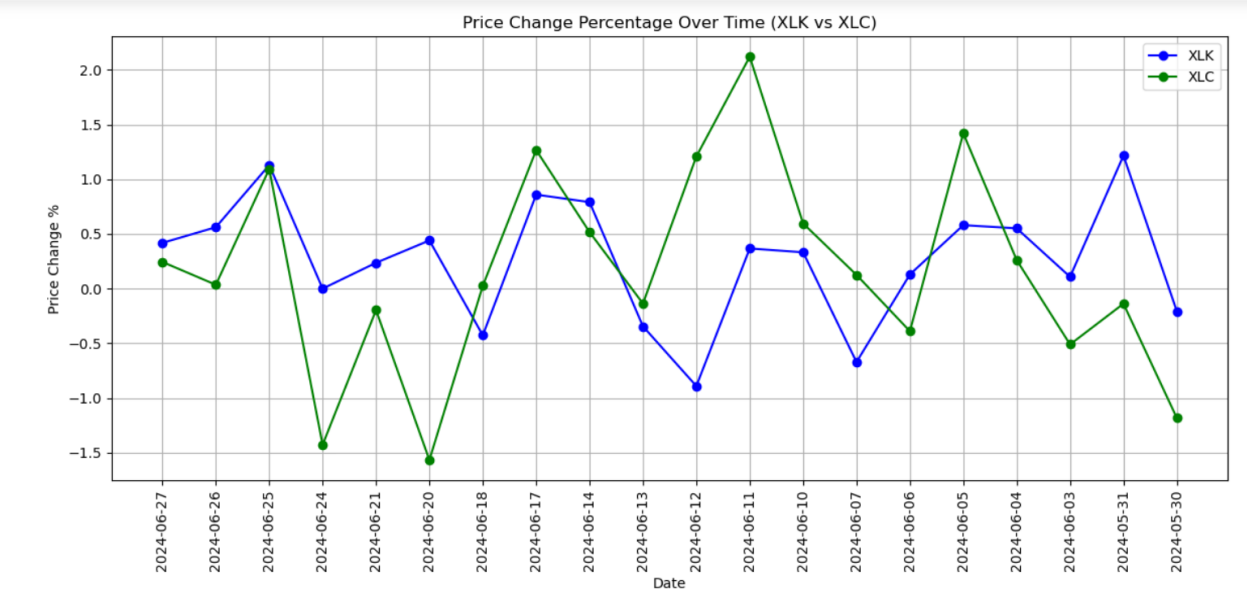


A key interesting fact about the dataset is that there are no null values and outliers as the API fetches clean real-time data, As indicated in the graph, there are no outliers in the dataset.

Sector Performance: The **technology sector (XLK)** has shown more price volatility and recovery throughout 2024 compared to the **communication services sector (XLC)**, which remained more stable.

Volatility in XLK: XLK experienced a noticeable dip in prices around August 2024, which might be due to market corrections, earnings reports, or other sector-specific events. The recovery that follows suggests that the market conditions improved, leading to increased investor confidence in the tech sector.

XLC Stability: XLC, on the other hand, exhibits consistent price behavior throughout this period, which may suggest less sensitivity to market volatility or slower growth compared to the technology sector.



This graph reveals that while both **XLK** and **XLC** faced similar overall market conditions, **XLC** exhibited higher volatility with more pronounced price changes, while **XLK** showed more stable, gradual shifts. **XLK** (blue line) tends to show smoother and smaller price changes compared to **XLC** (green line), which exhibits sharper price movements. This suggests that **XLC** may be more volatile than **XLK** during this period. The fluctuating price changes could be linked to broader market events, impacting both sectors during this time frame.

Proposed Predictions:-

Open Price (1. open): Useful for understanding the starting valuation for the day, which can be influenced by after-hours news or market conditions preceding the trading day.

High Price (2. high): Reflects the peak trading price during the day, indicating the maximum bullishness in the market for the stock.

Low Price (3. low): Indicates the lowest point reached during the day, showing the bearish pressures.

Close Price (4. close): Considered a critical metric as it determines the final settled price for the day and is often used as the reference price for the next trading day.

Candle Stick Chart will be implemented in order to give the investor a vague idea about how the stock is performing.