GROUP WORK PROJECT # 2 MScFE 642: Deep Learning for Finance

GROUP NUMBER: 7634

FULL LEGAL NAME	LOCATION (COUNTRY)	EMAIL ADDRESS	MARK X FOR ANY NON-CONTRIBUTING MEMBER
Chih wei Ng	Malaysia	ncw388@gmail.com	
Bharat Swami	India	bharatswami1299@gmail.com	
Pole Okuttu Stephen	Uganda	poleoks@gmail.com	

Statement of integrity: By typing the names of all group members in the text boxes below, you confirm that the assignment submitted is original work produced by the group (excluding any non-contributing members identified with an "X" above).	
Team member 1	**Chih wei Ng**
Team member 2	**Bharat Swami**
Team member 3	**Pole Okuttu Stephen**

Use the box below to explain any attempts to reach out to a non-contributing member. Type (N/A) if all members contributed. **Note:** You may be required to provide proof of your outreach to non-contributing members upon request.

Group Number: 7634

This report is done with reference to the Colab file.

Step 1

Five ETFs were used for this project, and their daily prices were downloaded using the yfinance library.

They are SPY (S&P500), TLT (20+ Year Treasury bond), SHY (1-3 Year Treasury Bond), GLD (Gold), DBO (Oil).

The period used is from 1 Jan 2012 to 31 Dec 2022. This period of analysis covers the training, validation, and test samples that will be used in the next steps. Some exploratory data analysis (EDA) was then performed to investigate some of the characteristics of the data.

For the price returns, all the 5 ETFs seem to be quite normally distributed and centred at zero, based on the histogram plot and the data description as seen below.

Ticker	DBO	GLD	SHY	SPY	TLT
count	2767.000000	2767.000000	2767.000000	2767.000000	2767.000000
mean	-0.000037	0.000076	0.000020	0.000530	0.000070
std	0.019416	0.009498	0.000698	0.010717	0.009098
min	-0.166453	-0.087808	-0.005088	-0.109424	-0.066683
25%	-0.009450	-0.004833	-0.000243	-0.003584	-0.005284
50%	0.001282	0.000341	0.000000	0.000599	0.000493
75%	0.010254	0.004972	0.000350	0.005496	0.005498
max	0.106227	0.049038	0.005452	0.090603	0.075196

Fig 1: Statistical description of the price returns of the 5 ETFs

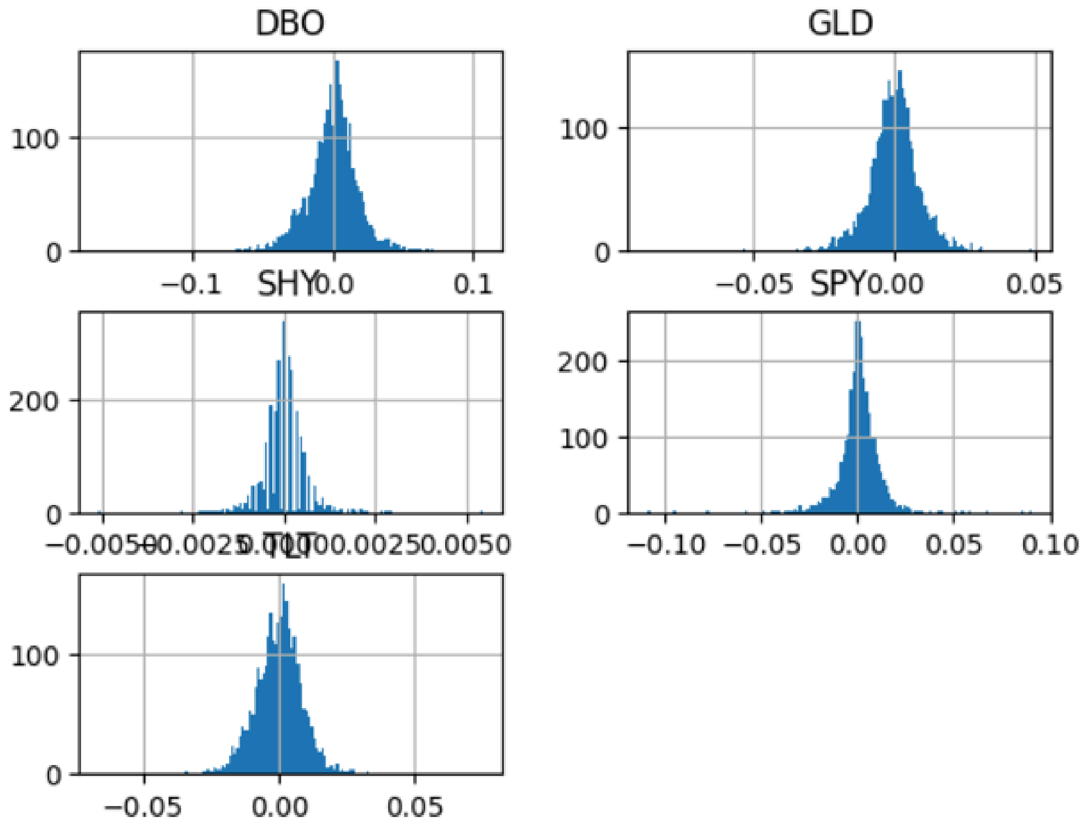

Fig 2: Histogram plots of the price returns of the 5 ETFs

Group Number: 7634

For an idea of the price movements of the 5 ETFs over the period, the SPY seems to be the only one to have gone up over time, while the others seem to be more or less stable over time, as shown below.

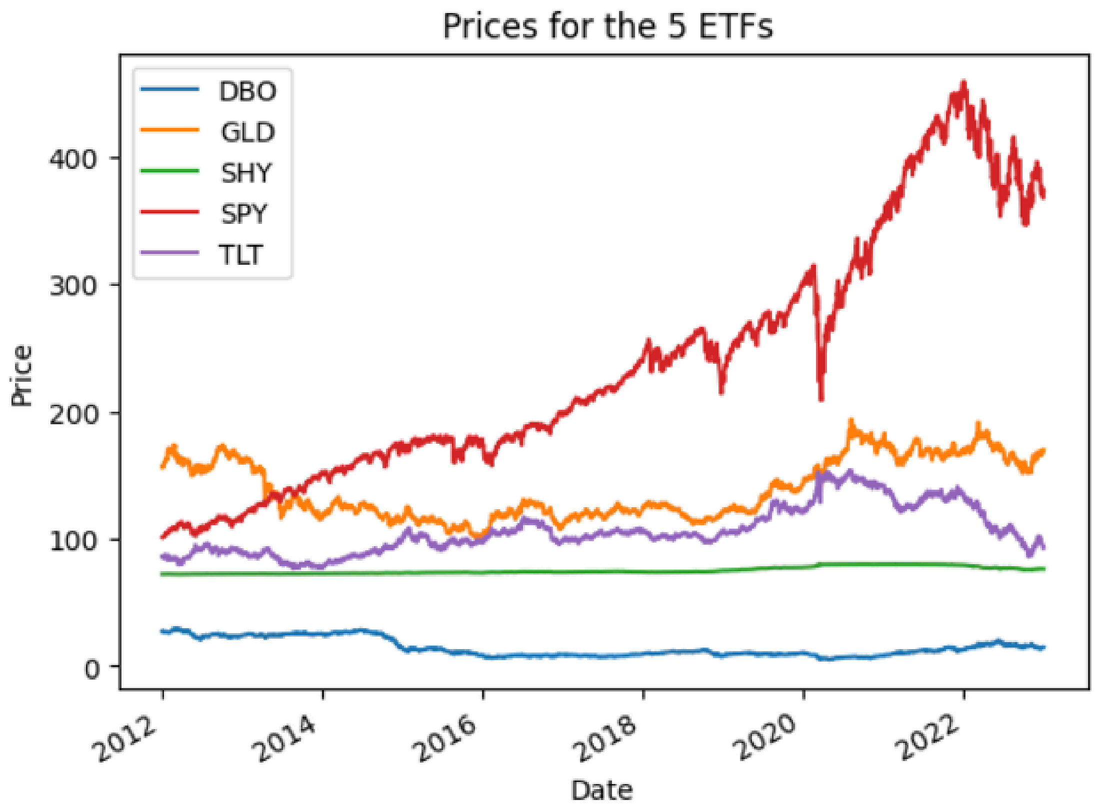

Fig 3: Plot of the prices of the 5 ETFs over the period.

A check on the stationarity of the data series using the Augmented Dickey Fuller (ADF) test showed that the price series of all the ETFs showed non-stationarity. However, using the daily price returns instead, all of them showed stationarity.

Group Number: 7634

Next, we did a check on the correlation of the 5 ETFs, as shown below.

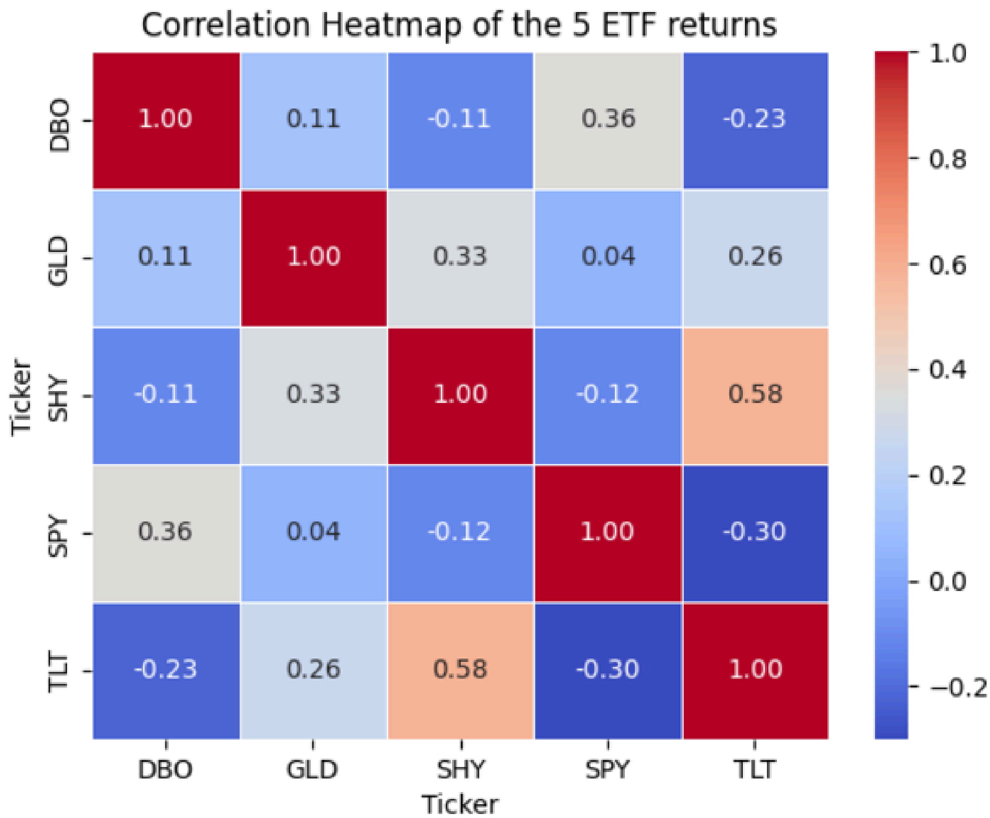

Fig 4: Correlation heatmap of the 5 ETFs.

The SHY and TLT ETFs showed quite a good correlation (0.58), which is to be expected as they are both referring to the similar financial instrument, i.e. Treasury bonds. The SPY and TLT ETFs showed quite a meaningful negative correlation, which is again expected, as SPY is related to the equity market, while TLT relates to the bond market, and under normal market conditions, the equity market is expected to be inversely correlated to the bond market.

Group Number: 7634

Following that, we investigated if there were any patterns in the data distribution based on certain selected frequencies. The data were aggregated and then averaged based on the selected frequencies.

Using the day of the month frequency, it is quite interesting to note that both TLT and SHY ETFs seem to generate negative returns during the first half of the month, while the second half of the month showed more positive returns.

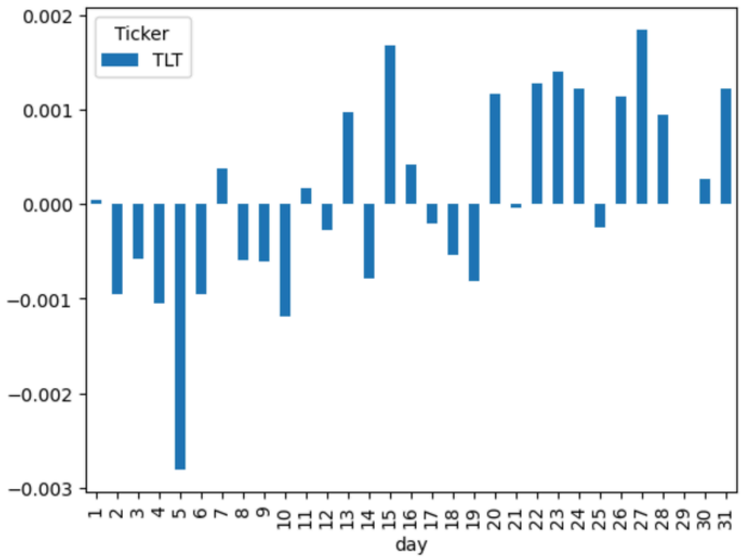

Fig 5: TLT aggregated returns for the different days in a month

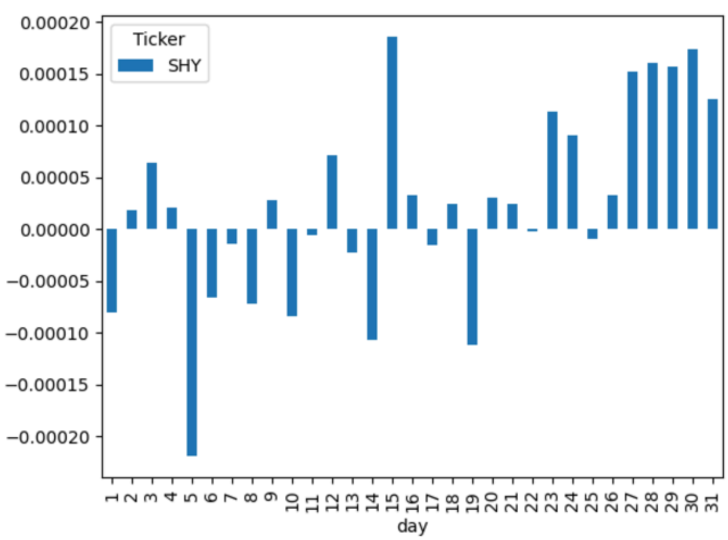

Fig 6: SHY aggregated returns for the different days in a month

Next, using the day of the week frequency, we can observe that for both TLT and SHY, they seem to get negative returns during the first two days of the week, while the remaining three days showed positive returns.

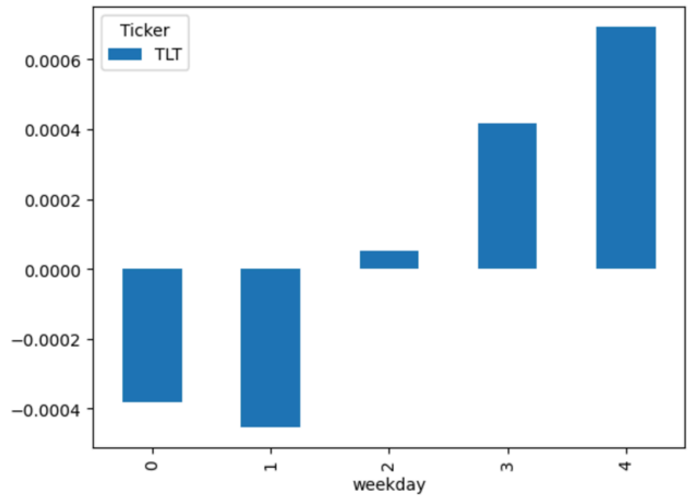

Fig 7: TLT aggregated returns for the different days of the week

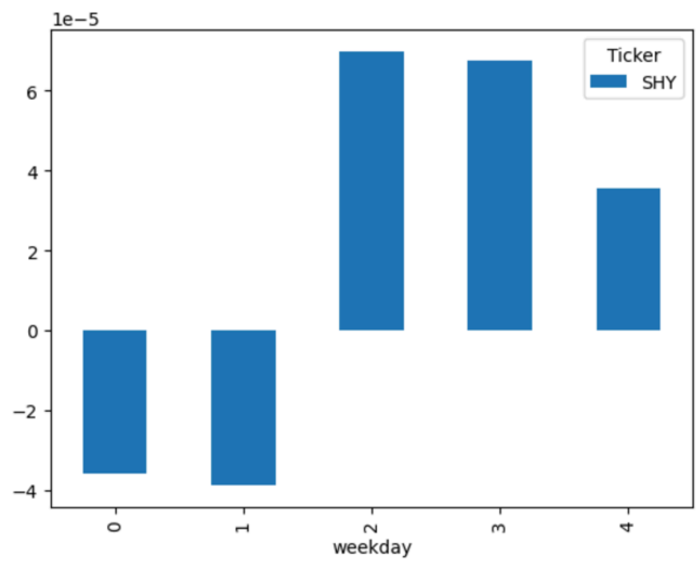

Fig 8: SHY aggregated returns for the different days of the week

Group Number: 7634

GLD also happens to show a similar pattern, as shown below.

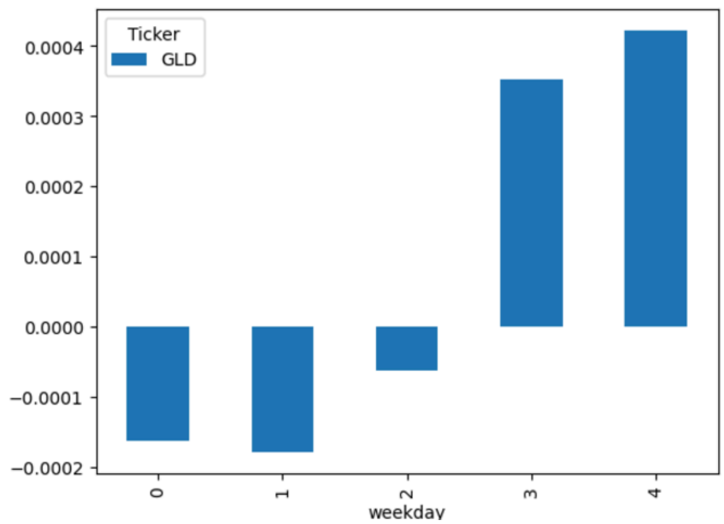

Fig 9: GLD aggregated returns for the different days of the week.

Step 2

In sample and out of sample performance

ETF	Train MSE	Test MSE
SPY	0.0108	0.0061
TLT	0.0147	0.0136
SHY	0.011	0.0072
GLD	0.0058	0.0122
DBO	0.0109	0.0

SPY and SHY assets have lower test MSE, indicating a better out of sample performance. They generalize well with new data.

On the other hand, TLT, GLD and DBO have a lower train MSE than test MSE, indicating the model predicts poorly with new data as compared to the train. Although, the difference between test MSE and train MSE is minimal indicating an almost balanced performance between in sample and out of sample data.

Trading Strategy

2 ETFs went long and 2 ETFs went short with the hope of better and worse performance respectively. The portfolio was rebalanced every 25 days from prediction information.

The strategy resulted in a cumulative loss of 34.49% during the test period.

The bad performance could be as a result of factors such as high volatility, tough market conditions due to COVID-19 pandemic, poor model tuning, etc.

This can be improved through further hyper parameter tuning, diversification of rebalancing periods (10 days, 40 days, etc.), model accuracy improvement, etc.

The buy and hold strategy performed well with a cumulative return of 32.58% in a similar time frame. This model performed better than the trading strategy.

Step 3

- **Overview of the Model:**
    - Using LSTM (Long Short-Term Memory) networks, we constructed a multi-output deep learning model in this stage to forecast the 25-day ahead returns for each of the five ETFs at the same time.
    - The model was created to forecast the future returns of all five asset classes (SPY, TLT, SHY, GLD, and DBO) using 30 days of historical data as input.
    - We made use of a number of features:
        - Returns: Changes in ETF prices as a percentage per day.
        - Log Prices: Prices that have been log-transformed to better accurately reflect percentage changes.
        - A 30-day rolling window is used to reflect recent volatility in the rolling standard deviation.
- **Preparing Data:**
    - In order for neural networks to function properly, the data was preprocessed using MinMaxScaler to scale the features to a [0,1] range.
    - 80% of the dataset was used for training, and 20% was used for testing.
- **Architecture of the Model:**
    - Using an LSTM layer with 50 units and a ReLU activation function, we employed a sequential model in Keras.
    - Five numbers are output by the dense layer, one for the expected return of each ETF.
    - As is common for regression assignments, the model was constructed using the Adam optimiser with mean squared error (MSE) as the loss function.
- **Training**:
    - A batch size of 32 was used to train the model over 20 epochs.
    - After every epoch of training, the test set was validated to keep an eye out for overfitting.
- **In-sample and out-of-sample evaluation:**
    - The test data was used to assess out-of-sample performance, and the training data was used to assess in-sample performance.
    - Mean Squared Error (MSE) was the evaluation metric that was employed.
- **Results:**
    - In-sample MSE: 7.999e-06
    - Out-of-sample MSE: 1.2752e-05
    - The model's ability to predict the 25-day ahead returns improves when more features (like volatility and log prices) are added.

Non- Technical Report

- **Overview of the Multi-Output Model**
    - We switched from employing separate models for every ETF to a multi-output model throughout this project phase. This means that we developed a single model that forecasts the future returns of all five ETFs at once, rather than having a different model for each ETF. This method is more effective and allows the model to make predictions for all asset classes while learning the links between them.
    - The model was created to forecast the returns for each ETF over the following 25 days and was trained using 60 days of historical data. The model can identify intricate patterns

in the data, including trends and volatility, by utilising a variety of features, including rolling standard deviation, log prices, and daily returns.

- **Model Evaluation**
  - The model's performance on the data it was trained on is referred to as in-sample performance. To get a sense of how well the model matches the data, we computed the Mean Squared Error (MSE) for the training set.
  - The model's performance on fresh, unseen data (the test set) is measured by out-of-sample performance. This is a more practical evaluation of the model's capacity to generalise and forecast future data with accuracy.
- **Trading Method Using Forecasts from the Model:**We used a trading strategy to make investment decisions based on the model's forecasts. The method works by:
  - purchasing the top two exchange-traded funds (ETFs) that the model indicates will yield the highest returns over the course of the following twenty-five days.
  - Selling the bottom two ETFs that the model indicates will yield the lowest returns is known as going short.
  - Every 25 days, this strategy is rebalanced in accordance with the model's revised projections.
- **Backtesting the Strategy:**Next, using historical data (January 2018–December 2022), we evaluated the trading method and contrasted its results with:
  - a buy-and-hold portfolio with identical weights for each ETF.
  - the results of a trading system that only predicts one ETF at a time, or single-output trading.

Group Number: 7634

Step 4

The results for single-output and multi-output architectures show significant differences in predictive performance and backtesting results. As seen above, the single output model registered low training MSE especially with low volatile assets like SHY and SPY, however MSEs were higher for assets like GLD and DBO which had higher volatility. This in summary led to a poor performance (-34.49%) in the portfolio returns from the backtest. On the other hand, the multi-output architecture had lower MSE across board for both in-sample and out of sample tests. This shows accuracy in training but also a possibility of overfitting.

The single-output model works with an asset independently which could be beneficial for unique asset types as compared to the multi-output approach that can take advantage of the interdependencies between assets to improve portfolio predictions.

References

1. Zhang, Ashton, et al. Dive into Deep Learning. https://d2l.ai/d2l-en.pdf