ISE 537 Final Project Report

Price Prediction: Comparison between ARIMA Model and LSTM

Name: Aditya Chavan | USC ID: 8741411805

1. Introduction

Predicting the price of a stock is a very complicated task. Using machine learning this task can be simplified. It can also be perfected using techniques like feature engineering. These algorithms will help investors to judge the nature of the market and assist them in the investment process. We will look at 2 methods namely ARIMA and LSTM to predict stock prices.

2. Data

For this project the aim is to look at stocks in the energy sector. I have chosen this sector as it is globally important and holds immense value for investing.

The stocks that I have chosen are:

a) Exxon Mobile (XOM): Oil and Gas
b) NextEra Energy (NEE): Renewable Energy

These are companies that contradict each other with respect to their environmental impact. One uses fossil fuels, and another is heavily invested in wind and solar energy.

Stock prices are obtained from the popular yahoo finance python library. The time period is from 1st Jan 2019 to 17th Nov 2023. This period is chosen for the following reasons:

a) To identify difference in the effect of COVID-19 on traditional fossil fuel companies and on renewable energy companies. (discussed ahead)
b) To see that if models are given enough data can they predict the trend of stock price in recession like environment.

Let's see the historical closing price for both the stocks.

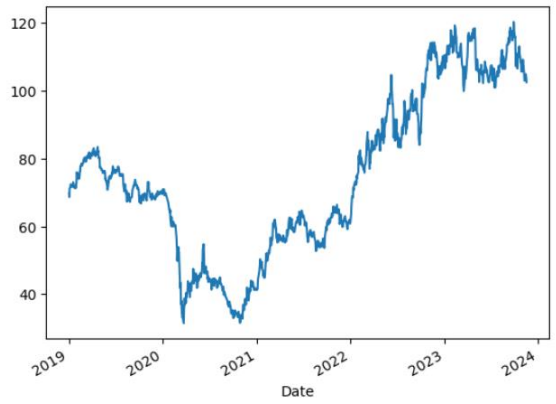

Fig 1: Closing prices for XOM

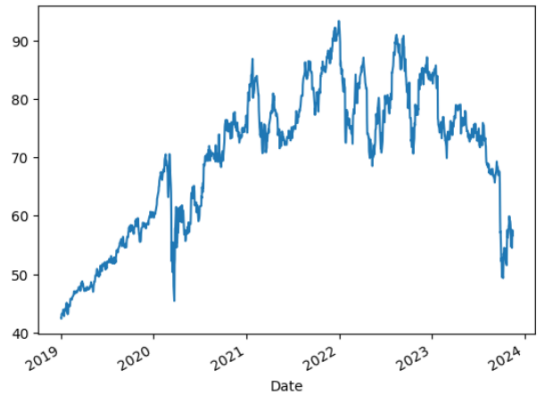

Fig 2: Closing prices for NEE

If we carefully observe both the graphs, we can see that after Covid time period XOM recovers slowly while NEE has a speedy recovery. One interesting thing to note is that in recent times Exxon has performed better than NEE, the stock price for which is depreciating.

3. ARIMA Model

The ARIMA model has 3 important parameters:

a) p is the number of autoregressive terms
b) d is the number of nonseasonal differences
c) q is the number of lagged forecast errors in the prediction equation

P and q were decided by looking at the AIC (Akaike Information Criterion) and BIC (Bayesian Information Criteria) of the ARIMA model.

```
p: 0  q: 0  aic: 4452.832078051973  bic: 4463.0583622693875
p: 0  q: 1  aic: 4454.737914981912  bic: 4470.077341308033
p: 0  q: 2  aic: 4455.736341211737  bic: 4476.188909646566
p: 1  q: 0  aic: 4454.743491016637  bic: 4470.082917342758
p: 1  q: 1  aic: 4456.56717157442  bic: 4477.019740009248
p: 1  q: 2  aic: 4457.022851795762  bic: 4482.588562339298
p: 2  q: 0  aic: 4455.635988897053  bic: 4476.088557331881
```

```
C:\Users\Aditya\anaconda3\Lib\site-packages\statsmodels\tsa\s
autoregressive parameters found. Using zeros as starting para
  warn('Non-stationary starting autoregressive parameters'
C:\Users\Aditya\anaconda3\Lib\site-packages\statsmodels\tsa\s
MA parameters found. Using zeros as starting parameters.
  warn('Non-invertible starting MA parameters found.'
```

```
p: 2  q: 1  aic: 4456.939883949868  bic: 4482.505594493404
p: 2  q: 2  aic: 4456.154949149463  bic: 4486.833801801706
p: 3  q: 0  aic: 4455.761066277738  bic: 4481.326776821274
p: 3  q: 1  aic: 4457.217249745388  bic: 4487.89610239763
p: 3  q: 2  aic: 4457.448041977179  bic: 4493.240036738129
p: 4  q: 0  aic: 4456.732527403665  bic: 4487.411380055907
p: 4  q: 1  aic: 4458.732693753409  bic: 4494.524688514359
p: 4  q: 2  aic: 4458.889915983093  bic: 4499.795052852749
p: 5  q: 0  aic: 4458.732295075475  bic: 4494.524289836425
p: 5  q: 1  aic: 4460.73213323533  bic: 4501.637270104987
```

Fig 3: AIC and BIC values for XOM ARIMA model

```
p: 0  q: 0  aic: 4020.738995859639  bic: 4030.965280077053
p: 0  q: 1  aic: 4022.1093133327017  bic: 4037.448739658823
p: 0  q: 2  aic: 4021.583670396012  bic: 4042.0362388308404
p: 1  q: 0  aic: 4022.054585278517  bic: 4037.3940116046383
p: 1  q: 1  aic: 4023.286731804922  bic: 4043.7393002397503
p: 1  q: 2  aic: 4017.743226970477  bic: 4043.308937514012
p: 2  q: 0  aic: 4021.7304556451204  bic: 4042.1830240799486
p: 2  q: 1  aic: 4023.685689101697  bic: 4049.251399645232
p: 2  q: 2  aic: 4003.858653967603  bic: 4034.5375066198453
p: 3  q: 0  aic: 4023.574919900707  bic: 4049.140630444242
p: 3  q: 1  aic: 4025.2682997966044  bic: 4055.9471524488467
p: 3  q: 2  aic: 4002.285618265455  bic: 4038.077613026405
p: 4  q: 0  aic: 4022.6585159256447  bic: 4053.337368577887
p: 4  q: 1  aic: 4006.631335879195  bic: 4042.4233306401447
```

```
C:\Users\Aditya\anaconda3\Lib\site-packages\statsmodels\tsa\sta
autoregressive parameters found. Using zeros as starting parame
  warn('Non-stationary starting autoregressive parameters'
C:\Users\Aditya\anaconda3\Lib\site-packages\statsmodels\tsa\sta
MA parameters found. Using zeros as starting parameters.
  warn('Non-invertible starting MA parameters found.'
```

```
p: 4  q: 2  aic: 4003.897634876159  bic: 4044.802771745816
p: 5  q: 0  aic: 4019.4424699401206  bic: 4055.2344647010705
p: 5  q: 1  aic: 4006.824457906373  bic: 4047.72959477603
p: 5  q: 2  aic: 4008.6978230892346  bic: 4054.7161020675985
```

Fig 4: AIC and BIC values for NEE ARIMA model

If we look at the AIC and BIC values they suggest (0,0) and (3,2) as (p,q) for XOM and NEE respectively. The p, q values are decided we will now see at the differential values for determining the d values.

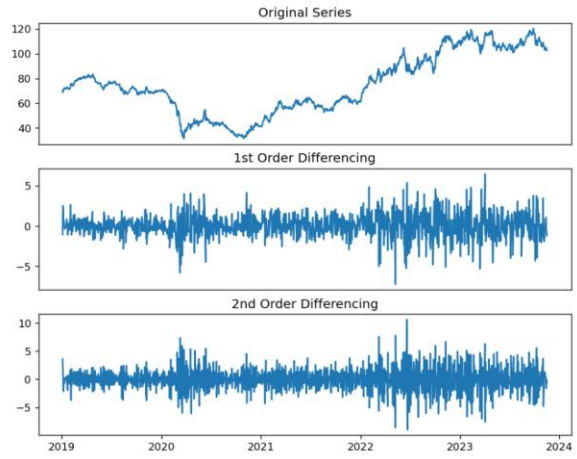

Fig 5: Differential comparison for XOM

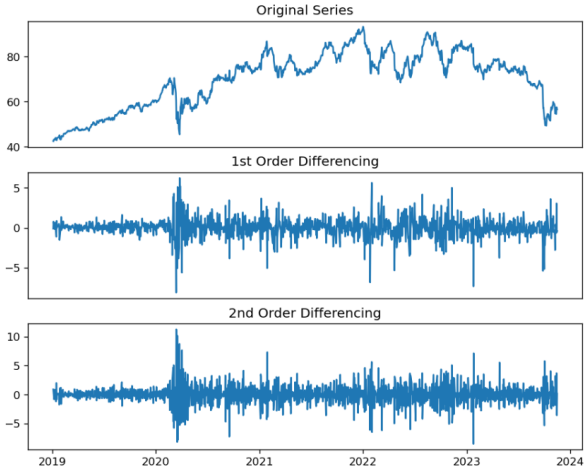

Fig 6: Differential comparison for NEE

We can see that the noise in the 2nd differential is significantly higher than the 1st in both cases. So we can set d = 1.

Below are the predictions that were made by the ARIMA model.

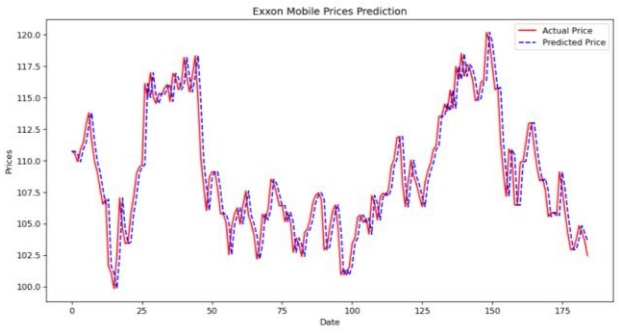

Fig 7: XOM Predictions using ARIMA

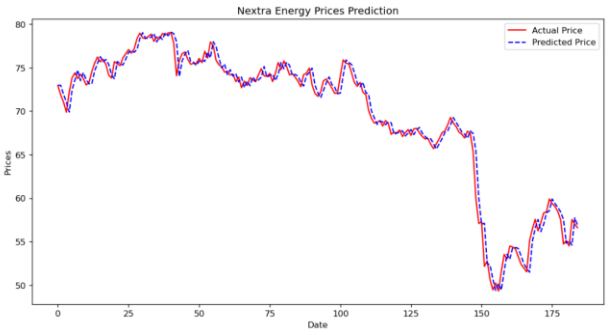

Fig 8: NEE Predictions using ARIMA

4. LSTM Model

LSTM stands for Long short-term memory. It is a type of a neural network model. They are well known for making predictions for time series data. In the project we feed LSTM with prices for last 60 days along with the close price for the last day. This creates 60 features for the model to train on.

We can see below that the training loss comes close to zero and becomes steady in that range around 10 epochs.

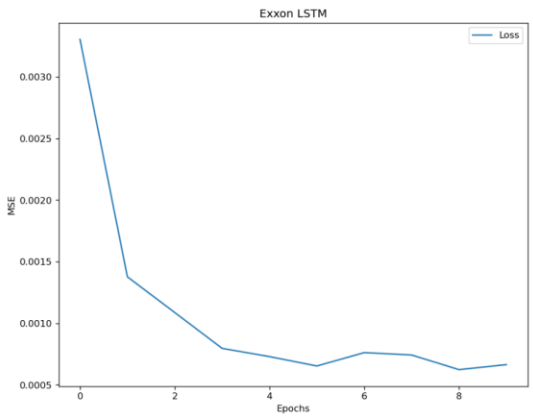

Fig 9: Loss plot for Exxon mobile.

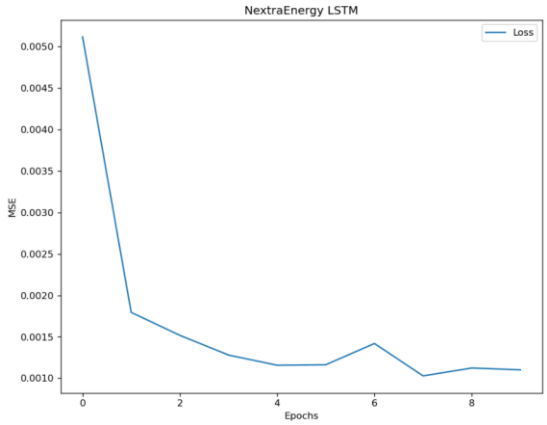

Fig 10: Loss plot for NextEra Energy

Below are the LSTM predictions for both the stocks.

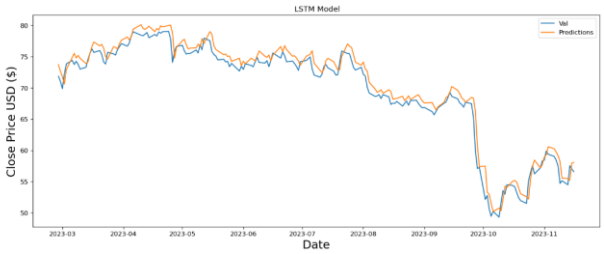

Fig 11: Predictions for NextEra energy

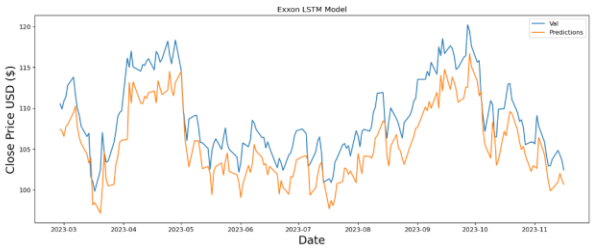

Fig 12: Predictions for Exxon mobile

5. Comparing Modelling Techniques

Exxon Mobile (XOM)		
Model	**MSE**	**RMSE**
ARIMA	**3.00**	**1.73**
LSTM	**13.69**	**3.70**

NextEra Energy (NEE)		
Model	**MSE**	**RMSE**
ARIMA	**1.30**	**1.14**
LSTM	**2.16**	**1.47**

You can see that ARIMA is performing better in our case. The Mean squared errors clearly show that. If you see the prediction graphs too it is clear that the ARIMA model is performing better than LSTM.

6. Conclusion

In this project we looked at LSTM and ARIMA models and tested it on stocks in Energy Sector. We found that ARIMA works better in our case.