AAST College of Artificial Intelligence 2023-2024

Time Series Data Analysis 12th Week Project Report

Names / Registration Numbers :

Abd El Rahman Ahmed Hossein Ahmed Abd El Halim Lasheen (20103045)

Youssef Hesham Shaheen (20107033)

Table of Contents:

- Introduction
- Forecast using traditional time series techniques.
- Forecast using Deep Learning Time Series Techniques (GRU)
- Clustering Techniques used in Time Series
- Conclusion

Section I : Introduction to Project & Description of Dataset

The Following is a report on the 12th week project where the procedures as well as screenshots and comparisons are discussed.

The Dataset Used is an infy stock dataset which is a part of dataset connected to the national stock exchange dataset.

We chose this dataset since the aim of this project is to forecast values according to time series and the dataset contains values perfect for this process.

Section II : Traditional Time Series Forecasting

First, we read the dataset using pandas and we perform pre-processing steps as well as data cleaning steps to ensure the minimum error in the dataset. This is done by checking for any NULL or duplicate values and removing them,

After this procedure, we perform the time series decomposition plot.
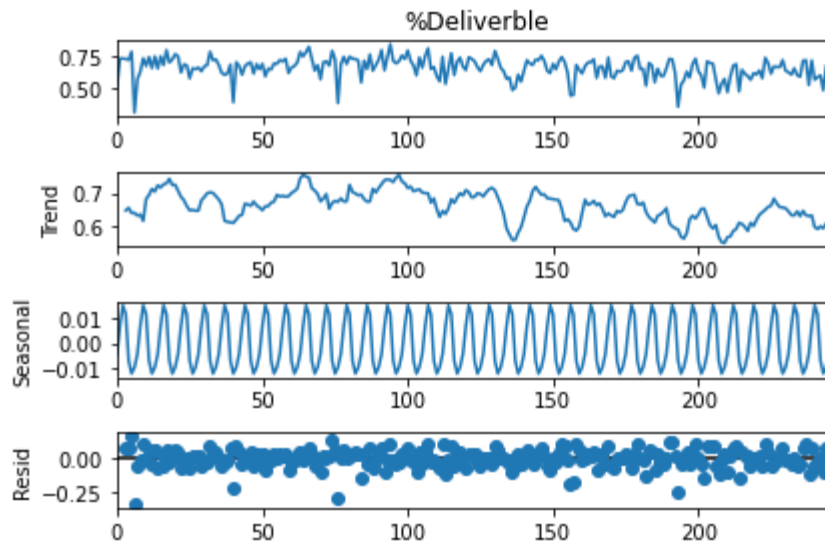
By applying decomposition on the '%Deliverable' attribute of the dataset,

We can visualize the data itself as well as its trend, seasonality, and residual.

Now we check for stationarity (in case if it is a random walk), this is done by applying the Augmented Dickey-Fuller test on the Deliverable column. This resulted in an ADF statistic of -6.916144464772943 but we also acquired a p-value of 1.17912e-09 (this is less than 0.05, indicating no existence of a unit root which rejects the null hypothesis. Therefore, It is stationary).

After Checking for stationarity, we check the correlations of the data through the ACF and PACF Plots.

Autocorrelation Function (ACF)


Partial Autocorrelation Function (PACF)
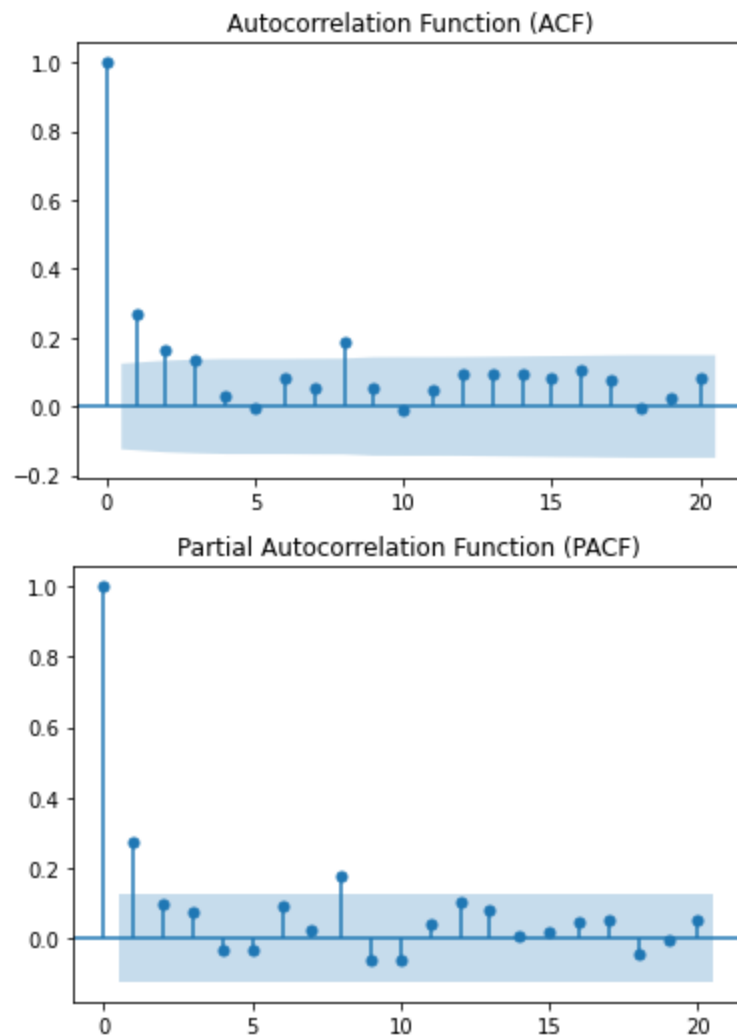
After Applying the following plots , we detect that value of p is 2 while the value of q is 1 (according to the ACF and PACF plots respectively).

Since we have both p and q values, it is not a moving average model nor an autoregressive model. But a model composed of both (ARMA or ARIMA)

To identify which one, we perform AIC test with respect to each p and q value.

(Where AIC = 2 * K(q + p) – (2 * ln(L)))

This results in the Best Order of (2,0,1)
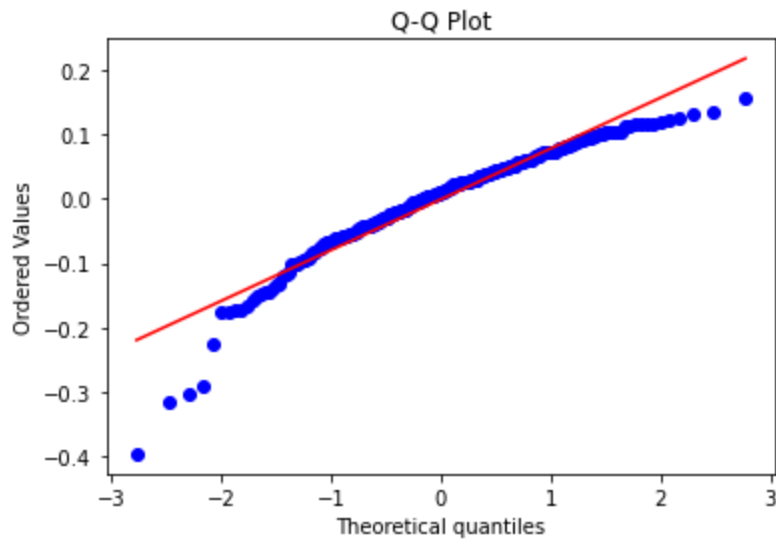
Q-Q Plot

According to the Q-Q plot, despite most p-values being more than 0.05, the data isn't normally distributed. Therefore, the data is not considered fit for forecasting.

Section III : Deep Learning in Time Series using Gated Recurrent Unit (GRU)

The Gated Recurrent Unit (GRU) is a type of traditional Recurrent Neural Network (RNN) Architecture used to resolve limitation issues of Usual RNNs and is used as an alternative to LSTM.

The Reason we Used GRU instead of LSTM is the fact that stock exchanges, while efficient, may require high computational power and a simplified code for efficient memory. So GRU outperforms LSTM in both conditions.

Comparison of Mean Squared Errors
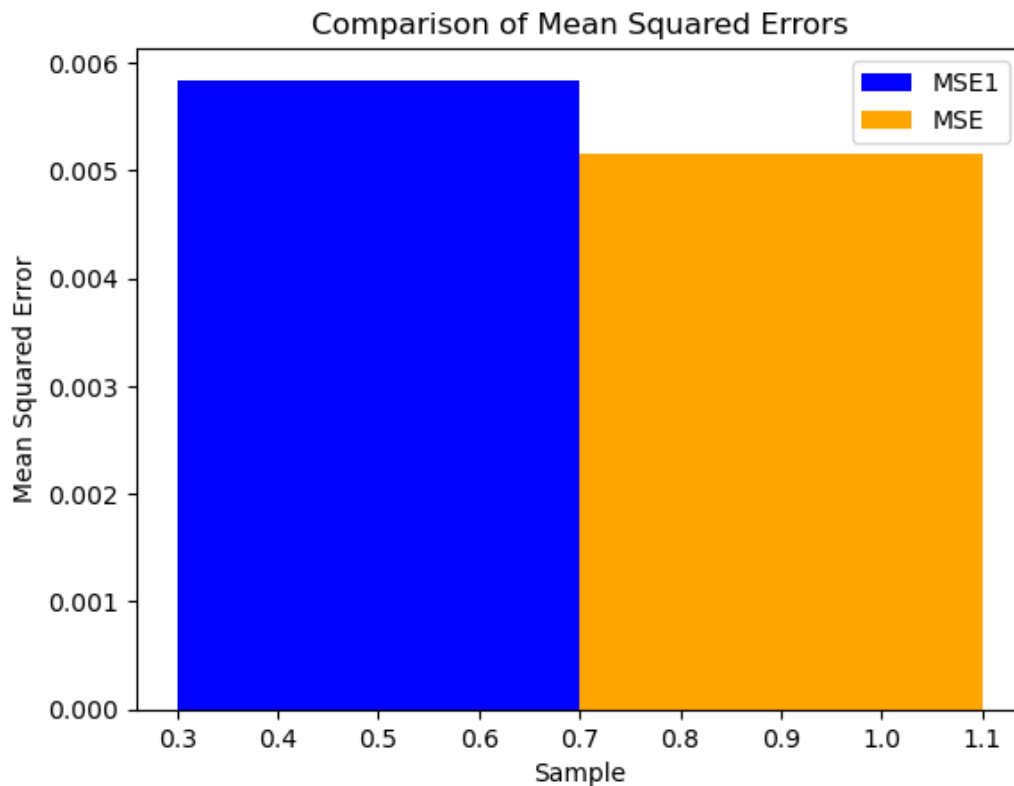
Next, we compare the mean squared error of the forecast (Traditional and Deep Learning) where MSE1 represents the MSE of the traditional method and MSE represents the deep learning method. This is done through a bar chart.

Section IV : Clustering Techniques Used

- K-Mean Clustering

A Clustering Technique which assigns points to clusters based on proximity to centroids, which becomes the mean of the assigned points. It is used for minimizing variance and focuses on efficiency.

- DBSCAN Clustering

Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is a clustering technique for grouping data into clusters. Like K-Mean,

What Makes DBSCAN Stand out from K-Mean is that DBSCAN doesn't require the number of cluster and can identify clusters of unique Shapes (not just circular like K-Mean),

- OPTICS Clustering

Ordering Points to identify the Clustering Structure (OPTICS), a density-based clustering algorithm which extends the concept of DBSCAN.

Its goal is to output a hierarchical clustering of data points based on density and rather than making one partition of clustering data points, it produces a reachability plot or distance that shows information about density-based structure of data.

Section V : Conclusion

In brief, the project has its challenges and outcomes. Using time series to forecast stock exchange is by all accountability, an essential core as it helps predict future sales as well as displaying statistics to identify the roots of a problem and resolve it quickly.

Thank You for your Attention and I look forward to our discussion.