Simulation of Stock Market to Forecast Price Indexes

*

Wajiha Tarannum Chaudhry
Department of Math and Natural Sciences (MNS)
School of Data and Sciences (SDS)
BRAC University
Dhaka, Bangladessh
wajiha.tarannum.chaudhry@g.bracu.ac.bd

***Abstract*—As Investing has become so important now a days and it's hard for people to make an informed decision due to the lack of tools to assess the market. Thus, This project aims to simulate the stock markets price indexes using a pipeline RNN.**

***Index Terms*—Forecasting, Time series, RNN**

I. Introduction

In this era of innovation and ever increasing inflation, It is a must for everyone to handle their wealth properly or it is much easy to lose money. While not losing in quantity rather losing in value. Thus, people look towards bank with minor risk to save for a minute amount of interest but unfortunately that is almost nothing now a days. Consequently, people start looking to other places to store/invest their money. Among them the top picks now would be Crypto Currency and NFT(Non-fungible Token), Unfortunately, this to has high volatility and risk as well. It is also hard to explain this to the groups of senior citizen to make them invest in them. Thus, Stocks is one of the best ways for people from all ages to invest in. Thus to give this people a tool to assess the market as well as to show them predicted future benefits. Thus, I forecasted using RNN,

A. Abbreviations and Acronyms

AR Auto Regression
ADF Augmented Dicky-Puller
A-VAR Advanced Vector Auto Regression
ARMA Auto Regressive Moving Average
ARIMA Auto Regressive Integrated Moving Average
BPNN Back Propagation Neural Network
Co-BPNN Connective BPNN
CIDM Computational Intelligence and Data Mining
CNN Convolutional Neural Network
DMAM Double Moving Average Mapping
FFNN Feed Forward Neural Network
GAF Gramian Angular Field
GRU Gated Recurrent Unit
MLP Multi-Layer Perceptron
MSE Mean Squared Error
RNN Recurrent Neural Network
SVM Support Vector Machine
STL Seasonal Trend Loss

II. Recent Works

Firstly, if talk about the work done by Sun and Xu which comparse different approaches for forecasting financial data. They concluded that the results show that the A-VAR model has lower RMSE than the ARIMA model with the same order of runtime performance. In the case of the BPNN model it is similar to the ARIMA model but has a longer runtime. The Co-BPNN performs better than the previous methods however it is slower than BPNN by one order of magnitude. The LSTM model performs the best by far while also having the longest runtime [1]. Aiming to present a reliable method to predict the fluctuations in stocks caused by seasonal changes by Sai and Sreela [2]. In there paper, SVM has been shown to be good with small datasets where there are fewer variations of instances. RF on the other hand is shown to be good for large datasets with a large number of instances and when generalizing new data. Their results show that about 30% of the data showed signs of following seasonal trends and are good places where investors are more likely to find profits. The remaining 70% being bad stock where even though there are seasonal variations, they are not easily predictable. In another paper a deep learning framework based on convolutional neural networks (CNN) to analyze financial time series data which results showed significant improvement from the expected performance which suggests that this approach allows visualization of patterns and correlations that are not apparent by typical methods. [3]. In the paper by M. Almuammar and M. Fasli talks of a method that uses Gated Recurrent Units (GRU) to forecast multivariate time series and compare it with the simple Multi-Layer Perceptron (MLP) [5].The authors conclude that even though the results suggest that GRU perform better than statistical methods they are out performed by MLPs in regards to time series forecasting. Tiwari, Bharadwaj and Gupta aim to apply various analysis techniques and models to compare and contrast between them and determine which approach is most suited for stock price index forecasting [6]. They concluded that the FFNN performed the best with the least amount of absolute percentage error with actual data. With seasonal data the Holt-Winters performed best. The fixed parameter ARIMA model did best with polynomial trend data. Fang Wang, Menggang Li b, Yiduo Mei e, and Wenrui Li proposed a time series analysis method to predict the future rise through the historical rise and fall probability distribution curve [7]. they predicted that there is a 84% probability that the stock's individual stocks will rise or fall between -4.5% and 4.35%. yang yujun, yang yime, Li jianping aimed to study on financial time series forecasting based on the support vector machine by conducting experiments with datasets [8]. From the result the authors deduce that This approach has better effect in the mature market (SP500 data sets which is an American stock market index) than in the immature market (HSI dataset which is a stock market of China), This is why I choose the SP500 dataset for this project.

III. Method

A. Time Series

The process of analysing of data point over a specific period of time is known as time series analysis. For time series data, it has to be collected after a certain period of time instead of the data being collected infrequently or randomly. The main feature that a time series brings in is the time series which lets us find patterns and trends of different sort. Some of the patterns would be Trend, Seaosnality, Auto correlation, Noise etc. All real life data would have some sort of ressembles to a mixture of this components. The following figure shows us a rough graph of all the components

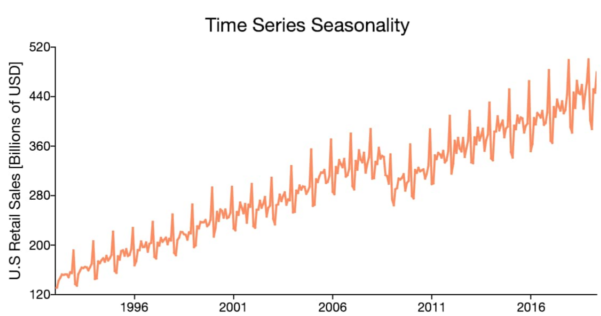

Fig. 1. Time series example

B. Neural Network

Neural Network is a series of algorithms that endeavors to finding underlying patterns in a set of data which mimic how our brains works. The following figure shows us how neural network is like. It has an Input layer, Output Layer and hidden layers and they are connected.

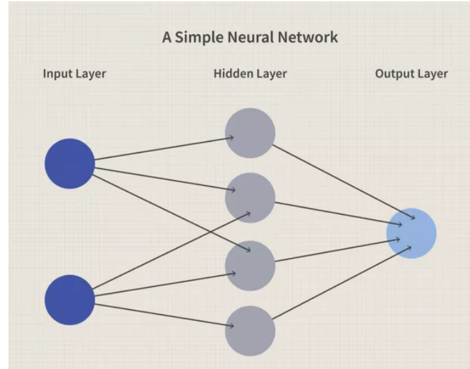

Fig. 2. Basic Neural Network structure

C. Recurrent Neural Network

Recurrent Neural Networks are a sort of FFNN which is focused on modeling in the temporal domain. Unlike FFNN or a basic NN the RNN it not only uses the current inputs that it has been exposed to but also all the previous inputs as well.

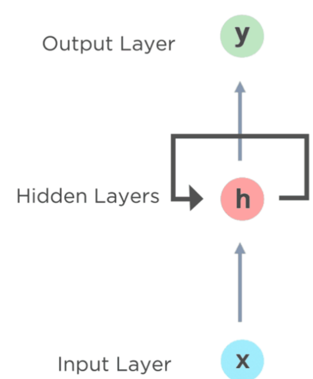

Fig. 3. Recurrent Neural Network structure

The figure above shows a basic structure of a neural network. One big thing to understand is that the decision a recurrent neural network reaches at timestep t-1 affects the decision it will reach at time step t. This sequential information is kept in the recurrent networks hidden state for future use. This manages to span many time steps. Such as in our humanly physique information transfers from place to place invisibly information in the hidden states circulates in the same fashion.

$$h_t = \phi(Wx_t + Uh_{t-1}) \tag{1}$$

h_t is the hidden state at time step t which is a function of x_t which is an input of the same time step and the W signifies that it was modified by a matrix which is further added to the h_{t-1} the hidden state of the previous time step which is also multiplied to the transition matrix U. The matrices work as filters that determine the importance that has to be accorded to both present input and it predecessing hidden state. Thus sum of the two are then squashed by the function ϕ. The ϕ

function in a normal case is a logistic sigmoid or tanh function. The figure 4 shows the data flow.

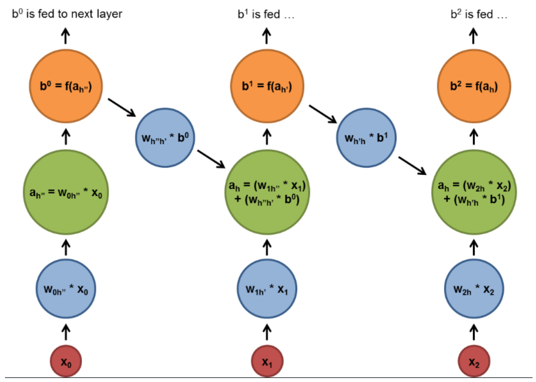

Fig. 4. Recurrent Neural Network dataflow

D. Methodologies

The ReLu activation function was used for the hidden layers. The total number of layers were 2 differing from type. A single unit dense layer was used for the output. I chose the Adam optimizer for the RNN.

IV. Results

The below figure shows us the result of our predicted value of stock which sort of simulates the stock market.

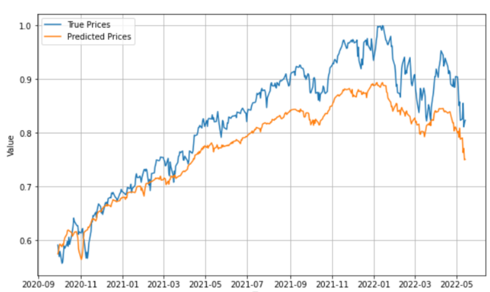

Fig. 5. Result of the predicted points

The values of Mean Squared Error and Mean Abolute error are given below:

Mean Squared Error	0.0033882614225149155
Mean Absolute Error	0.050406210124492645

V. Discussion

First I had to split the data set the I by using keras and given the parameters I completed the first sprint. When I ran the code for the first time it didn't seem so precise. The prediction was way off. Then by tuning the learning and loss rate then running it and by a bit of trial and error tinkering with the values of learning rate. I achieved the result.

VI. Conclusion and Future Work

We can see that via predicting we can use to help investor make informed decisions. As well as help people become motivated about investing. Thus helping the investors individually as well as the national and even the international economy as it facilitate the economic growth as well. Furthermore, This project can be taken a step forward by firstly optimizing this and introducing new variables help tinker with it more. In additon, a good GUI can be made for making it easier for people to directly interact with the simulation. We can also incorporate other statistical models such as ARRIMA, Prophet to get a more comparative view which might help in finding anomaly.

References

[1] H. Sun and J. Xu, "Improved approaches for financial market forecasting based on stationary time series analysis," in 2018 IEEE 3rd International Conference on Big Data Analysis (ICBDA), 2018, pp. 334–339. doi: 10.1109/ICBDA.2018.8367703.

[2] S. S. Ratakonda and S. Sasi, "Seasonal trend analysis on multi-variate time series data," in 2018 International Conference on Data Science and Engineering(ICDSE), 2018, pp. 1–6. doi: 10.1109/ICDSE.2018.8527804.

[3] J.-F. Chen, W.-L. Chen, C.-P. Huang, S.-H. Huang, and A.-P. Chen, "Financial time-series data analysis using deep convolutional neural networks," in 2016 7th International Conference on Cloud Computing and Big Data(CCBD), 2016, pp. 87–92. doi: 10.1109/CCBD.2016.027.

[4] C.-M. Guo, L.-Y. Xu, H.-F. Liu, L. Wang, X. Yu, and B. Han, "The financial data of anomaly detection research based on time series," in 2015 International Conference on Computer Science and Applications (CSA), 2015, pp. 86–89. doi: 10.1109/CSA.2015.42.

[5] M. Almuammar and M. Fasli, "Deep learning for non-stationary multivariate time series forecasting," in 2019 IEEE International Conference on Big Data(Big Data), 2019, pp. 2097–2106. doi: 10.1109/BigData47090.2019.9006192.

[6] S. Tiwari, A. Bharadwaj, and S. Gupta, "Stock price prediction using data analytics," in 2017 International Conference on Advances in Computing, Communication and Control (ICAC3), 2017, pp. 1–5. doi: 10.1109/ICAC3.2017.8318783.

[7] F. Wang, M. Li, Y. Mei, and W. Li, "Time series data mining: A case study with big data analytics approach," IEEE Access, vol. 8, pp. 14 322–14 328, 2020. doi: 10.1109/ACCESS.2020.2966553.

[8] Y. Yujun, Y. Yimei, and L. Jianping, "Research on financial time series forecasting based on svm," in 2016 13th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), 2016, pp. 346–349. doi: 10.1109/ICCWAMTIP.2016.8079870.