Stock Market Prediction for Time-series Forecasting using Prophet upon ARIMA

CH.RAGA MADHURI
Assistant Professor
Dept Of. Computer Science and Engineering
Velagapudi Ramakrishna Siddhartha Engineering College
Vijayawada
chragamadhuri@vrsiddhartha.ac.in

MUKESH CHINTA
Assistant Professor
Dept Of. Computer Science and Engineering
Velagapudi Ramakrishna Siddhartha Engineering College
Vijayawada
mukesh.chinta@vrsiddhartha.ac.in

V V N V PHANI KUMAR
Assistant Professor
Dept Of. Computer Science and Engineering
Velagapudi Ramakrishna Siddhartha Engineering College
Vijayawada
phanikumar.venna@gmail.com

***Abstract* - Since the beginning, the fundamental goal of man is to make life easy to live. The whole world believes that wealth would make life comfortable and luxurious. One of the most common notion among humans is that one of the best way to make money is to invest in stock markets which are expected to have tremendous results. There is a requirement to develop an intelligent system to perform predictions based on various indicators like fundamental, statistical and technical trends. However, there is no one good predictive model that has been successful to beat the trends in market continuously. Traditionally for time series data, the predictions are in general performed based on past historical data and market trends, historical correlation data and projections can be calculated. Above all said, there is no such system that calculates the predictions based on users selection on investment type and on risk criteria user is willing to take. So in this paper, we tried to demonstrate the technique(s) to get most accurate results.**

Keywords — Statistical indicators, Stock market, Time series forecasting, Future projections, Recurrence relation, Predict.

I. INTRODUCTION

Forecasting is one of the crucial tasks performed by data science which has been of prime concern to perform numerous works in any organization(s). For Example, gaint companies like Facebook has to perform capacity planning to allocate scarce resources efficiently and set their goals to measure the performance to the base line. System producing highly efficient forecasts is not an easy job for either a machine(s) or for most Data analysts. We have studied two major themes in order to perform business trend forecasts: automatic forecasting methodologies can be completely brittle and they are often too difficult to incorporate some useful important heuristics or assumptions. And another important methodology is the analysts who can provide with high quality results on forecasts which are quite rare because forecasting is a specialized skill which requires enormous expertise in data science.

One cannot know where to invest and how much to invest because he/she doesn't know whether there will profit or loss in their investment. This presents an interesting issue since many people in general eventually invest in any of the stock market sectors. The solution to this problem allows us, to know more about stock market options and even helps in making much more accurate decisions.

In the paper, here we tried to perform and demonstrate the efficient possible technique which can provide an appropriate solution to our problem.

This model will project the returns based on the user's selection of types of investments and the risk he/she is willing to take. Historic correlation data and projections are calculated. User will be able to change all the parameters that affect the final outcome and see how interest rate and inflation rate affects the final return. There is no such model which calculates the projections of all these following sectors:

- Fixed Deposit/Recurring Deposit/PPF-Public Provident Fund/ SSY- Sukanya Samriddhi Yojana/Senior Citizens Savings Scheme
- Commodities: Gold /Silver/Crude Oil Nifty 50 shares
- Returns are projected in the form of a graph. Updates are given for every 15 days according to the current interest rate and inflation rate.

The brief overview of above all the references that were taken are below: [1] discusses mainly on ARIMA (Auto Regressive Integrated Moving Average) model for stock market Prediction and to forecast the future values of a time series data. In [2] stock market predictions are tackled using both technical and fundamental analysis, which are combined together through data science. [3] Projected an example of best fitting time series models for volatility and also attempted to achieve equal error variances before performing modeling. [5] gave an efficient Data Mining (DM) Model for temporal data of Indian Stock Market whereas [4] gave the forecast on the consumer price index (CPI) in Belgium. [6],[7] gave the stock market predictions using data mining techniques.[8],[9] gave the stock market predictions using Machine Learning techniques. Data Flow Diagram describes the processes that are involved in the system for transferring data from the input to the file storage and result generation. We obtain the required results by correlating the historical data at the initial stage and then the data pre-processed and cleaned to obtain the required data. Further algorithms or operations are performed by installing the packages and then the rules are applied on the data, predictions are made and the resulting graph is the required output.

978-1-7281-7223-1/20/$31.00 ©2020 IEEE

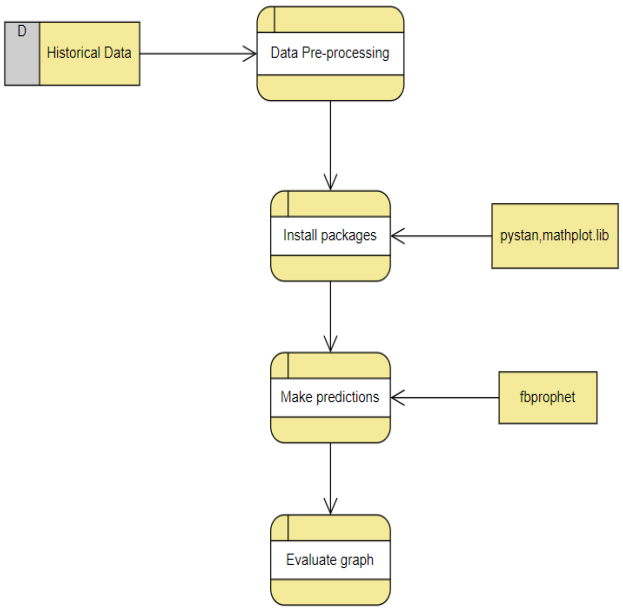

Fig. 1. Data Flow Diagram

Now, we install PyStan, Prophet and import prophet library into jupyter Notebook Ipython. The work is in general implemented using Jupyter, which is a part of software called IPython Project. The above Figure 1: Data Flow Diagram here is used to resemble the data flow and its processing.

II. METHODOLOGY

As one of the least complicated languages and simple syntaxes, we chose to use Python as our programming language. Though number of statistical programming languages like R provides many automated ways to give a solution to the problem, but they are yet to be officially ported over to program in Python. Pandas library is imported and is used to perform data analytics with back-end source written in python. Here, the data is a python dictionary with a key, value pair. From [1] it is observed that usually ARIMA is best used for prediction but it is not best suited for non-linear data patterns. Amazingly, the core main Data Science team of Facebook company published a interesting new method recently called by name Prophet, enables data developers and analysts to test or perform forecasting in Python at scale. Prophet is actually a procedure to perform forecasting of time series data mainly based on additive model unlike non-linear trends that in general works with daily, weekly, and yearly as well for seasonality, plus holidays. It gives best results for time series which has several season(s) of historical data and strong seasonal effects.

The input to prophet is always a dataframe with two columns. The major important requirement that Prophet has was pystan, which have its own installation procedures. We need to install pystan with command pip before making use of pip to perform installation of fbprophet. PyStan provides an interface for statistican modeling and high-performance statistical computation.

Prophet is actually an additive model which has the following :

$$m(t) = n(t) + o(t) + p(t) + \epsilon$$

- $n(t)$ models the trend(s), which projects long-term decrease or increase in data. Prophet has two trend related model(s), a piecewise linear model and a saturating growth model, which depends on the type of forecast.
- $o(t)$ models Fourier series with seasonality, that specifies how the data will be affected because of season related factor(s) like the time of the year
- $p(t)$ models the holidays effect or large events which highly impacts business time series data(e.g. Black Friday, New Product Launch, Superbowl etc.)
- ϵ represents an error term which is irreducible

III. IMPLEMENTATION

PRE-REQUISITES

- Time series analysis can be performed on either a local laptop/desktop or a server(s).
- Working on large dataset(s) be in general is memory intensive and will require a pc with atleast 2GB of main memory to perform calculations.
- For performing the above task, we will use Jupyter Notebook. If the software is already not installed, you need to install and perform setting up of Jupyter Notebook for Python 3.

Step 1 —Install Packages and Pull Dataset

For stock market analysis, we directly imported the dataset from Zen securities Pvt Ltd, a privately held company. Historical data of 37 years is collected and imported into a Excel sheet consisting of 2048 records.

Step 2 — Import Packages and Load Data

INSTALLING PYSTAN

Before installing and using fbprophet we need to install PyStan. PyStan and the required packages are installed using pip by the following command.

pip install pystan

INSTALLING PROPHET

To get started with Prophet, we installed it using the following command:

pip install fbprophet

LOADING CSV DATA IN PYTHON WITH PYSTAN

Pandas is one of the powerful data analysis Python library which was built on the top of *numpy* which was yet another important library that let you create 2d array and 3d arrays of in Python. The pandas library

978-1-7281-7223-1/20/$31.00 ©2020 IEEE

main object called as *dataframe*. Basically dataframe is a 2d numpy array which has rows and column(s), that even has label(s) for column(s) and rows. We create the dataframes for the input data format CSV. It should look like Figure 3 & Figure 4

	ds	y
0	11/8/2008	2645
1	12/8/2008	2011
2	1/9/2009	2142
3	2/9/2009	2059
4	3/9/2009	2392
5	4/9/2009	2518
6	5/9/2009	2822
7	6/9/2009	3304
8	7/9/2009	3137
9	8/9/2009	3461
10	9/9/2009	3311
11	10/9/2009	3461
12	11/9/2009	3611
13	12/9/2009	3492
14	1/10/2010	3542
15	2/10/2010	3463
16	3/10/2010	3608
17	4/10/2010	3746
18	5/10/2010	3461
19	6/10/2010	3480
20	7/10/2010	3496

Fig. 2. Sample dataframe format of PPF sector

In Figures 2 & 3, 'y' is the price of the particular sector on that particular day and 'ds' indicates the day 'y' relates to. Before we start doing any type of analysis with the data, we have to log transform 'y' variable and try to convert non-stationary data to a stationary one. It even tries to convert trend(s) to much more linear trend(s). This wasn't a perfect or a good way to deal with time-series data, but it generally works often can be tried first without worry.

	ds	y
0	31-03-81	1085
1	31-03-82	1177
2	31-03-83	1277
3	31-03-84	1392
4	31-03-85	1525
5	31-03-86	1677
6	31-03-87	1878
7	31-03-88	2104
8	31-03-89	2356
9	31-03-90	2639
10	31-03-91	2955
11	31-03-92	3310
12	31-03-93	3707
13	31-03-94	4152
14	31-03-95	4650
15	31-03-96	5208
16	31-03-97	5833
17	31-03-98	6533
18	31-03-99	7317
19	31-03-00	8195
20	31-03-01	8974

Fig. 3. Sample dataframe format of Crude oil sector

Step 3 — Time Series Forecasting with Prophet

MODELING:

The data set will be first split into 2 sets namely training and test data sets. The training set contains stock market prices from 1981–03–31 to 1991–03–31 while the test set contains prices from 1991–04–01 to 2001–03–31. Here we try to use training data to predict the next 10 years prices.

Modeling can be done by:

Instantiating the prophet to the model.

Fit the model with the data frame

Basic Syntax:

model=Prophet(intervalwidth=0.95) whichis 80% by default

'interval_width' sets the interval of uncertainty to get a confidence interval round the forecast.The following figure shows the sample code for modeling the data

```
from fbprophet import Prophet
import pandas as pd
import matplotlib.pyplot as plt
from matplotlib.backends.backend_pdf import PdfPages
```

```
gold_ds = pd.read_csv("D:\PythonCode\modified.csv")
gold_model = Prophet(interval_width=0.95)
gold_model.fit(gold_ds)
```

Fig. 4. Sample code for modeling the data

FORECASTING:

With Prophet, some future time data has been built by using make_future_dataframe method with a parameter period. The value of period depends upon the number of days, months or years we predict the values. Forecasting is done using predict command as shown in the Figure 5.

We need to verify about how will the model(s) works on historical data. Just by applying Prophet the model doesn't do a good job on fitting the data and catch the seasonality. So, we need to apply certain business insights like seasonality realted trends and holiday event and their affects, it's easy to project those input information into the Prophet. We tried to apply: weekly_seasonality, weekly_seasonality and holidays(manually created) .

```
gold_forecast = gold_model.make_future_dataframe(periods=3652)
gold_forecast = gold_model.predict(gold_forecast)
```

```
import matplotlib.pyplot as plt
plt.figure(figsize=(24, 6))
gold_model.plot(gold_forecast, xlabel = 'Date', ylabel = 'Gold Price')
plt.title('Gold Pricing');
```

Fig. 5. Sample code for forecasting the data

IV. RESULT

The output is given in the form of a graph. A graph is plotted using matplotlib. It is a plotting library in python. It is used for converting numerical data into a graph. This is imported using the command:

import matplotlib.pyplot as plt

The graph is obtained by using forecast_data command

978-1-7281-7223-1/20/$31.00 ©2020 IEEE

As we have selected 10 traditional sectors, the output is generated for those sectors in the form of a graph. The seasonality and the trend components of a sector are determined using data_forecast command. The result of gold sector prediction is shown in Figure 6, the result of gold sector prediction in table format is shown in the Figure: 7. Similarly, the results of Crudeoil prediction and PPF sector predictions are shown in the Figure : 8,9 and the table format prediction of Crudeoil sector is shown in Figure:10

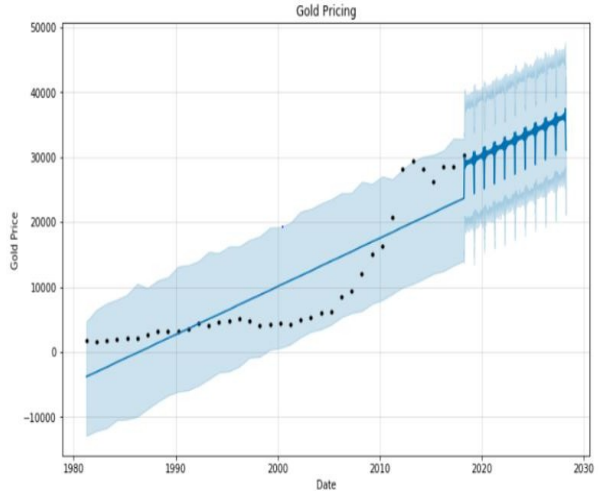

Fig. 6. Example output of gold sector prediction

	ds	trend	yhat_lower	yhat_upper	trend_lower	trend_upper	additive_terms	additive_terms_lower	additive_terms_upper	yearly
0	1981-03-31	1166.984388	-12925.472629	4668.952586	1166.984388	1166.984388	-4985.189953	-4985.189953	-4985.189953	-4985.189953
1	1982-03-31	1906.548263	-12188.177089	6494.640022	1906.548263	1906.548263	-5000.770175	-5000.770175	-5000.770175	-5000.770175
2	1983-03-31	2646.112130	-11744.147291	7470.798996	2646.112130	2646.112130	-5012.790375	-5012.790375	-5012.790375	-5012.790375
3	1984-03-31	3387.702205	-10540.793177	8065.729884	3387.702205	3387.702205	-4966.068665	-4966.068665	-4966.068665	-4966.068665
4	1985-03-31	4127.266078	-10419.784488	8821.395807	4127.266078	4127.266078	-4985.189953	-4985.189953	-4985.189953	-4985.189953
5	1986-03-31	4866.829951	-9978.653255	10494.195426	4866.829951	4866.829951	-5000.770175	-5000.770175	-5000.770175	-5000.770175
6	1987-03-31	5606.393822	-8811.535362	9856.191696	5606.393822	5606.393822	-5012.790375	-5012.790375	-5012.790375	-5012.790375
7	1988-03-31	6347.983891	-7693.523144	11018.256690	6347.983891	6347.983891	-4966.068665	-4966.068665	-4966.068665	-4966.068665
8	1989-03-31	7087.547765	-6731.610797	11513.433055	7087.547765	7087.547765	-4985.189953	-4985.189953	-4985.189953	-4985.189953
9	1990-03-31	7827.111655	-6116.556649	13179.930584	7827.111655	7827.111655	-5000.770175	-5000.770175	-5000.770175	-5000.770175
10	1991-03-31	8566.675546	-5910.101361	13369.200362	8566.675546	8566.675546	-5012.790375	-5012.790375	-5012.790375	-5012.790375
11	1992-03-31	9308.265634	-5197.168459	14010.984892	9308.265634	9308.265634	-4966.068665	-4966.068665	-4966.068665	-4966.068665
12	1993-03-31	10047.829520	-4209.202628	15413.833948	10047.829520	10047.829520	-4985.189953	-4985.189953	-4985.189953	-4985.189953
13	1994-03-31	10787.393404	-3179.881865	15262.118937	10787.393404	10787.393404	-5000.770175	-5000.770175	-5000.770175	-5000.770175
14	1995-03-31	11526.958492	-2984.300501	16253.740085	11526.958492	11526.958492	-5012.790375	-5012.790375	-5012.790375	-5012.790375
15	1996-03-31	12268.550142	-2246.263159	16276.555417	12268.550142	12268.550142	-4966.068665	-4966.068665	-4966.068665	-4966.068665
16	1997-03-31	13010.376070	-881.299109	17278.694818	13010.376070	13010.376070	-4985.189953	-4985.189953	-4985.189953	-4985.189953
17	1998-03-31	13754.195612	-772.500293	17809.349646	13754.195612	13754.195612	-5000.770175	-5000.770175	-5000.770175	-5000.770175
18	1999-03-31	14498.015840	331.815968	19155.575124	14498.015840	14498.015840	-5012.790375	-5012.790375	-5012.790375	-5012.790375
19	2000-03-31	15243.873931	601.446415	19248.667011	15243.873931	15243.873931	-4966.068665	-4966.068665	-4966.068665	-4966.068665
20	2001-03-31	15987.854595	1155.799704	19885.589104	15987.854595	15987.854595	-4985.189953	-4985.189953	-4985.189953	-4985.189953

Fig. 7. Example output of Gold sector in table format

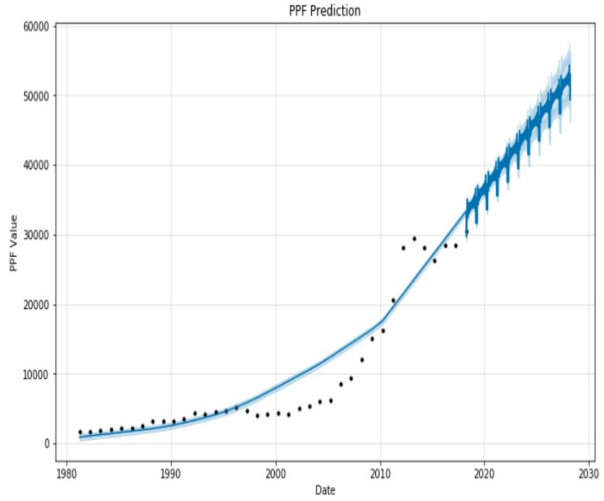

Fig. 8. Example output of PPF sector

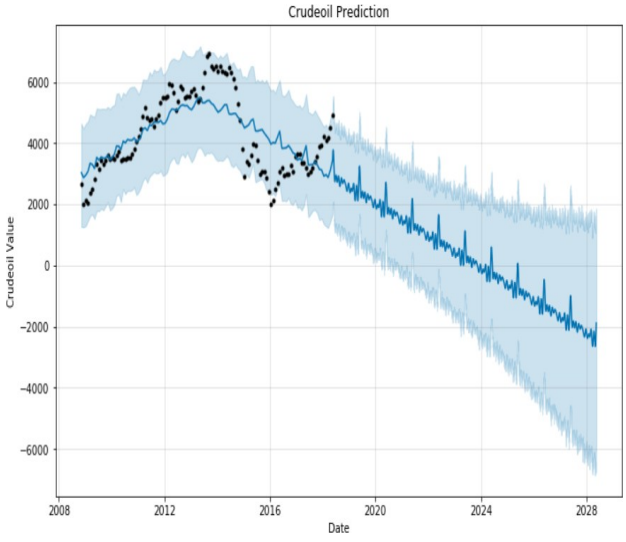

Fig. 9. Example output of Crudeoil sector prediction

	ds	trend	yhat_lower	yhat_upper	trend_lower	trend_upper	additive_terms	additive_terms_lower	additive_terms_upper	yearly	year
0	2008-11-08	3017.073909	1271.869067	4642.994499	3017.073909	3017.073909	16.809833	16.809833	16.809833	16.809833	1
1	2008-12-08	3064.222079	1260.027744	4476.563964	3064.222079	3064.222079	-188.711802	-188.711802	-188.711802	-188.711802	-18
2	2009-01-09	3114.513461	1296.191837	4635.733798	3114.513461	3114.513461	-155.799850	-155.799850	-155.799850	-155.799850	-15
3	2009-02-09	3163.233237	1464.540817	4715.702155	3163.233237	3163.233237	-80.869694	-80.869694	-80.869694	-80.869694	-8
4	2009-03-09	3207.238196	1618.439522	4939.370417	3207.238196	3207.238196	138.110093	138.110093	138.110093	138.110093	13
5	2009-04-09	3255.957970	1683.347409	4894.146471	3255.957970	3255.957970	43.101328	43.101328	43.101328	43.101328	4
6	2009-05-09	3303.106139	1523.610522	4855.149016	3303.106139	3303.106139	-129.801389	-129.801389	-129.801389	-129.801389	-12
7	2009-06-09	3351.825914	1794.915098	5142.373178	3351.825914	3351.825914	183.174100	183.174100	183.174100	183.174100	18
8	2009-07-09	3398.974083	1776.263041	5113.192552	3398.974083	3398.974083	49.229736	49.229736	49.229736	49.229736	4
9	2009-08-09	3447.693858	1864.242502	5117.959714	3447.693858	3447.693858	102.757835	102.757835	102.757835	102.757835	10
10	2009-09-09	3496.413633	1813.118971	5129.074182	3496.413633	3496.413633	9.728330	9.728330	9.728330	9.728330	
11	2009-10-09	3543.561802	1932.447094	5112.222148	3543.561802	3543.561802	-29.651358	-29.651358	-29.651358	-29.651358	-2
12	2009-11-09	3592.281577	2081.061416	5188.192674	3592.281577	3592.281577	11.505715	11.505715	11.505715	11.505715	1
13	2009-12-09	3639.429747	1807.256796	5119.552598	3639.429747	3639.429747	-179.384259	-179.384259	-179.384259	-179.384259	-17
14	2010-01-10	3689.721127	1869.486330	5212.616081	3689.721127	3689.721127	-167.029345	-167.029345	-167.029345	-167.029345	-16
15	2010-02-10	3738.440903	2004.502939	5251.261047	3738.440903	3738.440903	-87.303228	-87.303228	-87.303228	-87.303228	-8
16	2010-03-10	3782.445861	2283.792477	5615.560697	3782.445861	3782.445861	123.922363	123.922363	123.922363	123.922363	12
17	2010-04-10	3831.165636	2331.674095	5632.620432	3831.165636	3831.165636	68.561761	68.561761	68.561761	68.561761	6
18	2010-05-10	3878.313806	2075.829151	5425.571764	3878.313806	3878.313806	-86.340616	-86.340616	-86.340616	-86.340616	-9
19	2010-06-10	3927.033581	2506.758236	5854.521192	3927.033581	3927.033581	137.689423	137.689423	137.689423	137.689423	13

Fig. 10. Example output of Crudeoil sector in table format

978-1-7281-7223-1/20/$31.00 ©2020 IEEE

a) ADVANTAGES

One doesn't give a vast prior experience or knowledge on forecasting of time series data because it automatically will find seasonal related trends beneath data and will even offer a set of parameter(s) which are easy to understand.

It will allow even non statisticians to work using this and get good reasonable results which were often equal or sometime(s) even much better than the one projected by experts.

b) DISADVANTAGES

- Organizations should never rely 100 percent on any forecasting models
- It is not possible to accurately forecast the future
- Data may be flawed

V. CONCLUSION AND FUTURE WORK

In general, stock market data(s) are time-variant and are nonlinear in pattern, predicting the future price of stock is a challenging task. These types of Predictions will provide the users with good information about the current running status of stock price. With some historical data, you can use forecasting tools to predict into the future a specific metric. Prophet is specially designed to analyze time series data on daily observation(s) that will display the patterns on different types of time scale(s).It can be good tool which can be utilized in taking decisions for customers to finalize about either to buy or sell particular shares of a stock. Many researchers are been carried out for predictions on stock market prices with various data mining (dm) technique(s). The previous data of the stock selected will in general be used for building models and training them. The results of the model(s) are required to perform comparison with real time data to test the accuracy of the model.

Future work can be considered as using traditional time series analysis to perform forecast price of stock markets. It is an area of continuous research as investors and researchers strive to work with the market with an ultimate reason of acquiring higher returns. It is so unlikely that the new theoretical results will come out with the above projected works.

REFERENCES

[1] Idrees, Sheikh Mohammad, M. Afshar Alam, and Parul Agarwal. "A Prediction Approach for Stock Market Volatility Based on Time Series Data." IEEE Access 7 (2019): 17287-17298.

[2] Picasso, Andrea, et al. "Technical Analysis and Sentiment Embeddings for Market Trend Prediction." Expert Systems with Applications (2019).

[3] Somarajan, Siddarth, et al. "Modelling and Analysis of Volatility in Time Series Data." Soft Computing and Signal Processing. Springer,Singapore,2019.609-618.

[4] Nyoni, Thabani. "Time series modeling and forecasting of the consumer price index in Belgium." (2019).

[5] Abraham, Cerene Mariam, M. Sudheep Elayidom, and T. Santhanakrishnan. "Analysis and Design of an Efficient Temporal Data Mining Model for the Indian Stock Market." Emerging Technologies in Data Mining and Information Security. Springer, Singapore, 2019. 615-628.

[6] Kannan, K. Senthamarai, et al. "Financial stock market forecast using data mining techniques." Proceedings of the International Multiconference of Engineers and computer scientists. Vol. 1. 2010.

[7] Khedr, Ayman E., and Nagwa Yaseen. "Predicting stock market behavior using data mining technique and news sentiment analysis." International Journal of Intelligent Systems and Applications 9.7 (2017): 22.

[8] Shen, Shunrong, Haomiao Jiang, and Tongda Zhang. "Stock market forecasting using machine learning algorithms." Department of Electrical Engineering, Stanford University, Stanford, CA (2012): 1- 5.

[9] Choudhry, Rohit, and Kumkum Garg. "A hybrid machine learning system for stock market forecasting." World Academy of Science, Engineering and Technology 39.3 (2008): 315-318.

[10] Chong, Eunsuk, Chulwoo Han, and Frank C. Park. "Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies." Expert Systems with Applications 83 (2017): 187-205.

[11] Ologunde, Adedoyin O., David O. Elumilade, and T. O. Asaolu. "Stock Market Capitalisation and Interest Rate in Nigeria: A Time Series Analysis." Economic and Policy Review 13.2 (2007).

[12] Hussin, Mohd Yahya Mohd, et al. "Macroeconomic variables and Malaysian Islamic stock market: a time series analysis." Journal of Business Studies Quarterly 3.4 (2012):1.

[13] Deb, S. G., & Mukherjee, J. (2008). Does stock market development cause economic growth? A time series analysis for Indian economy. International Research Journal of Finance and Economics, 21(3), 142-149.

[14] Roberts, H. V. (1959). Stock-market" patterns" and financial analysis: methodological suggestions. The Journal of Finance, 14(1), 1-10.

[15] Rua, A., & Nunes, L. C. (2009). International comovement of stock market returns: A wavelet analysis. Journal of Empirical Finance, 16(4), 632-639.

978-1-7281-7223-1/20/$31.00 ©2020 IEEE