S. P. MANDALI's

PRIN. L. N. WELINGKAR INSTITUTE OF MANAGEMENT DEVELOPMENT & RESEARCH

SUMMER INTERNSHIP RESEARCH PROJECT

A report on

Stock Market Prediction and it's recovery during COVID-19

By

Sayantan Jana

PGDM - RESEARCH & BUSINESS ANALYTICS (2019 – 21)

Specialisation: Finance

Roll No: 42

PROJECT FACULTY GUIDE

PROF. DR. SONAL DAULATLKAR

Letter of Authorization

CERTIFICATE

<u>TO WHOSOEVER IT MAY CONCERN</u>

1. This is to certify that Mr. Sayantan Jana, a student of S.P. Mandali's Prin. L. N. Welingkar Institute of Management Development & Research and pursuing two years full time Post Graduate Diploma in Management – (PGDM- RBA), underwent two months of SIRP with me from 27/04/2020 to 20/07/2020.

2. During the SIRP, Mr. Sayantan Jana has successfully completed the project titled "Stock Market Prediction & its recovery during COVID-19" under my guidance of Prof. Dr. Sonal Daulatkar, Assistant Professor.

3. The students' performance during the SIRP and comments on his projects are as under:

…………………………………………………………………..

(Signature of Internal faculty member)

Name: Prof. Dr. Sonal Daulatkar

Designation: Associate Professor

Date: 20/07/2020

NO PLAGIARISM DECLARATION BY THE STUDENT

I, the undersigned, hereby declare that the project titled “Stock market prediction and its recovery during COVID-19”

a) Has been prepared by me towards the partial fulfilment for the award of Post Graduation Diploma in Management – Research & Business Analytics under the guidance of Prof. Dr. Sonal Daulatkar, S.P. Mandali’s Prin. L.N. Welingkar Institute of Management Development & Research, Mumbai.

b) This work is original and has not been submitted for any degree/diploma in this or any other Institute/Organization.

c) The information furnished is this dissertation is genuine and original to the best of my knowledge and belief.

d) I have not indulged in plagiarism. The project report has been checked plagiarism and output report has been attached.

Sayantan Jana

Sayantan Jana — Signature

Place: Mumbai — Date: 20/07/2020

PLAGIARISM CHECK OUTPUT REPORT

URKUND

Urkund Analysis Result

Analysed Document: Sayantan Jana_RBA_2019-21.pdf (D76281732)
Submitted: 7/12/2020 12:37:00 PM
Submitted By: sjana.chan@gmail.com
Significance: 2 %

Sources included in the report:

https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Exponential_Smoothing-Trend.pdf
https://core.ac.uk/download/pdf/39667613.pdf
https://towardsdatascience.com/stock-market-forecasting-using-time-series-c3d21f2dd37f
https://s3-ap-southeast-1.amazonaws.com/gtusitecirculars/uploads/02%20Thesis%20119997107006_931658.pdf

Instances where selected sources appear:

7

ACKNOWLEDGEMENT

The successful execution of this project was possible due to tremendous research, extensive work and dedication towards the project. This wouldn't have been accomplished without the rigorous support received from peers & surroundings directly or indirectly. I would like to exhibit a heartfelt thanking note and appreciation to all of them for their support.

I firstly thank my parents for all the sustenance and incitement they have shown towards me. Any project is incomplete without the support of one's family.

I owe my deep gratitude to my mentor Prof. Dr. Sonal Daulatkar for mentoring me and guiding me. She has been my mentor for the project enriching me with very useful insights and suggestions despite the busy schedule. I would also like to thank all the Research and Business Analytics faculty members for their guidance.

Also thanking the Dean of our department, Prof. C. Y. Nimkar for all the insights and guidance.

Lastly, a thanking note to all my family members, friends, individuals around & mother nature for the support.

Thanking all of them,

Sayantan Jana

Table of Contents

1. List of Tables

2. List of Graphs

3. List of Appendices

4. Executive Summary

The aim of this research is to analyse Indian stock market index through technical analysis. The two major and popular index SENSEX and NIFTY50 are taken into this report. The study aims at extracting the price movements of the stock market indexes just before the COVID-19 pandemic and during the pandemic over a significant period of time. These prices are analysed using some technical tools and indicators.

4.1) Major Findings:

1) The historical data has a significant role in helping investors to get an overview about the market behaviour.
2) Stock market went through a major trend shift during the COVID-19 pandemic.

4.2) Conclusion:

1) The seasonal trend and steady flow of any index will help both the existing and naïve investors to understand and make a decision to invest in stock market.
2) The trend chart may forecast on particular aspect but is not adequate for long term decision making.
3) Thus, the main target is not to govern whether to buy or sell a particular stock but to give a holistic view of the stock market by showing past trend, seasonal flow and variations of the stock.

4.3) Recommendations:

1) A broad research paper can be made including several other sector index and strategic index that were not taken into this study.
2) COVID-19 pandemic situation is expanding day by day. To get a trustworthy & convincing result dataset should be updated with the very recent data points.
3) A limited number of technical tools are used for analysis. More high-level machine learning algorithms can be used to strengthen the predictive model.

5. Introduction

The stock market is a general term which refers to the assembly of markets in which equities, bonds and other types of securities are issued and traded through various physical and electronic exchanges and through the counter market. They are the most important components of a market economy [1], because they provide companies with access to capital by enabling investors to purchase shareholdings in a company. The stock marketplace is continually changing as a result of the refining process.

5.1) Background to the problem:

Every stock exchange has its own stock index principles. Index is the average value that is calculated by combining a number of stocks. This helps to represent the entire stock market and predict market movements and stock prices [2] over time. The stock market can have a major effect on people and the economy of the country as a whole. As researchers have discussed in S.J. Grossmara et al., when new information arrives stock market value varies [3]. Therefore, forecasting stock movements in an efficient manner will reduce the risk of investing and increase profits.

SENSEX:

BSE SENSEX (S&P Bombay Stock Exchange Sensitive Index) is a free float market weighted stock market index of 30 well-established and financially good companies enumerated on the Bombay Stock Exchange. The 30 component firms, some of the largest and most actively traded stocks, are indicative of the different industries of the Indian economy. Sensex is a measure of the stock market index for BSE. This was released for the first time in 1986.

How SENSEX is calculated:

The SENSEX measurement is based on a free-float method that came into existence on 1 September 2003. The amount of SENSEX is a clear indicator of the market performance of 30 stocks. The free-float approach takes into account the proportion of securities that can be freely exchanged on the market. This does not include those owned by specific owners and promoters or other locked-in securities not available on the market.

Next, market capitalization is taken into account. It is achieved by averaging all the shares issued by the company at the price of its stock. BSE then calculates the free-float element, which is a multiple of the company's market capitalisation. It helps to assess the free-float market capitalization on the basis of the data provided by the client. The ratio and the proportion are then used for the base index of 100. This allows the SENSEX to be decided.

NIFTY:

The NIFTY50 index is India's benchmark Domestic stock market index for the Indian equity industry. Nifty is owned and managed by India Index Services & Products (IISL). NIFTY50 includes 13 sectors of the Indian economy and provides portfolio exposure to the Indian market to investment managers. NIFTY is the NSE sector tracker. Ideally, there is a set of 50 stocks, but at present there are 5 listed in it. It was also referred to by some as Nifty 50 and CNX Nifty as own.

How Nifty is calculated:

Nifty is also determined using a weighted form of free-float market capitalization. Like SENSEX, NIFTY also follows a mathematical formula dependent on knowledge of market capitalization. It multiplies the capital stock at a price to excerpt the capitalization of the market [4]. To order to determine free float market capitalisation, equity capital is multiplied by a price that is further compounded by the IMF, which is a factor to deciding the number of shares available for free trade on the market. The index is calculated on a regular basis by taking into account the current market value divided by the initial market capital and then multiplied by the initial index value of 1000.

5.2) Statement of the problem:

Stock price prediction is very important, because it is used by most business people as well as common people. People would either earn money or lose their entire life savings in the operation of the stock market. It's a method of uncertainty. Creating an effective model is difficult, as price volatility depends on several variables, such as company production, government bonds, historical prices and the economy of the region.

There are two common methods of forecasting stock market prices. Each of these is graphist or scientific theory, and the other is fundamental or intrinsic value analysis. This research paper is focused on the concepts of scientific theories. The fundamental supposition of this theory is that history repeats itself. The purpose of this research work is to develop a model that predicts movements in stock patterns using historical data.

COVID-19 Impact: The Corona Virus has appeared to the world in the late Dec’19 and is declared as a global pandemic on 12th Mar’20. Throughout the world it affected the economy. Stock market in every country sees a rapid decline in prices as global lockdown takes place. In India the major stock indices as well as every individual stock underwent a fall down, same scenery as the world. Though the impact in India was delayed, the declination started soon after covid is declared as pandemic.

6. Literature Review

For the stock market outlook, *K. Senthamarai Kannan et al.* [5] used a number of data mining techniques. The best thing about data mining is historical data, which holds the key memory to predict the future course. Through historical stock market data, investors uncover the secret trend of data that have predictive potential in investment decisions using data mining techniques. Forecasting the stock market outlook in the financial time series is seen as a difficult task. By means of a data analysis, we can predict potential stock price rises or decreases. Data analysis is also one way of forecasting events. Five methods of stock analysis have been integrated to estimate this. The suggested five approaches are Standard Price (TP), Relative Strength Index (RSI), Bollinger Bands, Moving Average (MA) and CMI. Combining such strategies would be useful in forecasting days when the closing price would increase or decrease.

Phichhang Ou and Hengshan [6] Wang used ten different data mining methods to forecast the price changes of the Hong Kong Stock Exchange Hang Seng Index. Such ten approaches are Quadratic Discrimination Analysis (QDA), Linear Discrimination Analysis (LDA), Naive Bayes based on kernel estimation, K-nearest neighbour classification, neural network, Tree based classification, Support Vector Machine (SVM), Bayesian Gaussian process classification, Logic model and Least Square Vector Support Machine (LS-SVM). Between all of these approaches, LS-SVM and SVM generate superior predictive efficiency. In most cases, SVM is higher than LS-SVM in the sample prediction. But in term of hit rate and error rate criteria LS-SVM is in turn superior than SVM for out sample forecast.

Aditya Gupta and Bhuvan Dhingra [7] used Hidden Markov Model (HMM) to forecast the stock market. By using historical stock prices, he presents a Maximum Posteriori HMM approach to forecast stock values for the next day. For continuous HMM instruction, the intraday high and low stock values and the fractional shift in stock values are considered. This HMM is used to make a maximum posteriori judgment on all potential stock values for the next day. Using some of the existing methods, such as HMMs and Artificial Neural Networks, using Mean Absolute Percentage Error (MAPE). They test their approach on a number of stocks and compare their performance. Ultimately, an HMM based Maximum a Posteriori (MAP) estimator for stock predictions is provided. The layout is using a Latency of days to estimate the stock value of the (d+1) day. All possible stock values shall be determined by means of a previously qualified continuous HMM MAP judgment. They assume four underlying hidden states that emit visible observations (fractional shift, fractional high, fractional low).

George S. Atsalakis and Kimon P. Valavanis [8] forecast stock market movements in the short term using neurofuzzy-based methods. Essentially, the neuro-fuzzy system consists of the Adaptive Neuro Fuzzy Inference System (ANFIS) controller used to control the stock market process model. The variety of stocks is also obtained and evaluated using an adaptive neuro-fuzzy technique.
The results obtained question the poor form of the Efficient Market Hypothesis (EMH) by demonstrating much improved and better forecasts, relative to other methods, of short-term patterns in the stock market and, in particular, of the next day's trend in chosen stocks. The ANFIS controller and the market model inputs are taken on the basis of a comparative study of fifteen different combinations of historical stock prices performed to determine the stock market model inputs that return the best stock market trend forecast for the next day in terms of the minimum root mean square error (RMSE). Gaussian 2 shaped membership functions are preferred over Gaussian bell-shaped and triangular ones to fuzzify system inputs due to the lowest RMSE. Real case studies using data from developing and well-developed Athens and New York Stock Exchange (NYSE) stock markets to train and test the proposed framework show that, compared to the buy and hold strategy and several other recorded approaches, the proposed approach and trade accuracy forecasting are far superior.

Binoy B. Nair, N. Mohana Dharini, V.P. Mohandas [9] suggested a hybrid decisions tree-neuro-fuzzy method to forecast the stock market. Automated stock market trend prediction system is suggested through the use of a decision tree adaptive neuro-fuzzy hybrid system. They used a number of methods, such as theoretical analysis and decision tree. First technical analysis which is commonly used by stock traders for the extraction of features and second decision tree for the collection of features. Using the theoretical model and decision tree used for the reduced data collection, the adaptive neuro-fuzzy method is applied to the next day's stock prediction. They tested their proposed system on four major international stock market. They tested their proposed program on four major international stock markets. Their experimental findings show that the given hybrid system gives output of much higher accuracy compared to the stand-alone decision-based system and the ANFIS-based system without features of selection and dimensional reduction.

J. Gong and S. Son [10] have implemented stock prediction model using logistic regression considering feature index variables. They have mentioned that daily stock trading prediction with logistic regression out performs other methods such as RBF – ANN prediction model.

7. Research Design

Research design is a conceptual structure and research is conducted within its bounds. It consists of the methods for data collection, measurement and analysis for an exploratory research. Exploratory research involves gathering relevant data from various sources like research papers and internet. For present study, the research is analytical and conclusion oriented.

7.1) Objectives of the study:

- To study the significance of technical analysis in Indian Capital Market.
- To analyse the performance of two major stock indices using advanced charts and predict the future trend in share prices.

Need for the study:

- To help apprehend the behaviour of share price movements in the past using charts. To predict future prices and trend in the near short term and be able to choose appropriate stop loss and target prices.

7.2) Type of Study:

The aim of the study is to analyse the Indian stock market index and do a near future prediction of 60 days. So, the researcher has used exploratory research method by which the collected information could be analysed and concluded. The researcher has used positivism philosophy as this is a quantitative study. In this study, the researcher has adopted inductive approach to reach to a conclusion. Conclusive research design is used which focuses to develop a conclusion.

7.3) Data collection:

This research is based on the analysis of secondary data. So, the relevant stock indices data has been collected from Yahoo finance website and formatted to fit into a csv format. Then its processed [11] and made ready for python analysis.

7.4) Data Description:

Historical data of the indices has been collected from Yahoo Finance website. Date ranging from 1st Jan'20 to 6th Jun'20 has been considered as a well time framed dataset for this research. The columns in the data are Date, open, high, low, close & volume.

7.5) Statistical Techniques:

In this study the stock market indices data used is fed to the system in a csv format. The whole analysis is based on python programming language [12]. For visualization line plots, candlestick charts are used. Augmented Dickey-Fuller test and OLS regression is used as a statistical tool for verifying certain hypothesis. For prediction & forecasting i.e. model building, exponential smoothing methods are used. They are simple exponential smoothing, Holt's linear smoothing, Holt's damped trend & Holt-Winter's exponential smoothing.

8. Data Analysis

The researcher divided the analysis part into two segments i.e. exploratory analysis and predictive analytics.

8.1) Exploratory analysis:

To understand the data more deeply and getting a general understanding of it, the researcher used a number of descriptive and visualizations on the dataset. Alongside simple line plots trend breakdown and seasonality analysis are also done. Statistical test like Augmented Dickey-Fuller test is conducted to reach a conclusion whether the dataset is stationary or not.

1. **Line & Scatter plot**: Basic line plots and scatter plot are done in python using time in x-axis and close prices of every day on y-axis.
2. **Candlestick**: A popular chart type in stock market is candlestick chart. It has every day high low prices plotted with respect to time.
3. **Rolling mean**: It is also known as moving average. Different time span is taken to get the rolling mean as well as its standard deviation.
4. **Trend separation**: Trend in the data poses a problem because it de-stabilizes the predictive learning and takes the model prediction into a biased version of it. Separating the trend from the data results in a more accurate future prediction of prices. Also, it shows how much trend component is there in the dataset.
5. **Auto-correlation**: Auto-correlation and partial auto-correlation check is important since we don't want either of them to be present in the dataset.
6. **Statistical Test**: Only statistical test used here is the Augmented Dickey-Fuller test which takes hypothesis and concludes whether the data is stationary or not. Stationary data is recommended to do time series analysis on it.

8.2) Predictive Analysis:

The primary goal of using predictive analytics in this research is to know and determine a possible path of the close prices of the indices for a short span of time. Simple regression, trend regression and several exponential smoothing [14] methods are used to make four exponential models using the dataset. In only one of them a 60 days prediction curve has been plotted to get an idea of the future prices based on the model.

The working principal behind smoothing methods is weighted averages. Near future forecasts are weighted averages of past observations. Weights can be of different types like a moving average where its uniform. It also can follow an exponential decay where more weight is given to recent observations and less to old observations.

1. **Simple exponential smoothing**:

This type of smoothing is easy and used where the dataset is small, irregular, completely free of trend, seasonality and other recurring things. The researcher used this to get a general understanding of how the model performs in this type of dataset.

SES has only one component called level with a smoothing parameter denoted as ‘alpha’. Mathematically, it’s the weighted average of previous and current observation.

$$p_t = l_t$$

$$l_t = \alpha y_t + (1-\alpha)\, l_{(t-1)}$$

p = forecast

l = level

α = smoothing parameter

2. **Holt’s linear smoothing**:

It’s a popular smoothing technique for trend data forecasting. Holt's model has three different equations that works together to produce a final prediction. The first is a normal smoothing equation that adjusts directly the last value for the trend of the last period. The trend itself is updated over time by the second equation, where the trend is expressed as the difference between the last two smoothed values. The third equation is used to produce the final prediction. The Holt model uses two components, one is for complete smoothing and second is for the trend smoothing equation. The forecast component is made of a level component and trend component.

(1) Forecast $p_{t+h|t} = l_t + hb_t$

(2) $l_t = \alpha y_t + (1-\alpha)\, l_{(t-1)}$

(3) $b_t = \beta\, (lt - l_{t-1}) + (1-\beta)\, bt_{-1}$

hb_t = trend

β = smoothing slope

3. **Holt’s damped trend**:

There are certain drawbacks for Holt’s linear trend method like the trend is constant in future means it’s increasing or decreasing indefinitely. For long forecast horizons, this turns out to be problematic. The damped trend method adds a dampening parameter so that the trend converges to a constant value in future i.e. it flattens the trend. The equation is similar to Holt’s linear method just the parameter b s replaced by φb.
As this data contains a trend component, it’s also known as double exponential smoothing or damped exponential smoothing. The trend part is flatted over time

instead of being linear as the simple exponential smoothing is literally used twice. Double exponential smoothing measures a trend equation via data using a special weighting function that puts the most importance on the most recent time periods. The predicted equation varies from time to time. The smoothing constant, alpha, dictates the amount of smoothing that takes place.

4. **Holt-winter's seasonal smoothing**:

The Holt-Winters seasonal approach requires the prediction equation and three smoothing equations. There are two variations in this method that vary in the nature of the seasonal portion. The additive method is preferred when seasonal variations are approximately constant across the series, whereas the multiplication method is preferred when seasonal variations differ in proportion to the series level. In the additive process, the seasonal variable is explained in absolute terms in the size of the series observed and, in the level equation, the series is modified seasonally by omitting the seasonal component. With the multiplication method, the seasonal part is expressed in relative ways (percentages) and the series is adjusted on a seasonal basis by dividing the seasonal component. The seasonal part will add up to nearly m in total over the years.
So, the Holt-winter's is like the Holt's linear except a seasonal component is introduced. To introduce the seasonal component, seasonal decomposition is used where it states that any series can be decomposed in a sum of 3 components: a trend, a seasonal component and residuals.

y = Trend + Seasonal + Residuals (Additive Version)
y = Trend * Seasonal * Residuals (Multiplicative Version)

In this analysis the researcher has used the multiplicative seasonal decompose function.

9. Results

The researcher has plotted the historical close prices on a graph using python. Also, several other visualization & analysis have been done on the data which have been discussed below.

1. **Line plots**: This is a line plot simply done using the close prices over time.

Fig 1. NIFTY50 Close prices

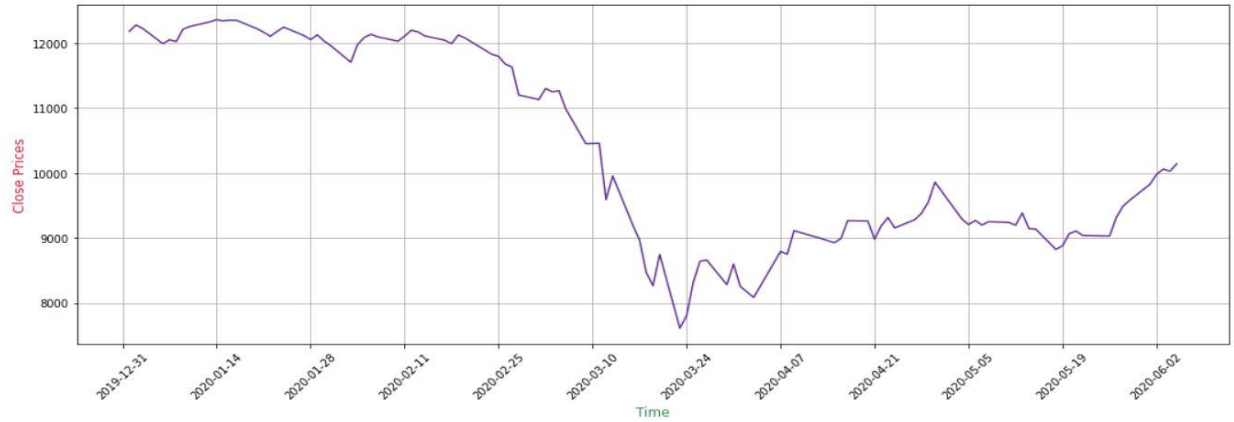
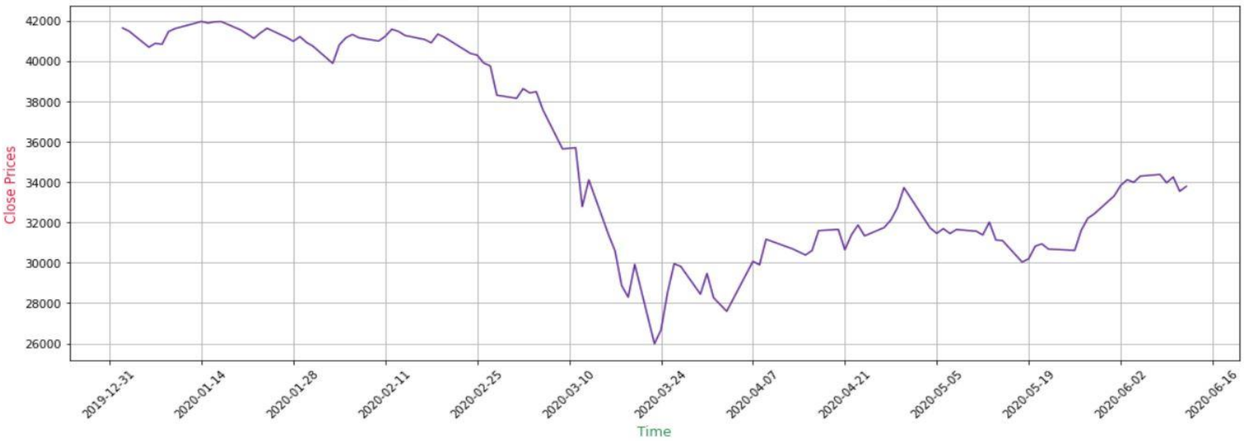


Fig 2. SENSEX Close prices



Both the indices experienced a massive downfall in prices starting from the last week of Feb'20 to late Mar'20. An obvious reason is that during these times the news of COVID-19 has spread all over the country. Many other countries at the same time were already hit by covid waves and were suffering a lot. The adverse effect in India started when it has been declared as a pandemic. Though since Apr'20 indices started to recover gradually. At the end of June, the indices recovered a lot.

2. **Candle stick charts**: As discussed before, candle stick charts feature a lot of details compared to line plots like the whether it's gone down or up showed by green or red. Also, longer bar means that day high & low price has a big difference.

Fig 3. NIFTY50 Candlestick

Fig 4. SENSEX Candlestick

There are a lot of small red bars in both the indices indicates every day the prices have gone down little bit but consistently. During the slide, the indices have gone down by up to 5% everyday.

3. **Moving average**: Moving average is a technical indicator used widely to smooth out price trends by filtering out 'noise' from random short-term fluctuations.

Fig 5. MA (NIFTY50)

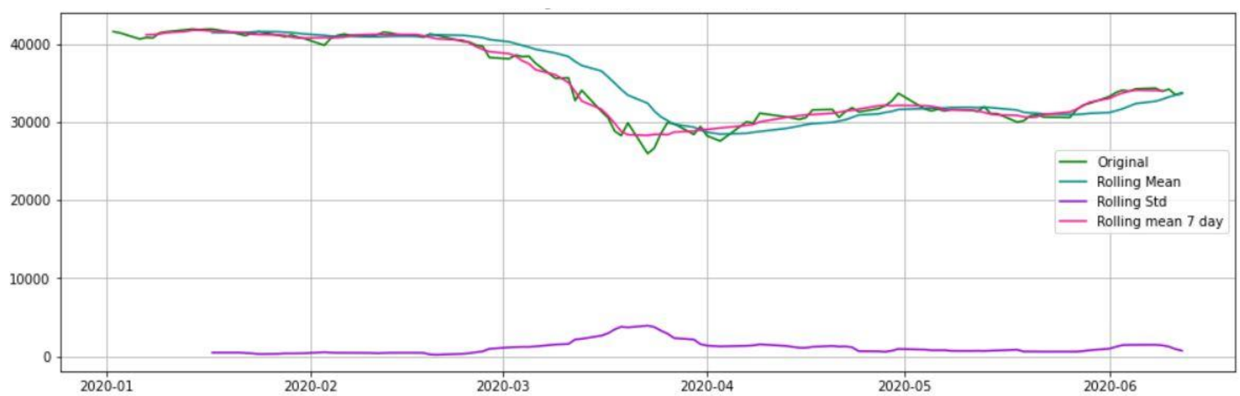
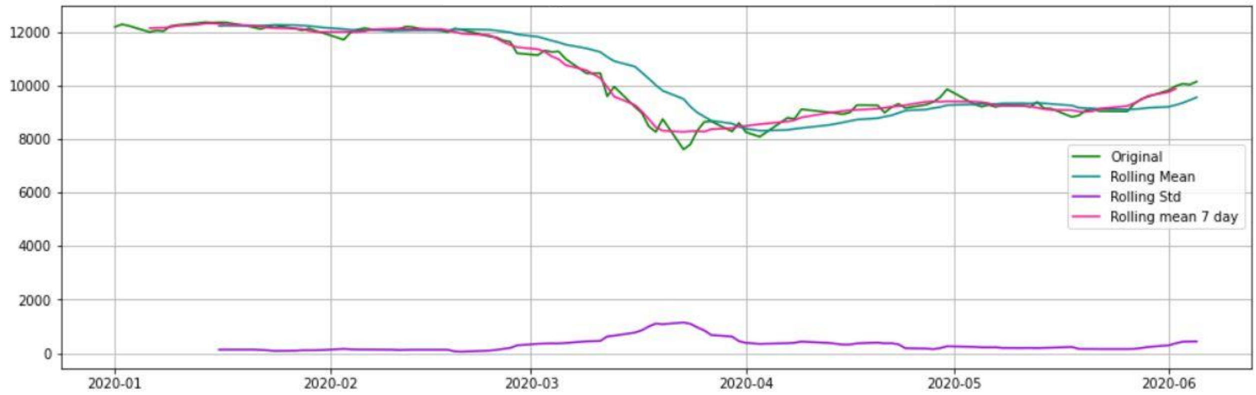


Fig 6. MA (SENSEX)

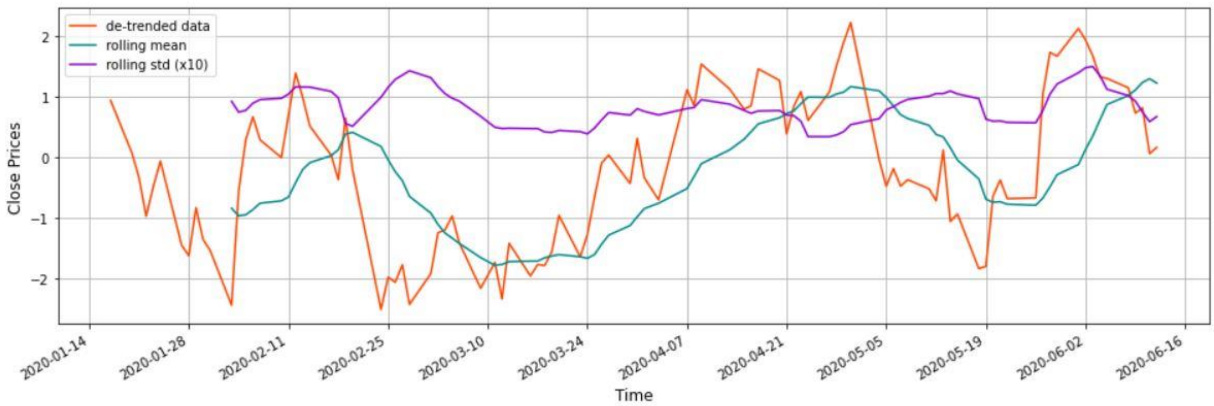


Moving average is also called rolling mean. The blue line indicates 30 day moving average where the pink line is a 7-day rolling mean. The implication of the moving average is that when the price is going over or under the moving average it is a signal for traders to buy or sell the stock.

4. **De-Trend**: It means the removal of trend component from the data to show only the absolute changes in prices and it helps to identify potential cyclical patterns.

Fig 7. De-Trended (NIFTY50)

Fig 8. De-Trended (SENSEX)



The trend also known as seasonality sometimes has to be removed to increase the degrees of freedom. The blue line shows the 30-day moving average and purple shows its corresponding standard deviation. The X-axis here is not the absolute close prices but a relative co-efficient to the actual close prices. Rolling mean of the de trended data further smoothers the fluctuations and provide a clear direction of upward or downward.

5. **Auto-correlation**: For a time series data auto-correlation is a thing that should not be present. It's a measure of correlation between the lagged values and the original values of a time series.

Fig 9. Autocorrelation (NIFTY50)

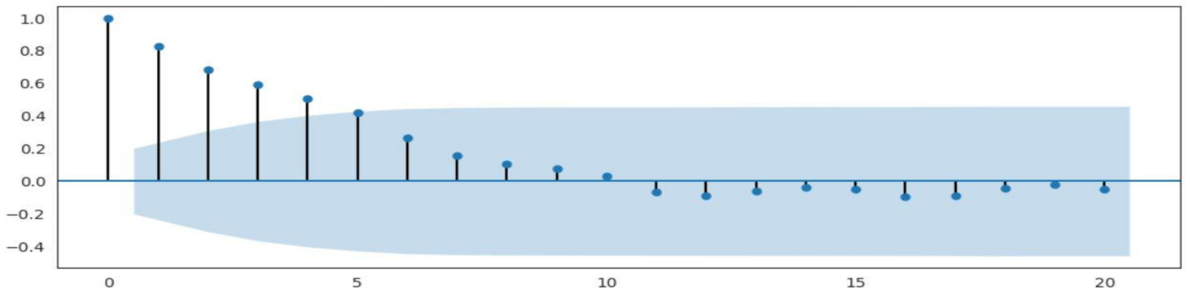
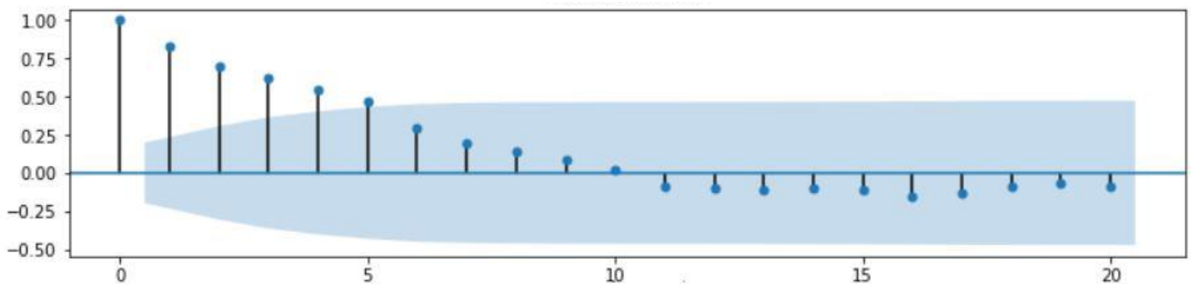


Fig 10. Autocorrelation (SENSEX)



In this case 20 day lagged values have been taken and checked the correlation with its own values. Both NIFTY50 & SENSEX shows a very low (near zero) auto-correlation which is a good sign. The blue shaded area indicates the ranges of permissible values. Means it's fine if those spikes are within the blue shade. Spikes over the limit values shows significant correlation with itself which may indicate a result of faulty data.

Fig 11. Partial Autocorrelation (NIFTY50)

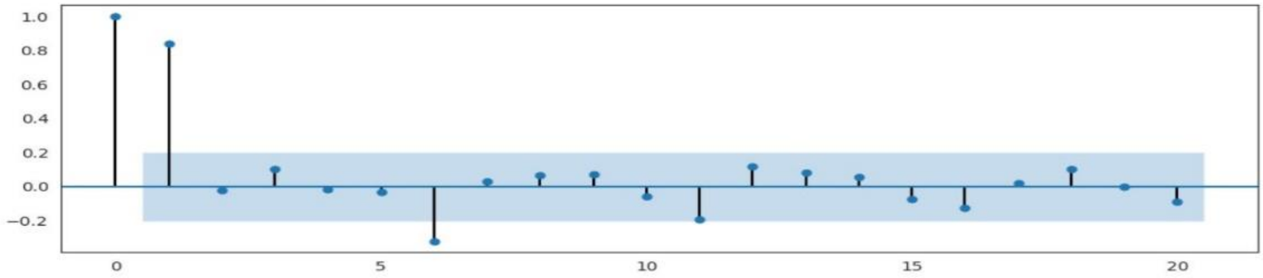
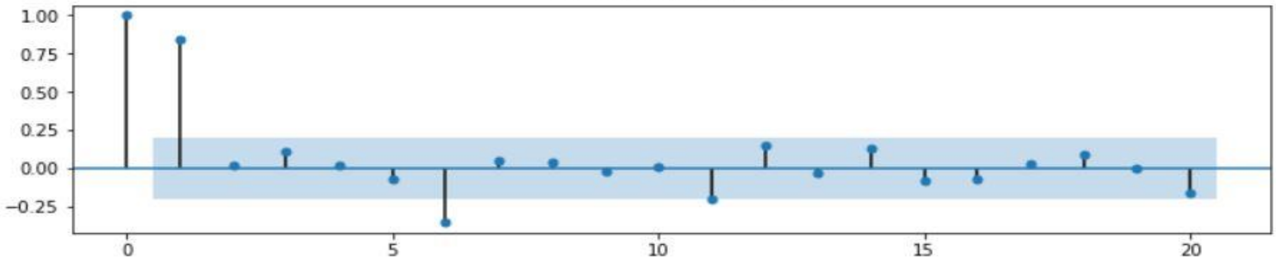


Fig 12. Partial Autocorrelation (SENSEX)



Unlike auto-correlation partial auto-correlation does control other lags. It measures the autocorrelation of the price and its lagged values separated by 'k' time units after removing any linear independence and also takes care of all the other terms of shorter lag. To pass the partial autocorrelation test the correlated values should be in the range of blue region which is exactly like in this case.

6. **Statistical test**: Before doing any analysis on a time series data, first the data is needed to be checked for stationarity. Here *Augmented Dickey-Fuller (ADF)* test is used to check the stationarity of the data.

a) Null hypothesis is that the time series is not stationary.

b) Alternate hypothesis is that time series is stationary.

If the absolute value of test statistic is less than critical value or p value is greater than the specified significance level, we fail to reject null hypothesis i.e. the data is not stationary.

Table 1. ADF test (NIFTY50)

```
 > Is the data stationary ?
Test statistic = -1.701
P-value = 0.430
Critical values :
        1%: -3.4996365338407074 - The data is not stationary with 99% confidence
        5%: -2.8918307730370025 - The data is not stationary with 95% confidence
        10%: -2.5829283377617176 - The data is not stationary with 90% confidence

 > Is the de-trended data stationary ?
Test statistic = -2.742
P-value = 0.067
Critical values :
        1%: -3.502704609582561 - The data is not stationary with 99% confidence
        5%: -2.8931578098779522 - The data is not stationary with 95% confidence
        10%: -2.583636712914788 - The data is  stationary with 90% confidence
```

Statistical test results show that the researcher fails to reject the null hypothesis. Means that the data is not stationary. The de-trended data is also not stationary.

Table 2. ADF Test (SENSEX)

```
 > Is the data stationary ?
Test statistic = -1.742
P-value = 0.410
Critical values :
        1%: -3.4968181663902103 - The data is not stationary with 99% confidence
        5%: -2.8906107514600103 - The data is not stationary with 95% confidence
        10%: -2.5822770483285953 - The data is not stationary with 90% confidence

 > Is the de-trended data stationary ?
Test statistic = -2.959
P-value = 0.039
Critical values :
        1%: -3.4996365338407074 - The data is not stationary with 99% confidence
        5%: -2.8918307730370025 - The data is  stationary with 95% confidence
        10%: -2.5829283377617176 - The data is  stationary with 90% confidence
```

The result for SENSEX data is indicating the non-stationarity of data in both the normal and de-trended case.

Predictive Analysis:

The historical data of the indices is taken in the time period from Jan'20 to Jun'20. For model training and testing the full data points divided in the proportion of 4:1 i.e. 80% for testing & 20% for training.

Fig 13. Train-Test Split

Statistical tests show that the data and de-trended processed data is not stationary. That makes the data incompatible for time series analysis. But luckily some exponential methods allow trend and seasonality component in the data. They perform quite well for non-stationary data. That's why the researcher used exponential methods for this type of dataset.

Simple Exponential Smoothing: This method is used where there are few input data points, features an irregularity and has no seasonality or trend. SES is not in any way useful or effective in time series which have trend & seasonal component. But it turns out effective for data sets where the variable in question does go up down very frequently and with a significant magnitude.

Fig 14. SES (NIFTY)

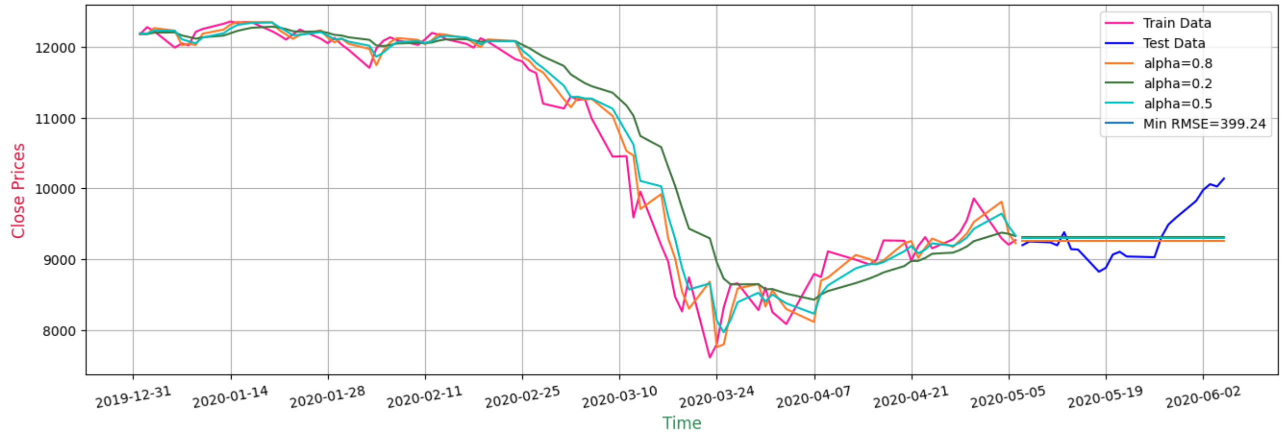
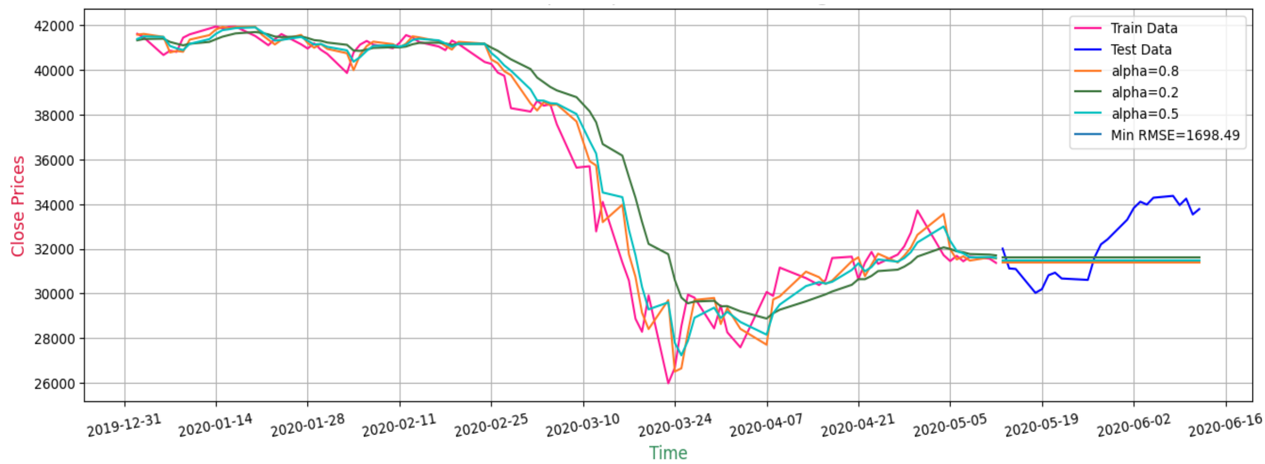


Fig 15. SES (SENSEX)



Thus, the RMSE of the model goes down by a good amount but this happened only for this unique set of data. It does not guarantee & has a very low chance to perform well compared to other higher-level methods. The forecast is generated with various levels of smoothing parameter (α) values. Then it's compared with the test values and RMSE is calculated for model performance. As expected, SES does not show a good prediction. The RMSE is also on the higher side. This is mainly because of the trend & seasonal component of the data which is unaccounted in the model.

Holt's Linear Smoothing: It considers the trend component in the data but no seasonal component. But in this method the trend component takes only one derived slope that goes on forever without changing.

Fig 16. Holt's Exp (NIFTY50)

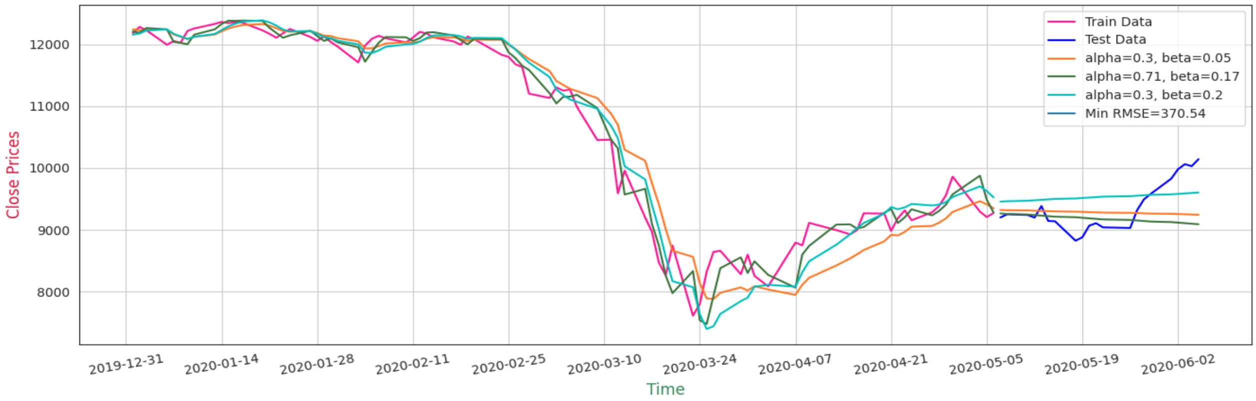
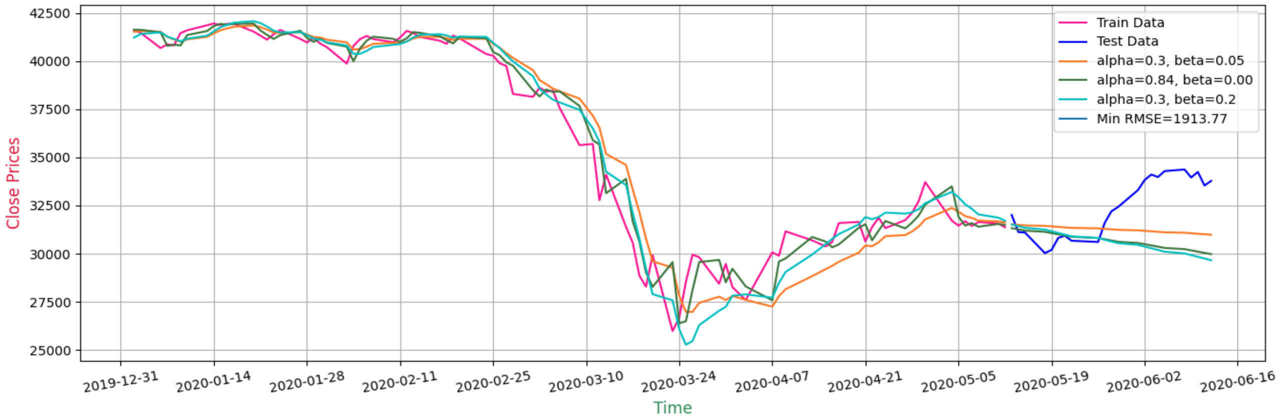


Fig 17. Holt's Exp (SENSEX)



As its seen that the prediction lines are very much straight lines with a constant slope at every point. This method is effective for a very short-term prediction and turns out badly for greater than 20 days prediction. Even with the various smoothing parameter values and introduction of smoothing slope the RMSE is hovering around the same values as previous. NIFTY50 data shows improvement compared to previous (RMSE) whereas SENSEX shows a poor RMSE.

Holt's Damped trend:

In this method a dampening parameter is also added to flatten the trend. From the past data trend is derived with a slope which is mentioned by beta parameter.

Fig 18. Holt's Damped trend (NIFTY50)

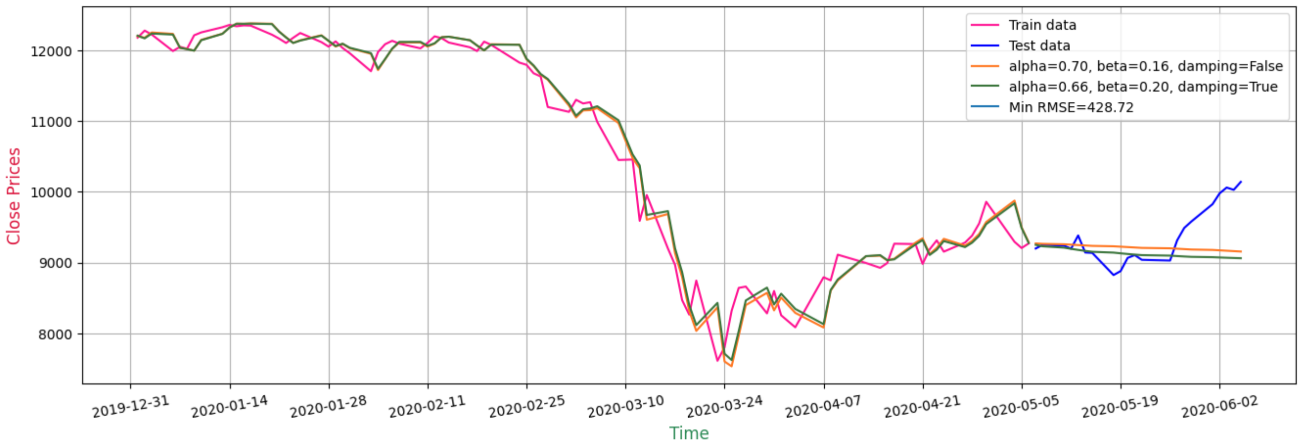
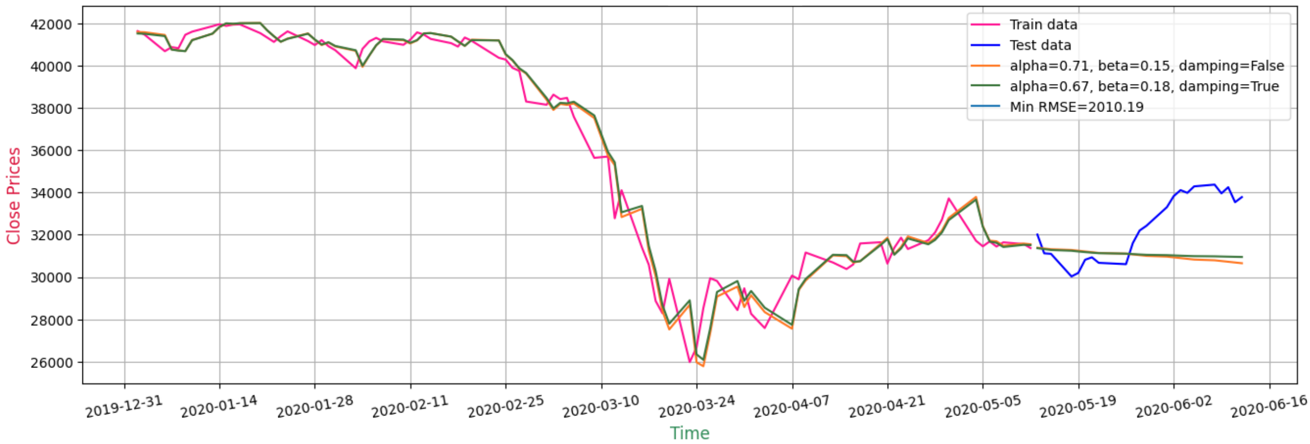


Fig 19. Holt's Damped trend (SENSEX)

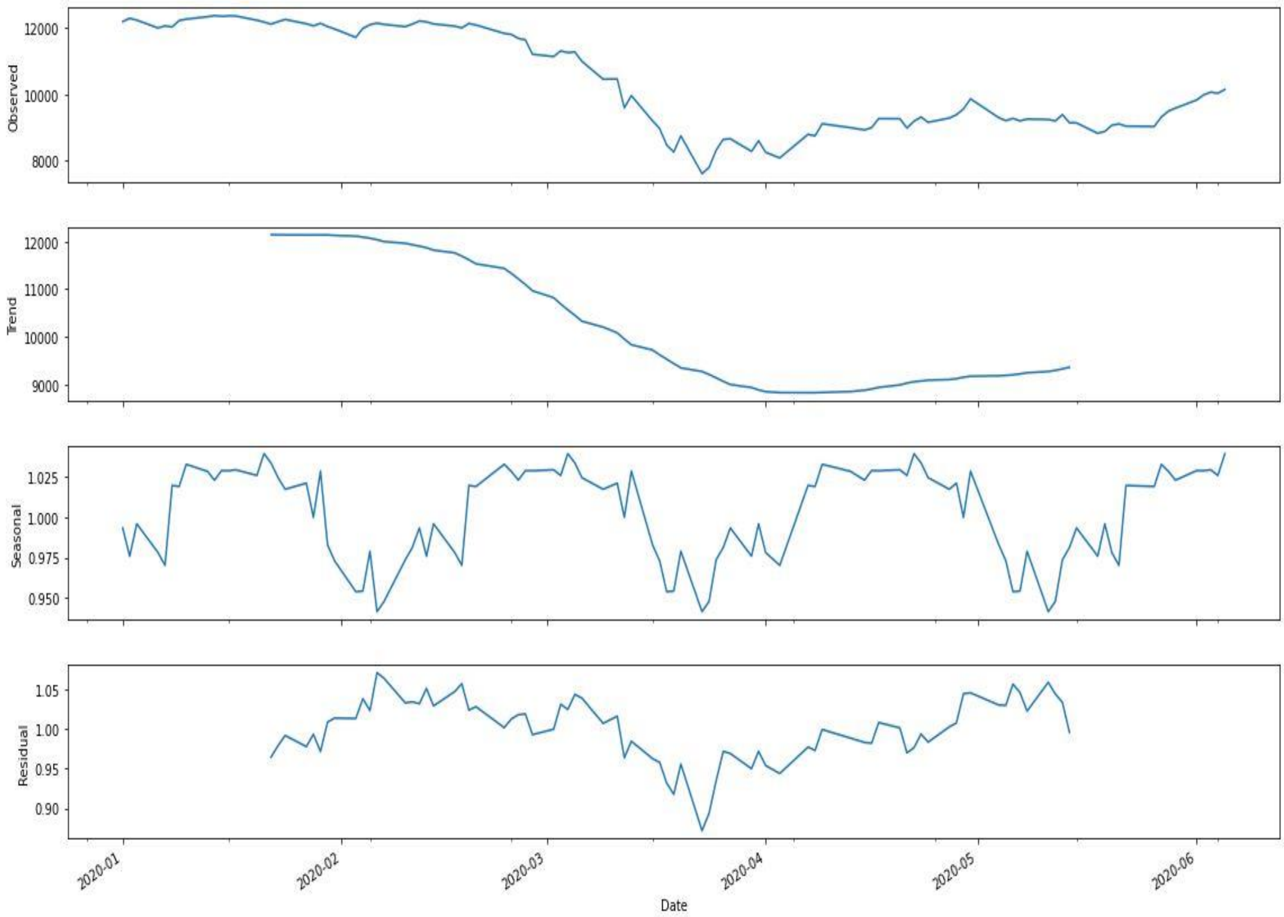


Flattening the trend with a damped parameter results a higher RMSE means further degradation of performance. Still the prediction curve shows no fluctuations at a glance. In both the case the RMSE increased by a little bit which tells that this model is not good enough handling the trend. Also, the seasonality is not taken into account which is affecting significantly.

Holt-winter's smoothing:

The seasonal component is taken into analysis by decomposing the series. Trend, seasonal component and the residual part is separated using multiplicative seasonality decompose function.

Fig 20. Multiplicative Decomposition



Based on this decomposition the trend is following a downward movement before plateauing at the end. In order to remove the trend STL decomposition approach is taken which consists in regressing the trend. The seasonal component shows a beautiful cyclical function but every cycle has slightly different magnitude. Residual is all time high but experiences a significant drop in middle and then again rises to its previous level.

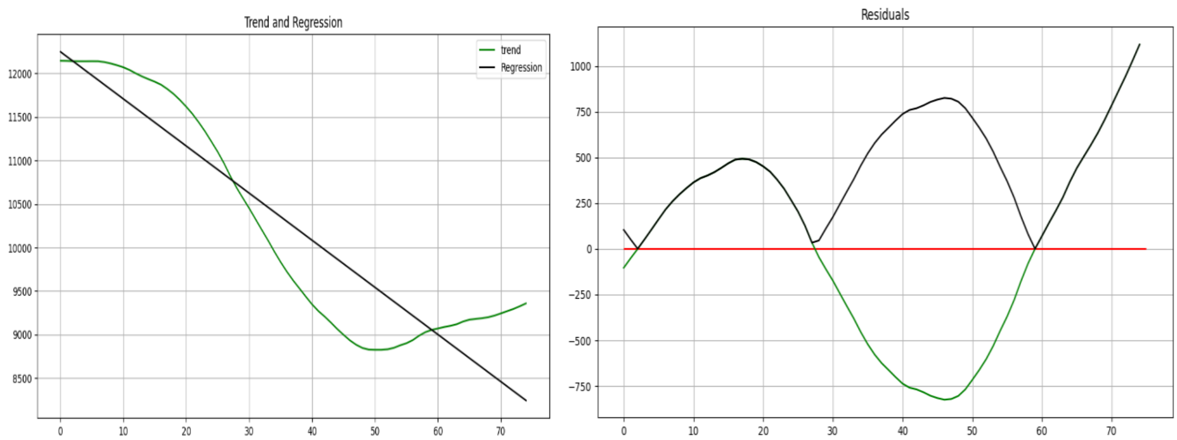
Table 3. OLS Regression Results (NIFTY50)

Dep. Variable:	Close	R-squared:	0.835
Model:	OLS	Adj. R-squared:	0.833
Method:	Least Squares	F-statistic:	369.3
Date:	Wed, 17 Jun 2020	Prob (F-statistic):	2.82e-30
Time:	17:41:24	Log-Likelihood:	-575.64
No. Observations:	75	AIC:	1155.
Df Residuals:	73	BIC:	1160.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	1.225e+04	120.817	101.387	0.000	1.2e+04	1.25e+04
x1	-54.1576	2.818	-19.216	0.000	-59.775	-48.541

Omnibus:	9.452	Durbin-Watson:	0.012
Prob(Omnibus):	0.009	Jarque-Bera (JB):	3.176
Skew:	-0.068	Prob(JB):	0.204
Kurtosis:	2.001	Cond. No.	84.9

Fig 21. Regression & Residual (NIFTY50)



The green line in the residual plot is the value of residuals form regression and the black line indicates the absolute value of the same.

The r-squared from the regression is quite good (0.835) but the absolute value of residuals increasing from left to right. This is a sign of heteroskedasticity which should not be present in the model. Similar results are seen from SENSEX data.

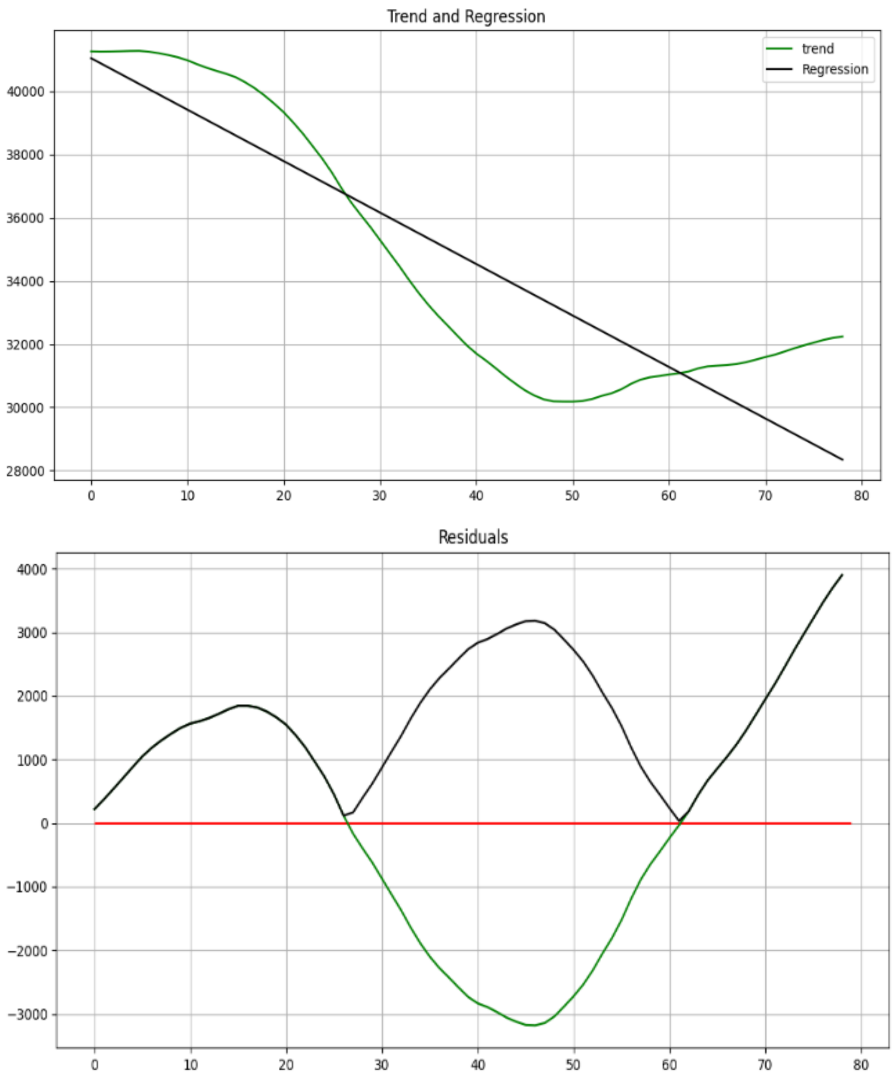
Table 4. OLS Regression Results (SENSEX)

Dep. Variable:	Close	R-squared:	0.783
Model:	OLS	Adj. R-squared:	0.780
Method:	Least Squares	F-statistic:	277.5
Date:	Tue, 30 Jun 2020	Prob (F-statistic):	3.02e-27
Time:	14:06:23	Log-Likelihood:	-710.84
No. Observations:	79	AIC:	1426.
Df Residuals:	77	BIC:	1430.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	4.105e+04	441.782	92.914	0.000	4.02e+04	4.19e+04
x1	-162.8866	9.779	-16.657	0.000	-182.359	-143.414

Omnibus:	14.329	Durbin-Watson:	0.010
Prob(Omnibus):	0.001	Jarque-Bera (JB):	4.121
Skew:	-0.151	Prob(JB):	0.127
Kurtosis:	1.923	Cond. No.	89.5

Fig 22. Regression & Residual (SENSEX)



As discussed, Holt-winter's model includes a seasonal component. Unlike the other methods the slope of the prediction line is not constant and is changing at every point. But it can be clearly seen that the seasonal component is integrated with the forecast so the curve follows a pattern similar to its historical data.

Fig 23. Holt-Winter's Seasonal (NIFTY50)

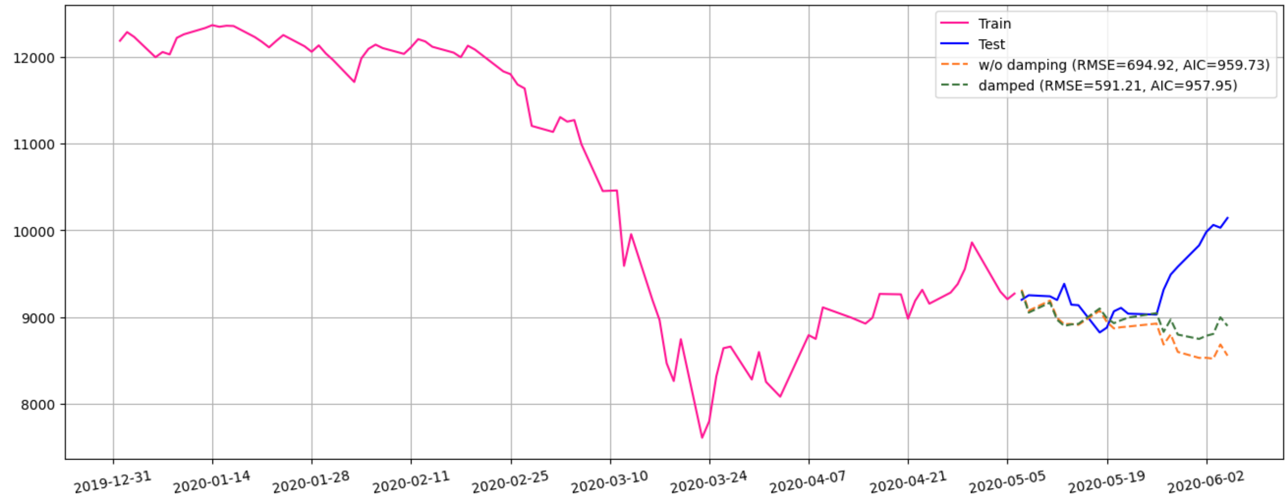
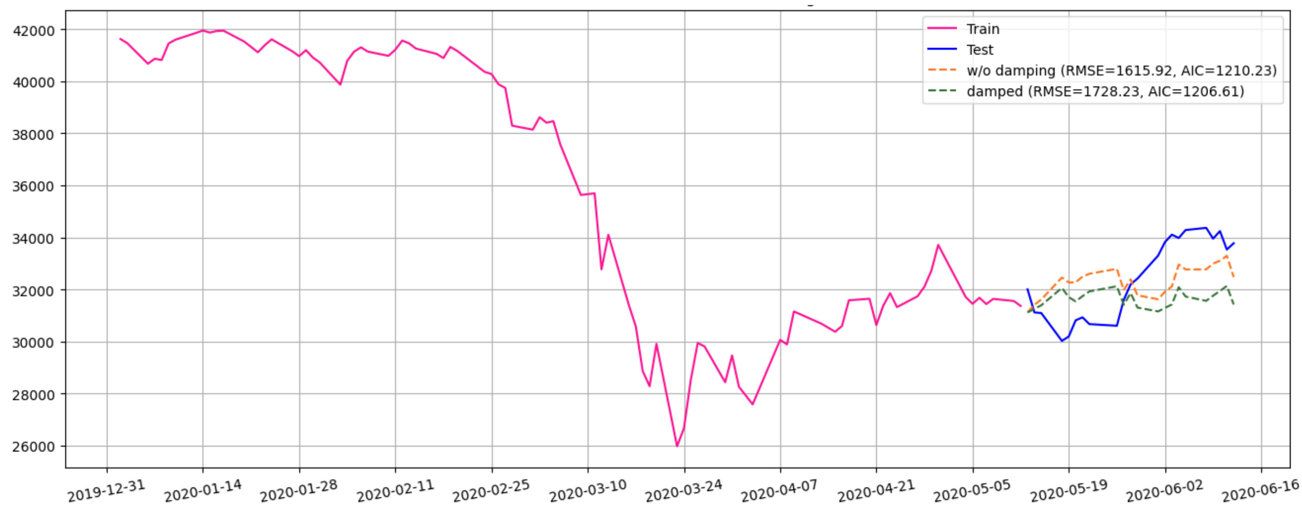


Fig 24. Holt-Winter's Seasonal (SENSEX)



At the end the prediction line is moving in a cyclical pattern which is not a good sign since it's a static pattern. Even with the seasonal component the NIFTY50 data model has the worst performance of all. Whereas the SENSEX data model has improved and is the best among all the SENSEX models. All these indicates that even if this model performs the best comparing with the other three, this also fails to delivers a convincing prediction.

Predictions:

The Holt-winter's method takes the seasonality and also is performing well or at least same as other models. So, a 60-day future prediction has been done based on this model only.

Fig 25. Holt-Winter's Seasonal prediction (NIFTY50)

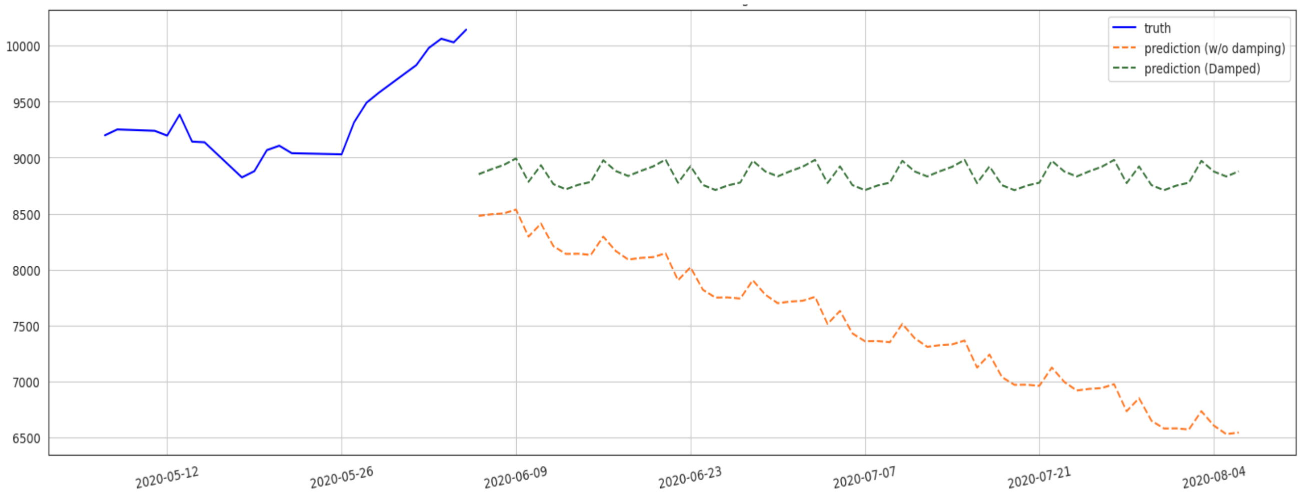
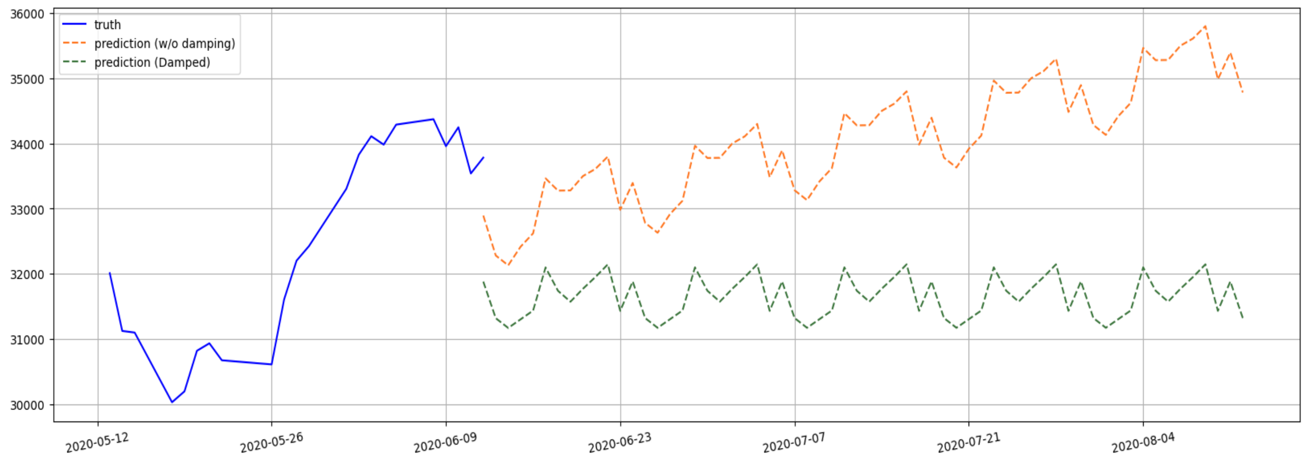


Fig 26. Holt-Winter's Seasonal prediction (SENSEX)

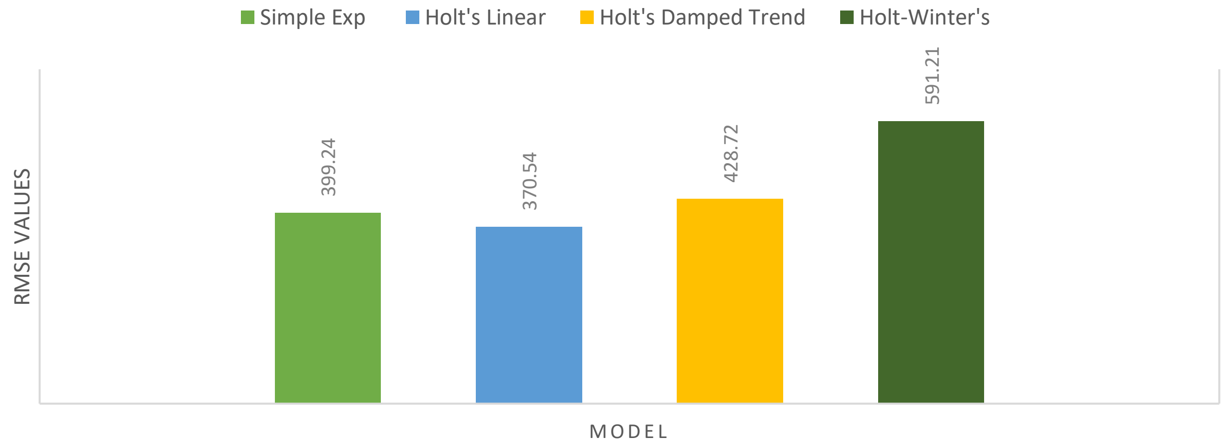


The prediction has been done with and without using the damping factor. The model has derived a cyclical seasonal component using train data. In the prediction curve it can be seen that the line is progressing by repeating the cyclical component. Without the damping factor the NIFTY50 prediction shows the future price to be continuously decreasing which is very less convincing. The curve with the damping factor is seen to be just around the moving average line of the previous data.

RMSE comparison:

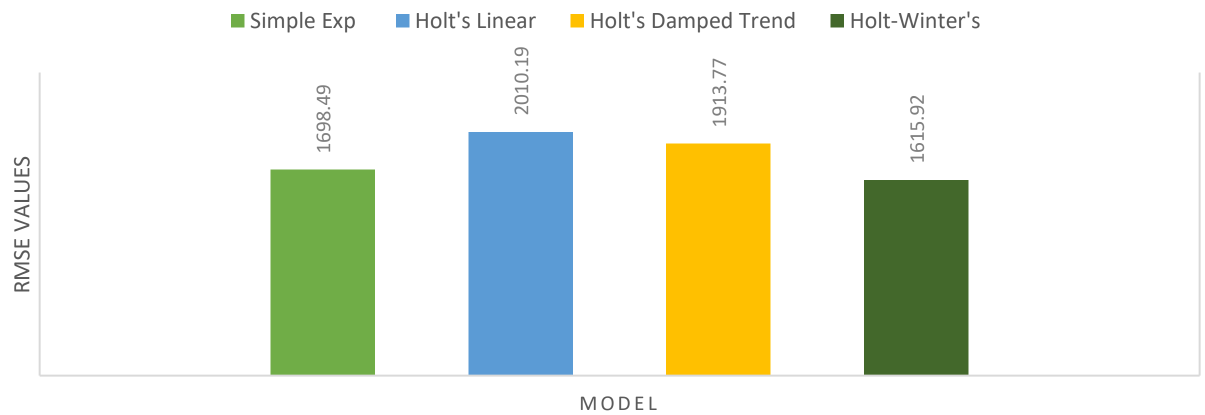
The model performance can be compared using their respective RMSE. Root Mean squared error is the absolute value difference between the predicted values and the original (test) values. The lesser the RMSE the better the model performance.

FIG 27. RMSE COMPARISON NIFTY50



For this particular time period of NIFTY50 data Holt's Linear smoothing shows the best performance.

FIG 28.RMSE COMPARISON SENSEX



The SENSEX data on the other hand displays a low RMSE on Holt-winter's seasonal smoothing. In both the cases simple exponential smoothing ranks second. RMSE parameter does tell how the model is performing against the test data. This doesn't necessarily mean that in future the model with the low RMSE will perform better than others rather it speaks the potential of the model to be truly useful in upcoming times.

10. Conclusion & Recommendations:

Conclusions: Exponential smoothing is a basic analysis tool in case of time series domain and works only up to a point. Among various exponential techniques used, Holt-winter's smoothing turns out to be pretty effective where in some cases simple exponential smoothing proves to be enough. These results can be hugely improved by using new, adaptive and complex machine learning algorithms. The prediction graphs are not at all compelling and is very less likely to be happen in real life. So, these forecasts are not reliable at all but somehow can give an idea about the near future trend and directions. If every day data is integrated with and the model is re-run every time, it is possible to get more sophisticated results. Time series analysis on a historical data renders the best result when the data is free of seasonality, trend & random fluctuations irrespective of the type of model used. But still some of the techniques allows the data to be non-stationary while doing the forecast. This results in a very poor prediction. That's why the exponential models which doesn't bother much about trends & seasonality, doesn't have much to offer. In recent times exponential smoothing is barely used in a time series analysis and fails to give trustworthy or at least a convincing prediction of the future.

Recommendation: Most of the time series data will have some trend & seasonality component within them. To overcome it several functions are present to get rid of them. For a superior and satisfactory predictions popular time series analysis methods like *ARIMA, Auto ARIMA, Seasonal ARIMA, SARIMAX, LSTM, Prophet (Additive & Multiplicative seasonality)* etc are proven to be top notch. These methods are far more complex than regression, exponential smoothing and are improving day by day. Judging the ongoing *COVID-19* pandemic situation stock market has long way to go to reach the level it was on. Stock market is basically a subset of countries' economic condition, but fluctuations in market always leading and the results are reflected sometime later in the economy. Thus, analysing and predicting the ups and down of prices can be very beneficial to potential investors. In this time pharmaceutical, agriculture and FMCG companies are the sectors that needs to be watched. By general sense sectors that are directly connected to the essential needs of human civilization will recover and flourish faster compared to others.

Research Limitations:

There are some limitations in this study that need to be paid attention to:

1. **Up to date Data**: At the time of the research the stock market was too much volatile and every point of data is important. Also, for a future prediction a real time updated database is much needed as based on the past data future predictions can change. So, at the time of this research report publication, the future prediction reported here will be outdated.
2. **Limitation on technical tools**: For several reasons advanced level machine learning and deep learning techniques [13] couldn't be used so the model performance is not good here.

11. REFERENCES

Research Papers:

[1] S. A. R. Nai-Fu Chen and Richard Roll, Economic Forces and the Stock Market, *The Journal of Business*, vol. 59, no. 3, pp. 383–403, (1986). [Online]. Available: http://www.jstor.org/stable/2352710.

[2] E. F. Fama, Random Walks in Stock Market Prices, *Financial Analysts Journal*, vol. 51, no. 1, pp. 75–80, (1995). [Online]. Available: http://www.jstor.org/stable/4479810.

[3] S. J. Grossman and R. J. Shiller, The Determinants of the Variability of Stock Market Prices, *National Bureau of Economic Research*, Working Paper 564, October (1980) [Online]. Available: http://www.nber.org/papers/w0564.

[4] A. W. Lo and A. C. MacKinlay, Stock Market Prices do not Follow Random Walks: Evidence from a Simple Specification Test, *Review of Financial Studies*, vol. 1, no. 1, pp. 41–66, (1988).

[5] K. Senthamarai Kannan, P. Sailapathi Sekar, M. Mohamed Sathik and P. Arumugam," Financial Stock Market Forecast using Data Mining Techniques", Proceedings of the International Multi Conference of Engineer and Computer Scientists 2010 Vol 1, March 17-19,2010, Hong Kong.

[6] Phichhang Ou and Hengshan Wang, "Predictions of Stock Market Index Movement by Ten Data Mining Techniques", 2009 International Conference on Advances in Recent Technologies in Communication and Computing, vol. 31, pp. 54-56

[7] Aditya Gupta, Non-Student Member, IEEE and Bhuwan Dhingra, Non-Student Member, IEEE, "Stock Market Predictions Using Hidden Markov Models", 2012 International Conference on Advances in Recent Technologies in Communication and Computing, vol. 34, pp. 37-39

[8] George S. Atsalakis and Kimon p. Valvanis "Forecasting Stock Market using a Neuro-Fuzzy based Technology", Journal Expert System with Application 36(2009) 1069610707

[9] Binoy B. Nair, N.Mohana Dharini and V.P. Mohandas," A Stock Market Trend Prediction System using a Hybrid Decision Tree Neuro-Fuzzy System",2010 International Conference on Advances in Recent Technologies in Communication and Computing

[10] J. Gong and S. Sun, A New Approach of Stock Price Prediction Based on Logistic Regression Model, In *2009. NISS '09. International Conference on New Trends in Information and Service Science*, pp. 1366–1371, June (2009).

[11] S.Arun Joe Babulo, B. Janaki, C. Jeeva,"Stock Market Indices Prediction with Various Neural Network Models" International Journal of Computer Science and Mobile Applications, Vol.2 Issue. 3, March-2014

[12] Dase R.K and Pawar D.D.”Applications of Neural Network for Stock Markety Prediction: A review of literature”, International Journal of Machine Intelligence, ISSN: 09752927, Volume 2, Issue 2, 2010

[13] Ronald J. Balvers, Thomas F. Cosimano and Bill McDonald,” Predicting Stock Returns in an Efficient Market”, Journal of Finance, Volume 45, Issue 4 (Sept.1990). 1109-1128

[14] David Enke and Suraphan Thawornwong, “The use of data mining and neural network for forecasting stock market returns”, Expert System with Applications 29 (2005)927940

Online Sources:

Andrea Laneti (11 Mar’20): The corona virus effect on stock market. Retrieved from: https://towardsdatascience.com/the-coronavirus-effect-on-the-stock-market-b7a4739406e8

Nagesh Singh Chauhan (27 Nov’19): Stock Market Forecasting Using Time series analysis. Retrieved from: https://towardsdatascience.com/stock-market-forecasting-using-time-series-c3d21f2dd37f

Ifeoma Ojialor (5 Dec’19): Performing a Time series analysis on AAPL stock Index. Retrieved from: https://towardsdatascience.com/performing-a-time-series-analysis-on-the-aapl-stock-index-3655da9612ff

Stock Market Prediction Using Time Series Analysis. Retrieved from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3168423

Prediction Models for Indian Stock Market. Retrieved from: https://www.sciencedirect.com/science/article/pii/S1877050916311619

Aishwarya Singh (25 Oct’18): Stock prices prediction using machine learning and deep learning techniques (with python codes). Retrieved from: https://www.analyticsvidhya.com/blog/2018/10/predicting-stock-price-machine-learningnd-deep-learning-techniques-python/

D. Soni, S. Agarwal, T. Agarwel, P. Arora and K. Gupta, "Optimised Prediction Model for Stock Market Trend Analysis," *2018 Eleventh International Conference on Contemporary Computing (IC3)*, Noida, 2018, pp. 1-3, doi: 10.1109/IC3.2018.8530457.

Selva Prabhakan: Complete guide to time series forecasting in Python. Retrieved from: https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/

Thomas Vincent (23 Mar’17): A guide to time series forecasting with ARIMA in python 3. Retrieved from: https://www.digitalocean.com/community/tutorials/a-guide-to-time-series-forecasting-with-arima-in-python-3

Thomas Vincent (14 Mar’17): A guide to time series visualization with python 3. Retrieved from: https://www.digitalocean.com/community/tutorials/a-guide-to-time-series-visualization-with-python-3

Benjamin Etienne (9 Feb’19): Time series in Python – Exponential smoothing and ARIMA processes. Retrieved from: https://towardsdatascience.com/time-series-in-python-exponential-smoothing-and-arima-processes-2c67f2a52788?gi=15c62475bb1c

Alina Zhang (21 Sep’18): How to build exponential smoothing methods using python- Simple exponential smoothing, Holt, and Holt-winters. Retrieved from: https://medium.com/datadriveninvestor/how-to-build-exponential-smoothing-models-using-python-simple-exponential-smoothing-holt-and-da371189e1a1

Benjamin Etienne (15 Feb’19): Time series in Python – part 2 dealing with seasonal data. Retrieved from: https://towardsdatascience.com/time-series-in-python-part-2-dealing-with-seasonal-data-397a65b74051

Jason Brownlee (2 Jan’17): Autoregression models for time series forecasting with python. Retrieved from: https://machinelearningmastery.com/autoregression-models-time-series-forecasting-python/

Jason Brownlee (6 Aug’2018): 11 classical time series forecasting methods with python (cheat sheet). Retrieved from: https://machinelearningmastery.com/time-series-forecasting-methods-in-python-cheat-sheet/

Varshita Sher (10 Jun’20): Time series analysis using pandas in python. Retrieved from: https://towardsdatascience.com/time-series-analysis-using-pandas-in-python-f726d87a97d8

Melina Mackey (13 Apr’20): Building a technical analysis chart in python. Retrieved from: https://medium.com/analytics-vidhya/building-a-technical-analysis-chart-with-python-17107b78b297

Appendix A

Statistical Parameters

A.1. Relative Strength Index: The Relative Strength Index (RSI) is a momentum yardstick used in a technical analysis that tests the extent of recent price shifts in order to determine over-purchased or over-sold conditions in the price of a commodity or other asset. The RSI is seen as an oscillator (a line graph that travels between two extremes) and can be read from 0 to 100. The indicator was originally designed by J. Welles Wilder Jr. introduced "New Concepts in Technical Trading Systems" in his seminal 1978 book. Standard meaning and application of the RSI is that the values of 70 or higher suggest that the security is being over-purchased or over-valued and can be pre-empted by a reversing of the trend or a retracting of the amount. The RSI reading of 30 or below suggests an over-sold or undervalued situation.

A.2. Mean Absolute Percentage Error: The mean absolute percentage error (MAPE), also known as mean absolute percentage deviation (MAPD), is a measure of the accuracy of the forecasting method in statistics, e.g. in the trend estimation, which is also used as a loss function in machine learning for regression problems. Typically, the accuracy is expressed as the ratio specified by the formula:

$$\mathrm{M} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|$$

Where A_t is the real value and F_t is the predicted value. The MAPE is also often stated as a number, which is multiplied by 100 by the above equation. The difference between A_t and F_t is again separated by the real A_t value. The absolute value of this equation is added up for each predicted point in time and divided by the number of points n. Multiplying by 100% is a percentage error.

A.3. Root Mean Squared Error: The root-mean - square variance (RMSD) or root-mean - square error (RMSE) is a widely used calculation of the variations between the values (sample or population values) expected by the model or estimator and the values observed. The RMSD represents the square root of the second observation moment of the difference between the expected values and the observed values or the quadratic mean of these differences. These anomalies are referred to as residuals when the measurements are conducted over the data set used for estimation and are referred to as errors (or measurement errors) when measured out of the set. The RMSD is used to aggregate the magnitude of prediction errors for different times into a single measure of predictive power. RMSD is a measure of precision, a measurement of forecasting errors in various models for a single dataset, though not between datasets, since it is scale-dependent.

RMSD is often non-negative, so a value of 0 (almost rarely obtained in practice) will suggest a good match for the results. In general, a lower RMSD value is better than a higher one. However, comparisons between different types of data would be invalid because the measure depends on the scale of the numbers used. RMSD is the square root of the maximum square error. The effect of increasing error on RMSD is equal to the magnitude of the squared error; thus, greater errors have a relatively large effect on RMSD. As a consequence, RMSD is vulnerable to outsiders.

A.4. Efficient Market Hypothesis: The Efficient Market Hypothesis (EMH), also known as the Efficient Market Principle, is the belief that share prices represent all knowledge and continuous alpha generation is unlikely. According to the EMH, stocks are always trading at fair value on exchanges, making it unfeasible for investors to purchase undervalued stocks or sell stocks at inflated prices. This would also be difficult to outperform the overall market by skilled asset picking or market timing, and the only way for an investor to achieve better returns is through purchasing riskier stocks. While it is a pillar of mainstream financial philosophy, the EMH is particularly contentious and sometimes contentious. Believers argue that there is no point in looking for undervalued stocks or trying to predict market trends through either a fundamental or a technical analysis.

Theoretically, neither a scientific nor a fundamental analysis can deliver risk-adjusted excess returns (alpha) reliably, and only inside details can result in overweight risk-adjusted returns.

A.5. Augmented Dickey-Fuller Test: The Augmented Dickey Fuller Test (ADF) is a stationary root test unit. Unit roots can produce unforeseeable effects in an examination of the time sequence. The Augmented Dickey-Fuller test may be used with a serial correlation. The ADF test can accommodate more complex models than the Dickey-Fuller test, and is also more efficient. That said, it should be used with caution because—like other unit root tests—it has a fairly high error rate of Type I.

Hypothesis for the test:

a) Null: There is a unit root
b) The alternative hypothesis diverges vaguely from the equation in use. The basic alternative is that the time series is stationary (or trend-stationary).

Null hypothesis says that there is a unit root which also implies that the time sequence is not stationary.