# Application of Algorithmic Strategies For Predicting Stock Prices and Algorithmic Trading


## Abstract

Stock market price prediction has always been an exciting subject that has piqued researchers' interest worldwide to develop better predictive models. Our study has two parts. In the first part we have used Random Forests to predict the stock prices of 8 NIFTY50 stocks(from 4 different sectors) and compared the results with traditionally used algorithms like linear regression and Autoregressive Integrated Moving Average.

In the second part, we have used an algorithmic trading strategy where buy and sell calls are made using Recurrent Reinforcement Learning(RRL) with Gradient Ascent. The cumulative normalised returns with this strategy for these 8 stocks are then compared with a buy and hold strategy for comparative study.

# Chapter 1: Objective and Introduction

This project aims to utilize the potential of algorithmic strategies optimum strategy to predict future stock prices and time the buy-sell decisions to maximise profits. Constructing a portfolio, in general, has three aspects(if we ignore leverage for simplicity)- a selection of stocks depending on risk appetite(standard deviation/volatility), predicting future price expectations, and deciding when to buy/sell(long/short). As part of this project, we have dealt only with the last two aspects as risk-taking ability is subjective in nature.

This project consists of two parts-

**Part A**- Comparative study of Random forest Model with traditional algorithms to predict stock prices.

**Part B**- Application of Recurrent Reinforcement Learning with Gradient Ascent compared to a buy & hold strategy

For both parts, our analysis involves applying the concept to diverse companies and industries to validate our results.

# Chapter 2: Theoretical Background

## Part A- Stock Price Forecasting

Stock market price prediction has always been an exciting subject that has piqued researchers' interest to develop better predictive models. Our comparative study of stock market forecasting has two parts –

a) Time series forecast using ARIMA

b) Multivariate Timeseries Forecasting model using Linear regression and Random Forest

a) **Time series forecast using ARIMA**

ARIMA is a univariate time series-based forecasting technique. They are known to be robust and efficient in financial time series forecasting. It combines two models- autoregression and moving average model. In an autoregression model, the regression-based prediction is made using a linear combination of the target variable's past values against itself. On the other hand, a moving average model uses the average of a subset of the data in the previous continuous days to predict the future stock price. This model also incorporates some other features like seasonality and trend.

b) **Multivariate Timeseries Forecasting model using Linear regression and Random Forest**

The **variables** used as part of the multivariate analysis are- price, volume, day, month, and year

**Correlation matrix** is a grid that describes the strength of the relationship between the relative movements of multiple pairs of variables.

**Linear Regression**

Linear regression is a quantitative method that enables us to study the relationship between two continuous (quantitative) variables. If the data shows a linear relationship between two variables Y and X, then the straight line which maps their relationship closest is the regression line. It involves finding and minimizing the squared distances between predicted and targeted values.

Y = a + bX

The values of a and b are calculated using the following formulas-

$$a = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}$$

$$b = \frac{1}{n}(\sum_{i=1}^{n} y_i - a\sum_{i=1}^{n} x_i)$$

**Random Forest**

It is a technique that utilizes multiple decision trees to determine the final output rather than instead of a single decision tree and can be used for regression and classification problems.

## Part B- Comparison of Recurrent Reinforcement Learning Strategy with Gradient Ascent Against a Buy and Sold Strategy

The **Sharpe ratio** measures an investment's performance (could be a security or portfolio) compared to a risk-free asset after adjusting for its risk.

$$S_T = \frac{Average(R_t)}{Standard\ Deviation(R_t)} \quad \text{for interval} \quad t = 1, \ldots, \mathrm{T}$$

In the above formula, we have assumed the risk-free rate to be 0 for simplicity

The Sharpe ratio will be used as the reward function but to time the trade, we will determine position $F$ at time $t$:

$F_t = \tanh(\theta^T x_t)$

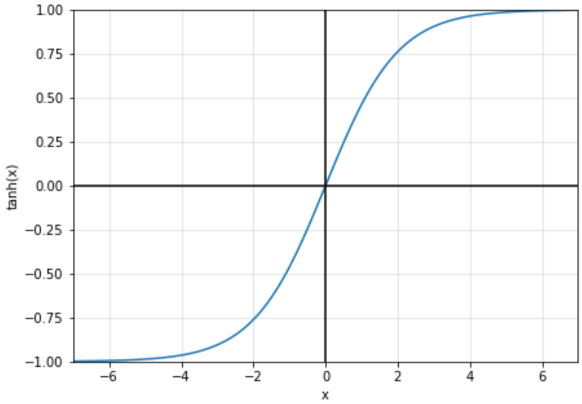


**Figure 4.1: Graph of tanh function**

This function generates a value between 0 and 1, indicating what percentage of the portfolio should buy the asset.

There are three types of positions that can be held: long, short, or neutral.

When $F_t > 0$, a long position is chosen. In this case, the expectation is that the price appreciates by period t+1.

When $F_t < 0$, a short position is chosen. In this case, an asset that is not owned is sold, with the expectation that the shares will be available at a lower price at period t+1 to fulfill the contract. If the price at t+1 is lower, then the trade makes a profit.

At $F_t = 0$, a neutral position is chosen. In this case, irrespective of the outcome at time t+1, there will be no gain or loss.

In the above formula for $F_t$, $x_t$ is a vector with value,

$x_t = [1, r_1 \ldots\ldots r_{t-M}, F_{t-1}]$

where M is the number of time series inputs being considered by the trader. The value of this variable is fixed during parameter optimisation. In our paper, we have termed the parameter as a "lookback window", during parameter optimisation.

$r_t$ is the difference in asset value at time t and (t-1).

The total number of shares that can be bought/sold can be denoted by the formula $n_t = \mu * F_t$, where μ is the maximum number of shares per transaction.

We then calculate the returns using the below formula-

$$R_t = \mu \cdot \left(F_{t-1} \cdot r_t - \delta \left|F_t - F_{t-1}\right|\right)$$

Where F represents the trader's position and δ is the cost for a transaction at period t. In our paper, we have termed the parameter as "commission", during parameter optimisation. The formula for returns will be used for calculating Sharpe ratio

$$S_T = \frac{E[R_t]}{\sqrt{E[R_t^{\,2}] - (E[R_t])^2}} = \frac{A}{\sqrt{B - A^2}} \quad \text{where } A = \frac{1}{T}\sum_{t=1}^{T} R_t \quad \text{and} \quad B = \frac{1}{T}\sum_{t=1}^{T} R_t^{\,2}$$

After taking partial derivates and simplifying, we get the following equations-

$$\frac{dR_t}{dF_t} = \frac{d}{dF_t}\left\{\mu \cdot \left(F_{t-1} \cdot r_t - \delta\left|F_t - F_{t-1}\right|\right)\right\} = \frac{d}{dF_t}\left\{-\mu \cdot \delta \cdot \left|F_t - F_{t-1}\right|\right\} = \begin{cases} -\mu \cdot \delta & F_t - F_{t-1} > 0 \\ \mu \cdot \delta & F_t - F_{t-1} < 0 \end{cases}$$
$$= -\mu\delta \cdot \operatorname{sgn}(F_t - F_{t-1})$$

$$\frac{dR_t}{dF_{t-1}} = \frac{d}{dF_{t-1}}\left\{\mu \cdot \left(F_{t-1} \cdot r_t - \delta\left|F_t - F_{t-1}\right|\right)\right\} = \mu \cdot r_t - \frac{d}{dF_{t-1}}\left\{-\mu \cdot \delta \cdot \left|F_t - F_{t-1}\right|\right\} = \begin{cases} \mu \cdot \delta & F_t - F_{t-1} > 0 \\ -\mu \cdot \delta & F_t - F_{t-1} < 0 \end{cases}$$
$$= \mu \cdot r_t + \mu\delta \cdot \operatorname{sgn}(F_t - F_{t-1})$$

$$\frac{dF_t}{dw} = \frac{d}{dw}\left\{\tanh(w^T x_t)\right\} = (1 - \tanh(w^T x_t)^2) \cdot \frac{d}{dw} \cdot \left\{w^T x_t\right\} = (1 - \tanh(w^T x_t)^2) \cdot \left\{x_t + w_{M+2}\frac{dF_{t-1}}{dw}\right\}$$

We then used gradient ascent in our optimization to converge to the local maximum of the partial derivative function.

The following formula has been used for gradient ascent-

$$\frac{dS_T}{d\theta} = \sum_{t=1}^{T}\left(\frac{dS_T}{dA}\frac{dA}{dR_t} + \frac{dS_T}{dB}\frac{dB}{dR_t}\right) \cdot \left(\frac{dR_t}{dF_t}\frac{dF}{d\theta} + \frac{dR_t}{dF_{t-1}}\frac{dF_{t-1}}{d\theta}\right)$$

We then update our $\theta$ for each epoch using the formula-

$\theta$=$\theta$+($dS_T/d\theta$), where $\alpha$ is our learning rate.

# Chapter 3: Data and Research Methodology

In the research, historical data of stock prices of companies has been used. The stocks' closing prices were extracted from the NSE website and the yahoo finance portal. Although most research papers have used data from the NSE, we found that yahoo finance has also been used for some of the papers due to the ease of data extraction through dedicated software packages. We compared the NSE data with the one from Yahoo Finance and they were a perfect match. Hence, in this paper, the data has been mostly extracted from the yahoo finance portal.

The weekly data of stock prices has been used for a period of 10 years from 01-01-2011 to 01-01-2021. We have used the adjusted closing prices for all calculations.

For our analysis, we have chosen eight stocks from the NIFTY50. These eight companies belong to four different sectors. The companies selected for this analysis are-

| S.No. | Company Name | Industry | Symbol | ISIN Code |
|---|---|---|---|---|
| 1 | HDFC Bank Ltd. | FINANCIAL SERVICES | HDFCBANK | INE040A01034 |
| 2 | Bajaj Finance | FINANCIAL SERVICES | BAJFINANCE | INE296A01024 |
| 3 | JSW Steel Ltd. | METALS | JSWSTEEL | INE019A01038 |
| 4 | Tata Steel Ltd. | METALS | TATASTEEL | INE081A01012 |
| 5 | Britannia Industries Ltd. | CONSUMER GOODS | BRITANNIA | INE216A01030 |
| 6 | ITC Ltd. | CONSUMER GOODS | ITC | INE154A01025 |
| 7 | Infosys Ltd. | IT | INFY | INE009A01021 |
| 8 | Tata Consultancy Services Ltd. | IT | TCS | INE467B01029 |

**Table 5.1 – List of NIFTY50 Stocks Selected for Analysis**

## Assumptions

1. Shares of stocks are infinitely divisible, allowing trades in fractional values
2. The cost of transacting is 0

## Part A- Stock Price Forecasting

**Step 1**- The weekly data for the stock was downloaded for the period 01/01/2011 to 01/01/2021 from yahoo finance, and the graph was plotted

**Step 2**- The data was analyzed for seasonality and trend

**Step 3**- ARIMA parameters were found through an iterative process, and the observed price of the stock was plotted along with the ARIMA forecast.

**Step 4**- The Mean Squared Error(MSE), Root Mean Squared Error (RMSE), and Average accuracy was calculated for ARIMA.

**Step 5**- We then move on to building the multivariate linear regression and random forest models. We first check for Outliers through outlier detection and ascertain the reason for it in case the model needs to be adjusted.

**Step 6**- A correlation matrix is created to identify the relation between the variables.

**Step 7**- The linear regression and random forest models are created

**Step 8**- The R-square value, MSE, RMSE, and average accuracy are calculated for both models

**Step 9**- The process is repeated for the other seven companies

## Part B- Comparison of Recurrent Reinforcement Learning Strategy in comparison with a Buy and Sold Strategy

**Step 1**- Two sets of data are used in this model. The stock's weekly data was downloaded for the time period 01/01/2019 to 01/01/2021 from yahoo finance and is used as the training set. The stock's weekly data was downloaded for the time period 01/01/2011 to 01/01/2021 from yahoo finance and is used as the test set.

**Step 2**- After setting up the iterative process as mentioned in the previous section, through a hit and trial process, the values of the parameters are found for the normalized data. The parameter optimization process should ensure that the Sharpe ratio converges to the highest value. In this study the fixed parameters used for all companies are- epochs(number of iterations)=2000, M(lookback window)=10, commission(cost of transaction)=0, learning_rate($\alpha$)=.05

**Step 3**- The plot for the RRL model with gradient ascent vs buy&hold data is created for both training and test models.

**Step 4**- The difference in cumulative normalized return between RRL model with gradient ascent vs buy&hold data is calculated for both training and test data set and the results are analyzed.

**Step 5**- The process is repeated for the other seven companies

# Chapter 4: Observations and Analysis

## Part A- Stock Price Forecasting

### Price Charts

**HDFC Bank**

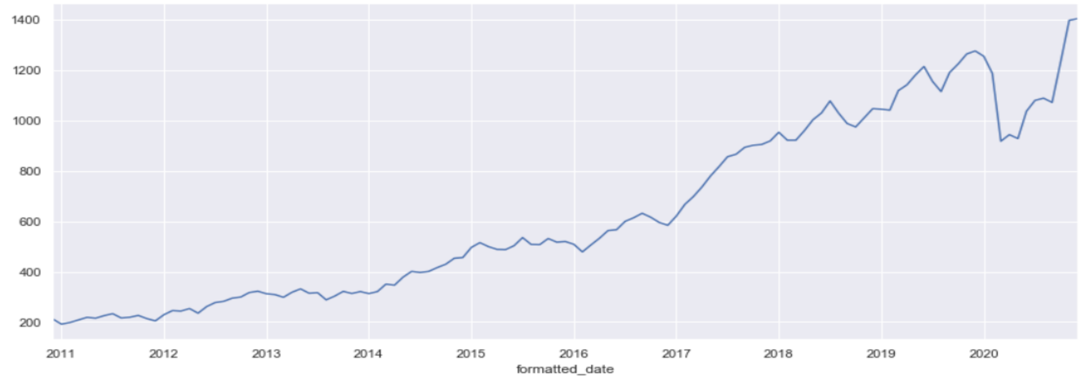


**Figure 6.1: Price Chart for HDFC Bank**

**Bajaj Finance**

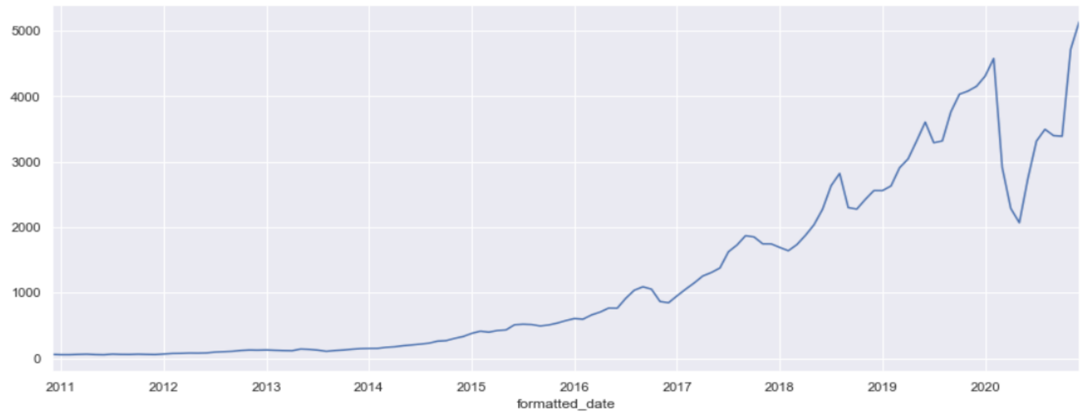


**Figure 6.2: Price Chart for Bajaj Finance**

**JSW Steel**

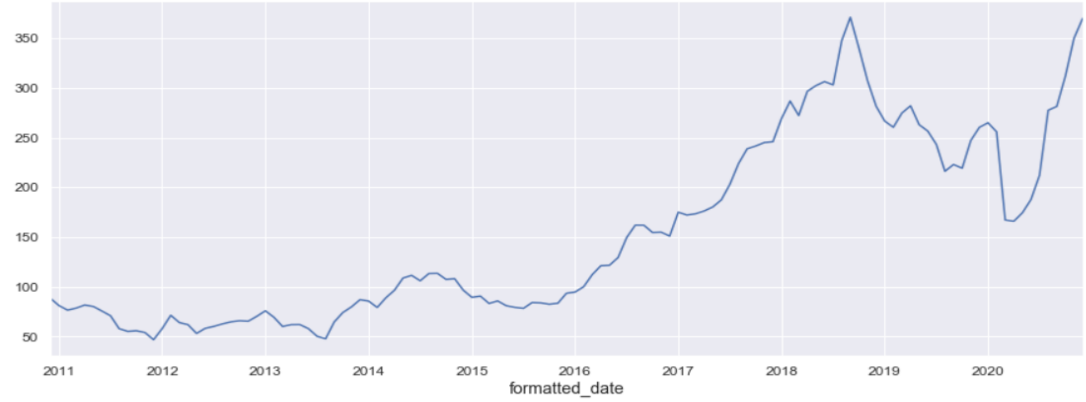


**Figure 6.3: Price Chart for JSW Steel**

**Tata Steel**

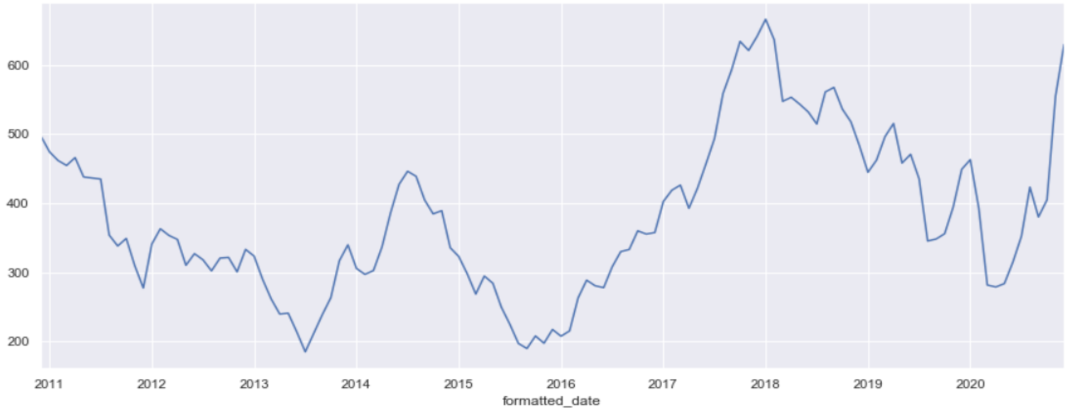


**Figure 6.4: Price Chart for Tata Steel**

**Brittania**


**Figure 6.5: Price Chart for Brittania**

**ITC**

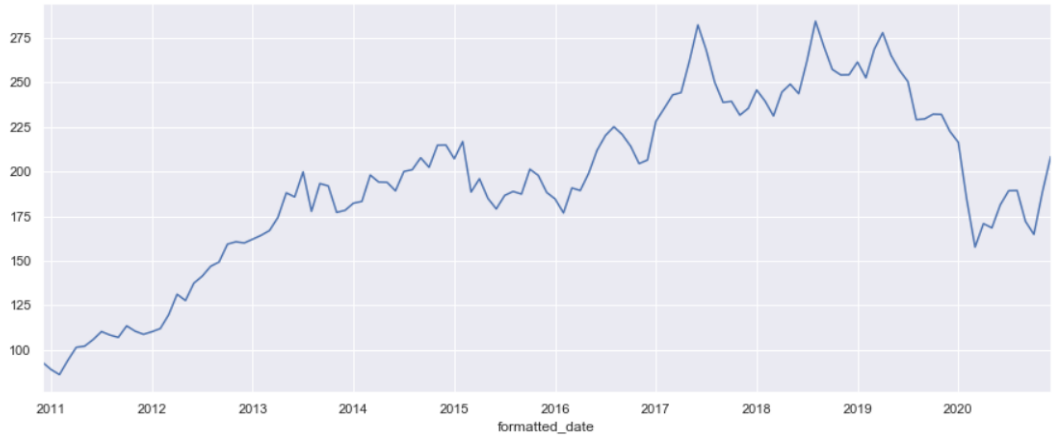


**Figure 6.6: Price Chart for ITC**

## Infosys

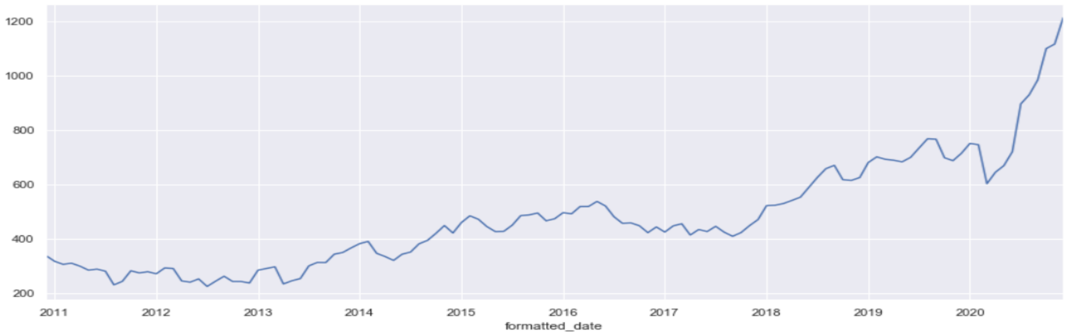


**Figure 6.7: Price Chart for Infosys**

## Tata Consultancy Services


**Figure 6.8: Price Chart for Tata Consultancy Services**

## ARIMA Parameters and Forecast

**HDFC Bank**

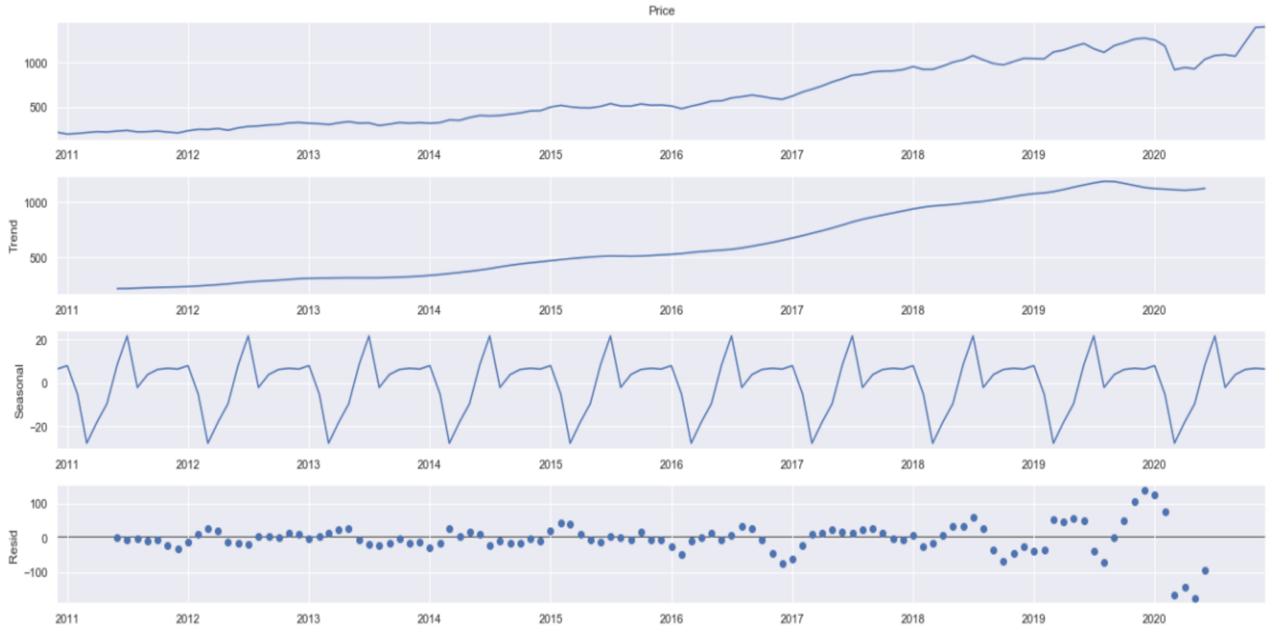


**Figure 6.9: Seasonality and Trend for HDFC Bank**

ARIMA(0, 1, 1)x(1, 1, 1, 12)12 - AIC: 992.310212472688

| | coef | std err | z | P>\|z\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| ma.L1 | 0.2160 | 0.090 | 2.404 | 0.016 | 0.040 | 0.392 |
| ar.S.L12 | -0.7616 | 0.287 | -2.651 | 0.008 | -1.325 | -0.199 |
| ma.S.L12 | -0.3216 | 0.299 | -1.076 | 0.282 | -0.907 | 0.264 |
| sigma2 | 2042.3559 | 168.190 | 12.143 | 0.000 | 1712.709 | 2372.003 |

**Table 6.1: ARIMA Statistics for HDFC Bank**

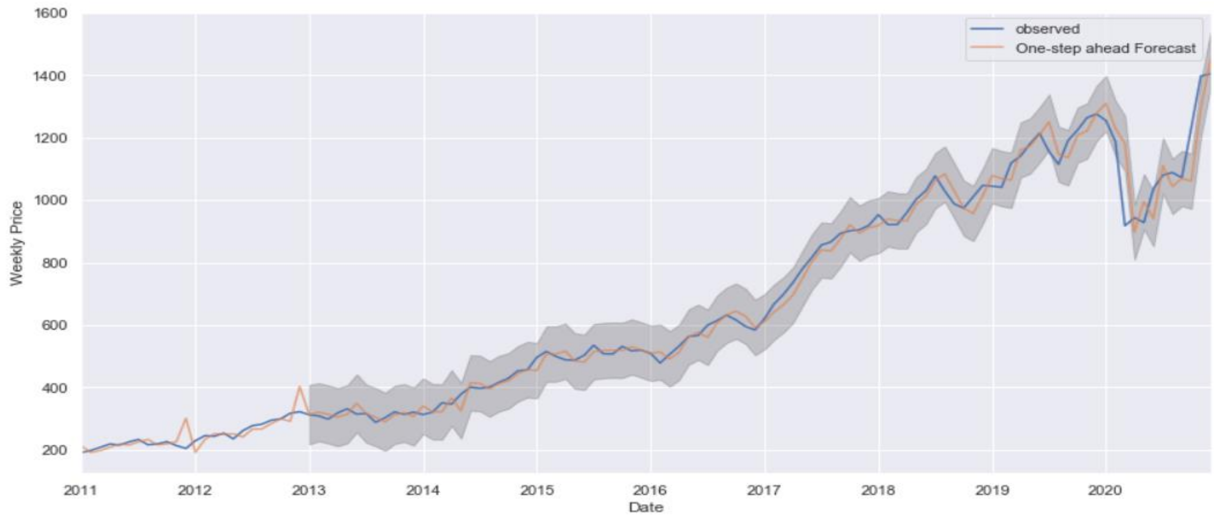


**Figure 6.10: Graph of Observed vs One Step Ahead Forecast for HDFC Bank**

## Bajaj Finance

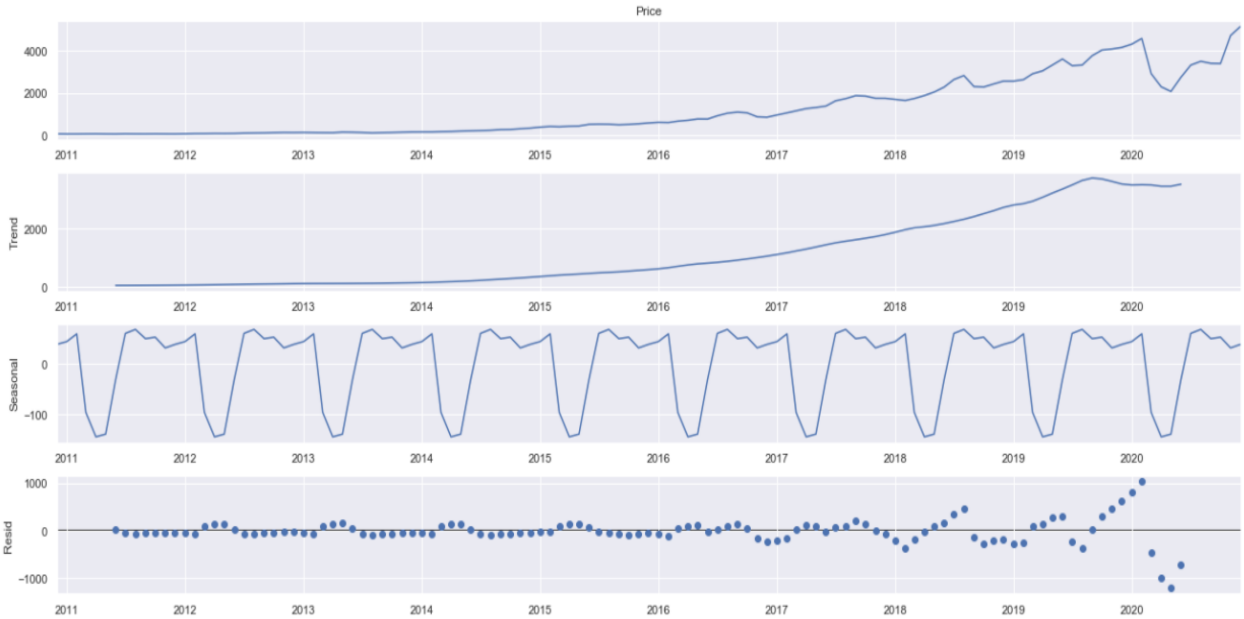


**Figure 6.11: Seasonality and Trend for Bajaj Finance**

ARIMA(0, 1, 1)x(1, 1, 1, 12)12 - AIC: 1327.7424537059544

| | coef | std err | z | P>\|z\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| ma.L1 | 0.2523 | 0.122 | 2.066 | 0.039 | 0.013 | 0.492 |
| ar.S.L12 | -0.8815 | 0.224 | -3.932 | 0.000 | -1.321 | -0.442 |
| ma.S.L12 | -0.2778 | 0.167 | -1.668 | 0.095 | -0.604 | 0.049 |
| sigma2 | 7.28e+04 | 4380.572 | 16.619 | 0.000 | 6.42e+04 | 8.14e+04 |

**Table 6.2: ARIMA Statistics for Bajaj Finance**

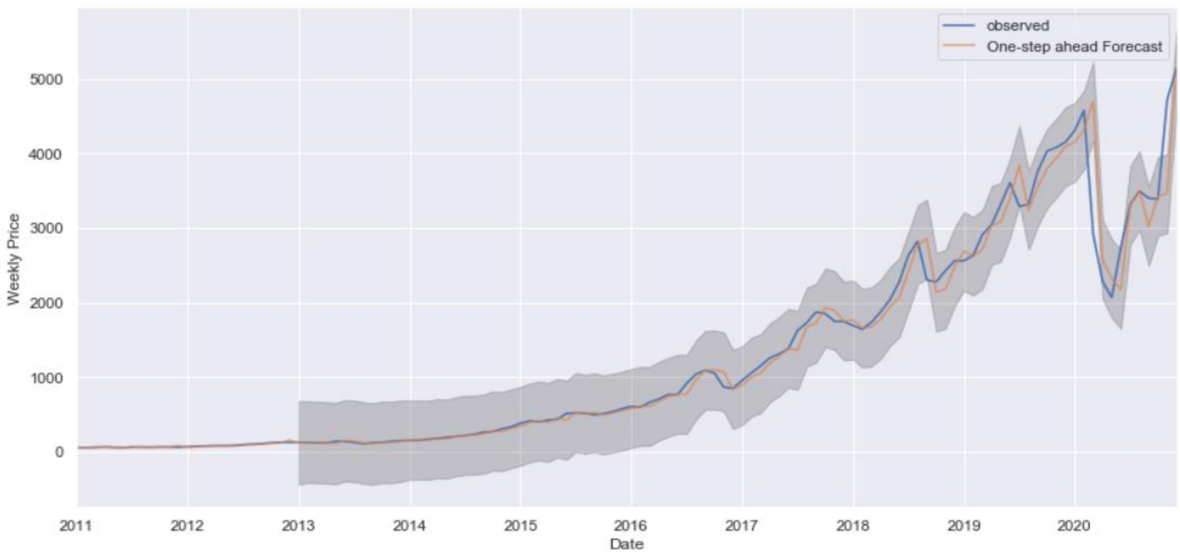


**Figure 6.10: Graph of Observed vs One Step Ahead Forecast for Bajaj Finance**

## JSW Steel

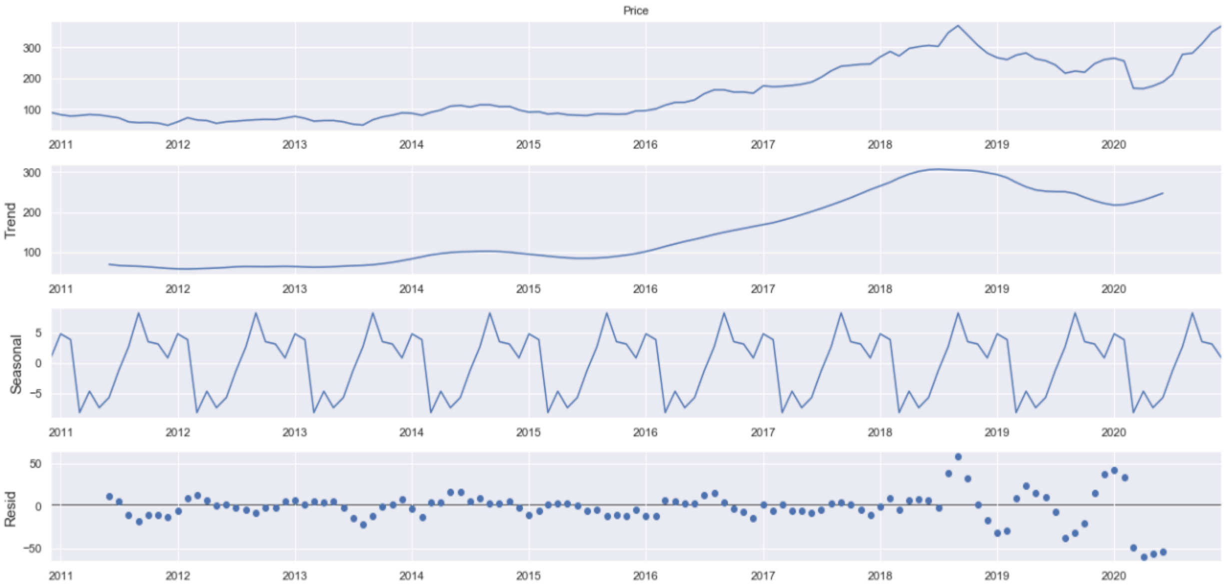


**Figure 6.13: Seasonality and Trend for JSW Steel**

ARIMA(0, 1, 1)x(1, 1, 1, 12)12 - AIC: 804.9395458302852

| | coef | std err | z | P>\|z\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| ma.L1 | 0.3930 | 0.079 | 4.979 | 0.000 | 0.238 | 0.548 |
| ar.S.L12 | -0.5409 | 0.102 | -5.285 | 0.000 | -0.741 | -0.340 |
| ma.S.L12 | -0.6718 | 0.108 | -6.240 | 0.000 | -0.883 | -0.461 |
| sigma2 | 262.8997 | 22.468 | 11.701 | 0.000 | 218.863 | 306.937 |

**Table 6.3: ARIMA Statistics for JSW Steel**


**Figure 6.14: Graph of Observed vs One Step Ahead Forecast for JSW Steel**

## Tata Steel

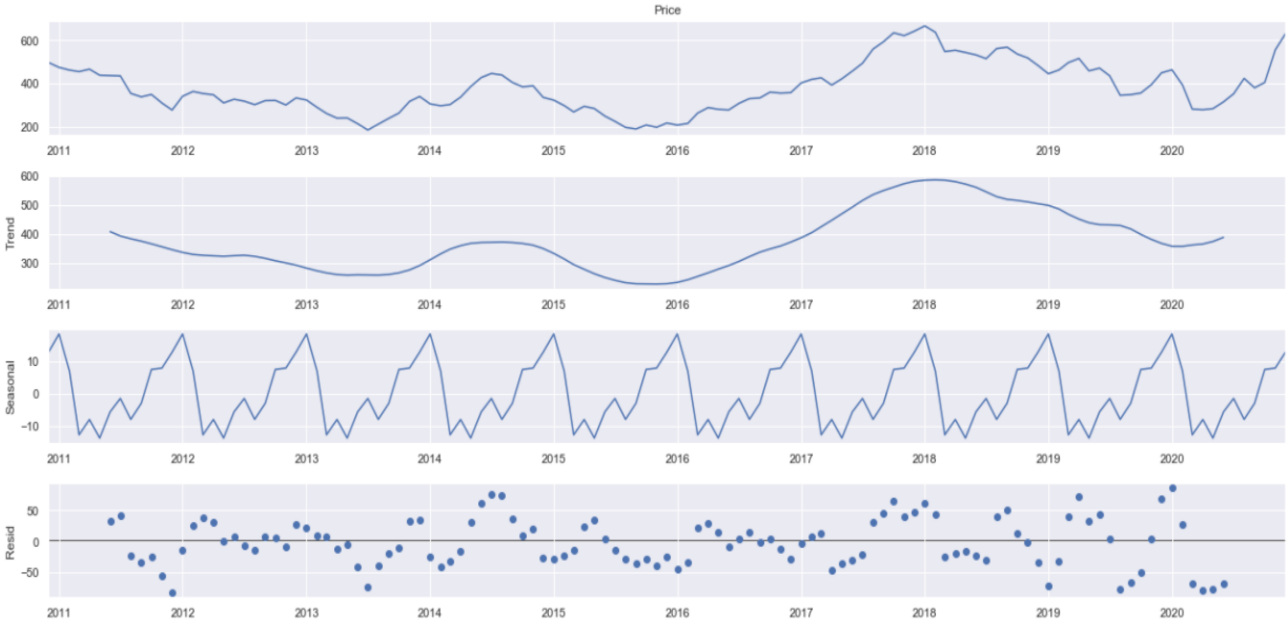


**Figure 6.15: Seasonality and Trend for Tata Steel**

ARIMA(0, 1, 1)x(1, 1, 1, 12)12 - AIC: 958.0929381747515

| | coef | std err | z | P>\|z\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| ma.L1 | 0.4966 | 0.077 | 6.425 | 0.000 | 0.345 | 0.648 |
| ar.S.L12 | -0.1801 | 0.124 | -1.448 | 0.148 | -0.424 | 0.064 |
| ma.S.L12 | -1.0000 | 0.152 | -6.589 | 0.000 | -1.297 | -0.703 |
| sigma2 | 1109.5762 | 0.000 | 8.11e+06 | 0.000 | 1109.576 | 1109.576 |

**Table 6.4: ARIMA Statistics for Tata Steel**

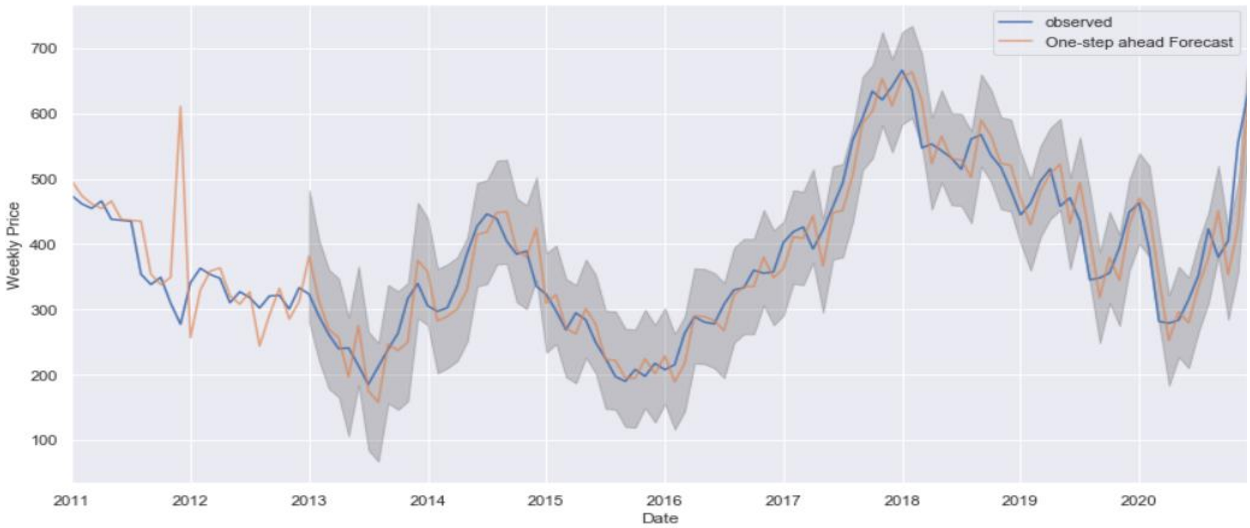


**Figure 6.16: Graph of Observed vs One Step Ahead Forecast for Tata Steel**

## Brittania

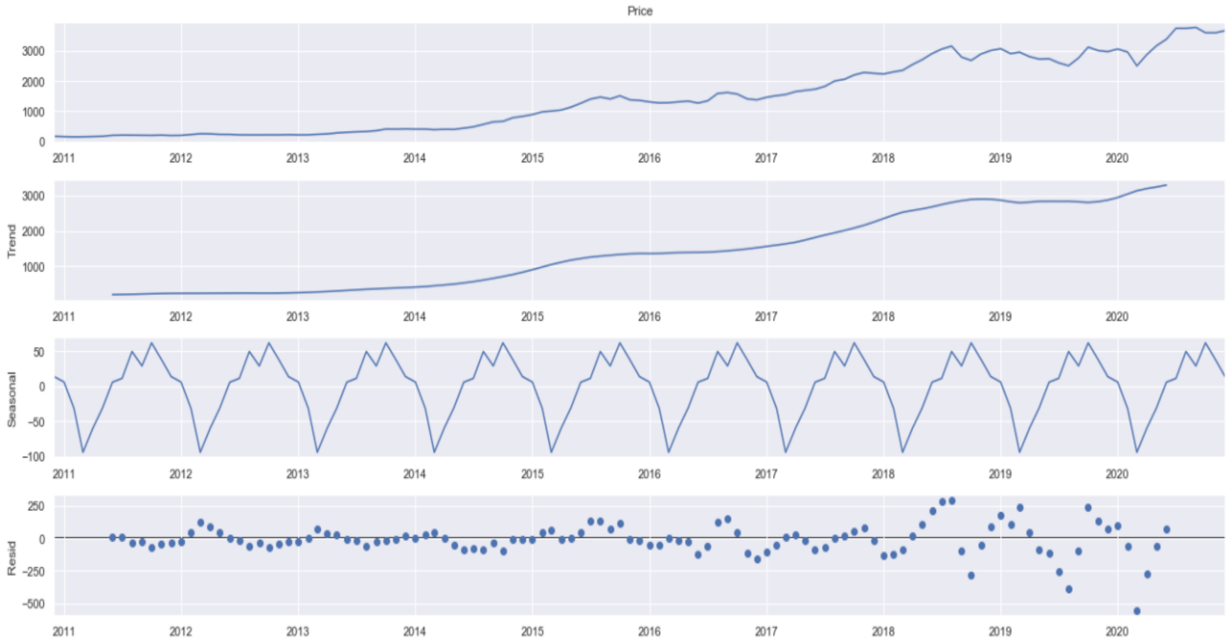


**Figure 6.17: Seasonality and Trend for Brittania**

ARIMA(1, 1, 1)x(1, 1, 1, 12)12 - AIC: 1178.0572352667546

| | coef | std err | z | P>\|z\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| ar.L1 | -0.6196 | 0.147 | -4.201 | 0.000 | -0.909 | -0.331 |
| ma.L1 | 0.8555 | 0.112 | 7.612 | 0.000 | 0.635 | 1.076 |
| ar.S.L12 | -0.5494 | 0.085 | -6.479 | 0.000 | -0.716 | -0.383 |
| ma.S.L12 | -0.7877 | 0.139 | -5.672 | 0.000 | -1.060 | -0.516 |
| sigma2 | 1.313e+04 | 1536.127 | 8.550 | 0.000 | 1.01e+04 | 1.61e+04 |

**Table 6.5: ARIMA Statistics for Brittania**


**Figure 6.18: Graph of Observed vs One Step Ahead Forecast for Brittania**

## ITC

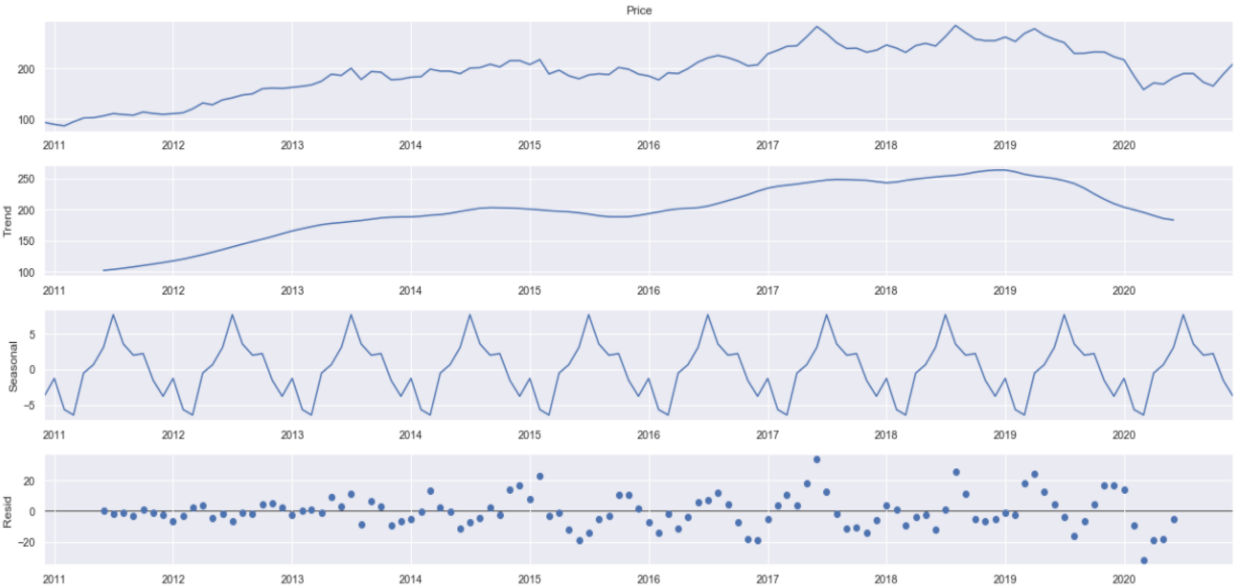


**Figure 6.19: Seasonality and Trend for ITC**

ARIMA(0, 1, 1)x(0, 1, 1, 12)12 - AIC:747.5239714846524

| | coef | std err | z | P>\|z\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| ma.L1 | 0.1038 | 0.090 | 1.158 | 0.247 | -0.072 | 0.279 |
| ma.S.L12 | -1.0000 | 7011.905 | -0.000 | 1.000 | -1.37e+04 | 1.37e+04 |
| sigma2 | 127.8254 | 8.96e+05 | 0.000 | 1.000 | -1.76e+06 | 1.76e+06 |

**Table 6.6: ARIMA Statistics for ITC**

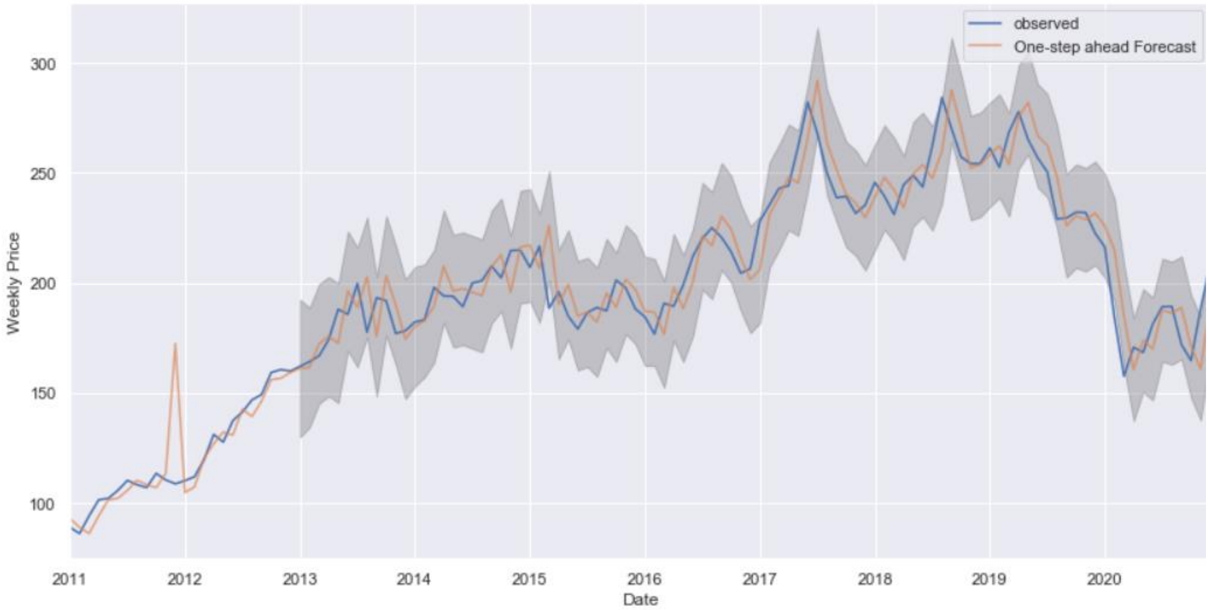


**Figure 6.20: Graph of Observed vs One Step Ahead Forecast for ITC**

<u>**Infosys**</u>

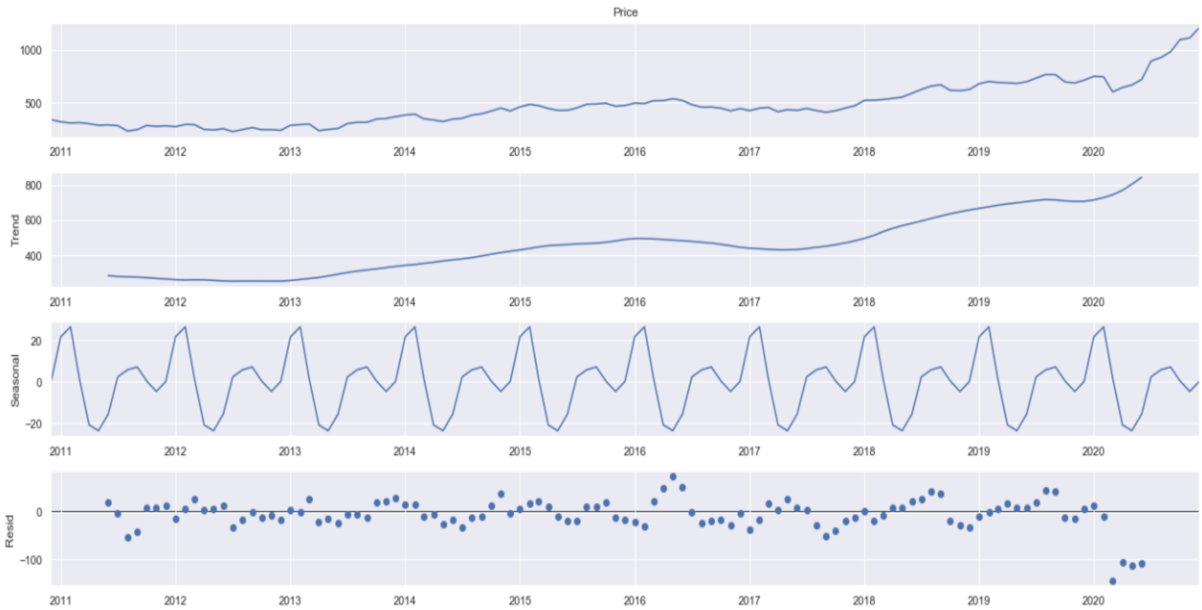


**Figure 6.21: Seasonality and Trend for Infosys**

ARIMA(1, 1, 1)x(0, 1, 1, 12)12 - AIC:952.0335929142682

| | coef | std err | z | P>\|z\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| ar.L1 | 0.8153 | 0.210 | 3.885 | 0.000 | 0.404 | 1.227 |
| ma.L1 | -0.6318 | 0.245 | -2.582 | 0.010 | -1.112 | -0.152 |
| ma.S.L12 | -0.6794 | 0.106 | -6.438 | 0.000 | -0.886 | -0.473 |
| sigma2 | 1307.1809 | 116.726 | 11.199 | 0.000 | 1078.401 | 1535.961 |

**Table 6.7: ARIMA Statistics for Infosys**

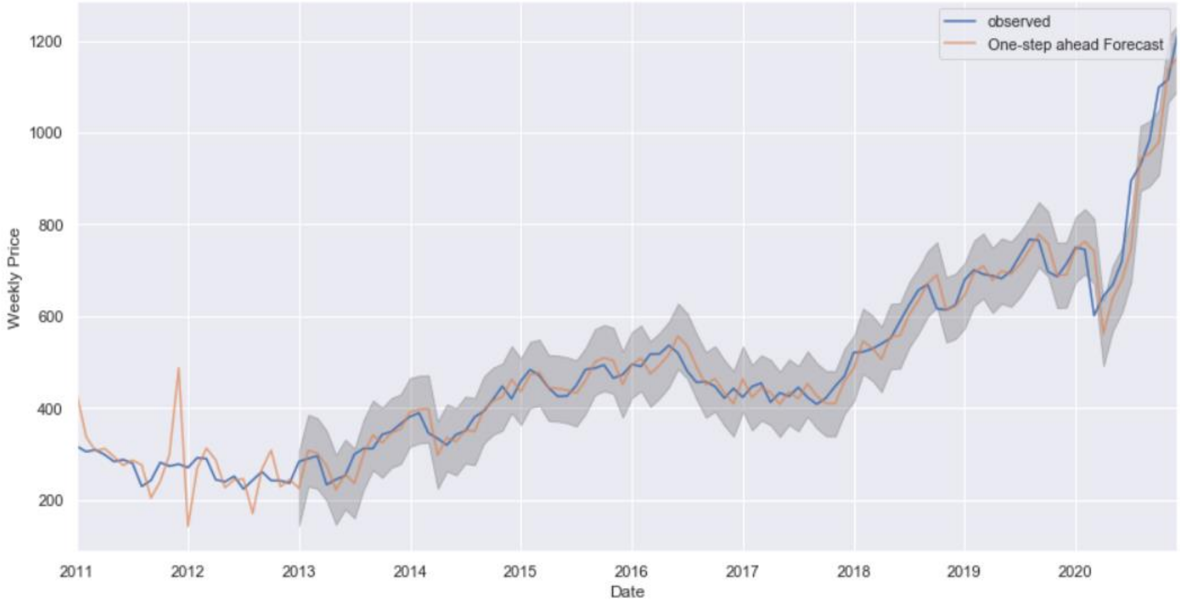


**Figure 6.22: Graph of Observed vs One Step Ahead Forecast for Infosys**

## Tata Consultancy Services

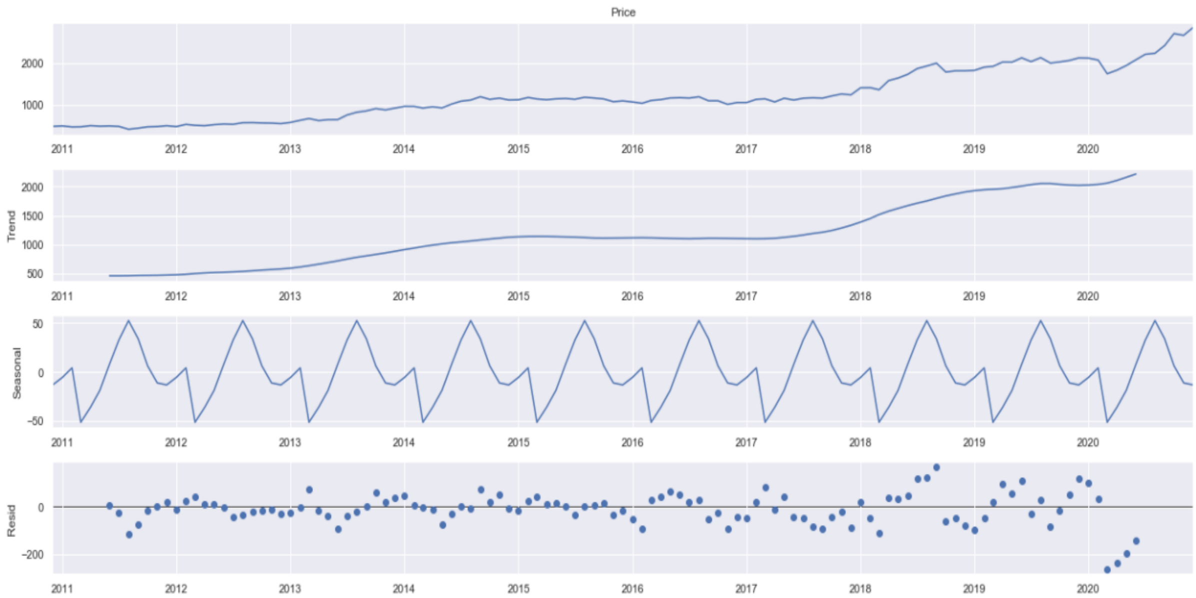


**Figure 6.23: Seasonality and Trend for Tata Consultancy Services**

ARIMA(0, 1, 1)x(0, 1, 1, 12)12 - AIC:1115.4382733388795

| | coef | std err | z | P>\|z\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| ma.L1 | -0.0048 | 0.105 | -0.045 | 0.964 | -0.210 | 0.201 |
| ma.S.L12 | -1.0565 | 0.587 | -1.800 | 0.072 | -2.207 | 0.094 |
| sigma2 | 6027.3225 | 4158.015 | 1.450 | 0.147 | -2122.237 | 1.42e+04 |

**Table 6.8: ARIMA Statistics for Tata Consultancy Services**


**Figure 6.24: Graph of Observed vs One Step Ahead Forecast for Tata Consultancy Services**

## Correlation Matrix

**HDFC Bank**


**Figure 6.25: Correlation Matrix of HDFC Bank**

There is a strong correlation(0.95) between price and year. There is a weak correlation(0.2) between volume and price

**Bajaj Finance**


**Figure 6.26: Correlation Matrix of Bajaj Finance**

There is a strong correlation(0.89) between price and year. There is also moderate correlation(0.48) between volume and price.

**JSW Steel**


**Figure 6.27: Correlation Matrix of JSW Steel**

There is a strong correlation(0.86) between price and year. There is also negative correlation(-0.41) between volume and price.

## Tata Steel


**Figure 6.28: Correlation Matrix of Tata Steel**

There is a moderate correlation(0.65) between volume and year. There is a weak to moderate correlation(0.39) between price and year.

## Brittania


**Figure 6.29: Correlation Matrix of Brittania**

There is a strong correlation(0.96) between price and year. There is a moderate correlation(0.52) between volume and price.

## ITC


**Figure 6.30: Correlation Matrix of ITC**

There is a strong correlation(0.73) between price and year.

## Infosys


**Figure 6.31: Correlation Matrix of Infosys**

There is a strong correlation(0.88) between price and year.

## Tata Consultancy Services


**Figure 6.32: Correlation Matrix of Tata Consultancy Services**

There is a strong correlation(0.88) between price and year.

## Comparative Study of Results

| | | Finance | | Metals | |
|---|---|---|---|---|---|
| | | HDFC Bank | Bajaj Finance | JSW Steel | Tata Steel |
| **ARIMA** | MSE | 1775.29 | 56993.97 | 244.29 | 2254.25 |
| | RMSE | 42.13 | 238.73 | 15.63 | 47.48 |
| | **Average Accuracy (%)** | **95.60** | **92.47** | **92.06** | **91.34** |
| **Linear Regression** | R Square Value | 0.91 | 0.78 | 0.73 | 0.15 |
| | MSE (Standarized) | 9617.91 | 347753.01 | 2133.26 | 10646.98 |
| | RMSE (Standarized) | 98.07 | 589.71 | 46.19 | 103.18 |
| | **Average Accuracy (%)** | **76.61** | **74.31** | **59.75** | **77.63** |
| **Random Forest** | R Square Value | 0.99 | 0.96 | 0.95 | 0.91 |
| | MSE (Standardized) | 584.11 | 23632.07 | 143.00 | 445.33 |
| | RMSE (Standardized) | 24.17 | 153.73 | 11.96 | 21.10 |
| | **Average Accuracy (%)** | **98.31** | **96.55** | **96.32** | **96.55** |
| | | **Consumer Goods** | | **IT** | |
| | | **Brittania** | **ITC** | **Infosys** | **Tata Consultancy Services** |
| **ARIMA** | MSE | 10732.60 | 158.21 | 1848.92 | 6828.61 |
| | RMSE | 103.60 | 12.58 | 43.00 | 82.64 |
| | **Average Accuracy (%)** | **94.31** | **95.33** | **92.72** | **94.84** |
| **Linear Regression** | R Square Value | 0.93 | 0.61 | 0.77 | 0.89 |
| | MSE (Standarized) | 89425.85 | 802.17 | 8464.19 | 38045.01 |
| | RMSE (Standarized) | 299.04 | 28.32 | 92.00 | 195.05 |
| | **Average Accuracy (%)** | **71.55** | **88.32** | **85.12** | **86.13** |
| **Random Forest** | R Square Value | 0.99 | 0.95 | 0.98 | 0.98 |
| | MSE (Standardized) | 4256.60 | 50.60 | 272.57 | 1996.88 |
| | RMSE (Standardized) | 65.24 | 7.11 | 16.51 | 44.69 |
| | **Average Accuracy (%)** | **97.82** | **97.68** | **98.06** | **98.06** |

**Table 6.9: Comparative Performance Analysis of Random Forest Model with Linear regression and ARIMA Across Sectors**

Finance Sector- For HDFC Bank, we see that the average accuracy in the case of ARIMA is 95.60% compared to 76.6% in LR and 98.31% in RF. For Bajaj Finance, we see that the average accuracy in the case of ARIMA is 92.47% compared to the 74.31% in LR and 96.55% of RF. Based on the result, the results from LR are far inferior compared to the other two algorithms. While the level of accuracy of RF and ARIMA is close, RF yields slightly better results.

Metals- For JSW Steel, we see that the average accuracy in the case of ARIMA is 92.06% compared to 59.75% in LR and 96.32% in RF. For Tata Steel, we see that the average accuracy in the case of ARIMA is 91.34% compared to the 77.63% in LR and 96.35% of RF. Based on the result, the results from LR are far inferior compared to the other two algorithms. While the level of accuracy of RF and ARIMA is close, RF yields slightly better results.

Consumer Goods- For Brittania, we see that the average accuracy in the case of ARIMA is 94.31% compared to the 71.55% in LR and 97.82% in RF. For ITC, we see that the average accuracy in the case of ARIMA is 95.33% compared to the 88.32% in LR and 97.68% of RF. Based on the result, the results from LR are far inferior compared to the other two algorithms. While the level of accuracy of RF and ARIMA is close, RF yields slightly better results.

IT- For Infosys, we see that the average accuracy in the case of ARIMA is 92.72% compared to 85.12% in LR and 98.06% in RF. For Tata Consultancy Services, we see that the average accuracy in the case of ARIMA is 94.84% compared to 86.13% in LR and 98.06% in RF. Based on the result, the results from LR are far inferior compared to the other two algorithms. While the level of accuracy of RF and ARIMA is close, RF yields slightly better results.

The results for LR are significantly better for IT companies as compared to other sectors. ARIMA worked best for the consumer goods sector and worst for the metals sector. RF performed well across all sectors.

From the above observations, we see that linear regression returns the worst results. Both ARIMA and Random Forest provide reasonably good results in predicting the stock prices. However, the results yielded by Random Forest are slightly better than ARIMA across all the sectors where the algorithm was applied.

## Part B- Comparison of Recurrent Reinforcement Learning Strategy in comparison with a Buy and Sold Strategy

Training Set – Weekly data from 01/01/2019 to 01/01/2021

Test Set- Weekly data from 01/01/2011 to 01/01/2021

### Finance Sector

**HDFC Bank**


**Figure 6.33: RRL Output for HDFC Bank**

**Bajaj Finance**

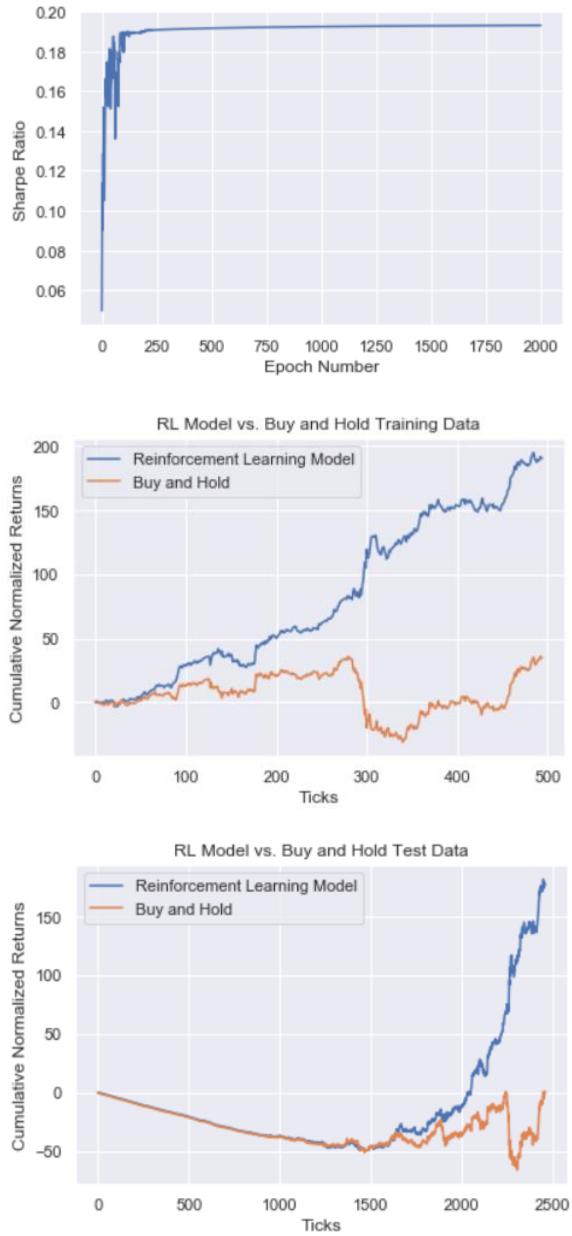


**Figure 6.34: RRL Output for Bajaj Finance**

## Metals

**JSW Steel**


**Figure 6.35: RRL Output for JSW Steel**

**Tata Steel**

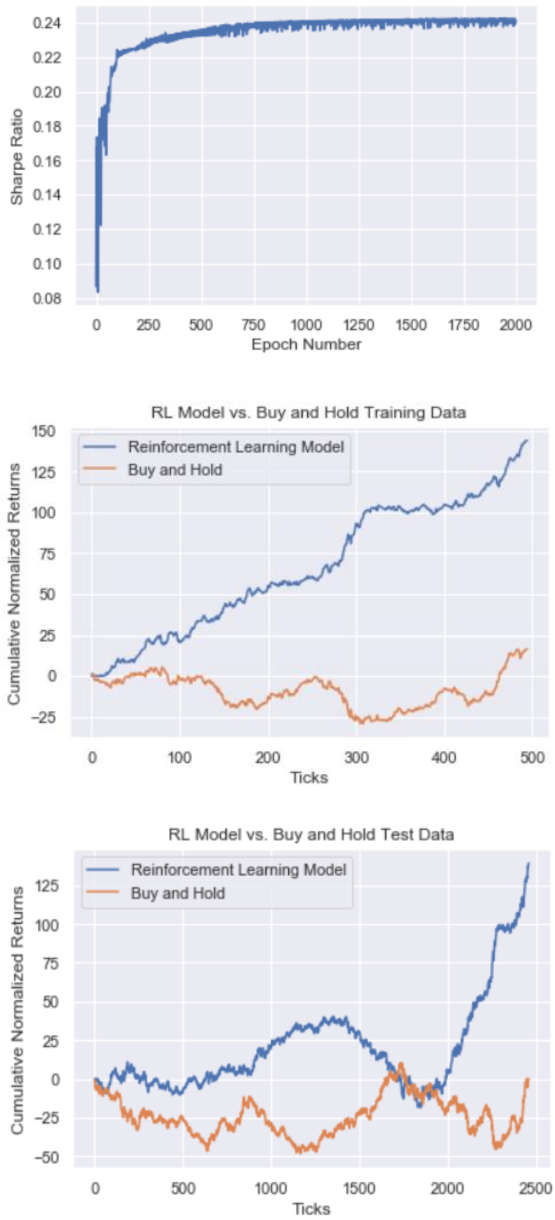


**Figure 6.36: RRL Output for Tata Steel**

## Consumer Goods

### Brittania

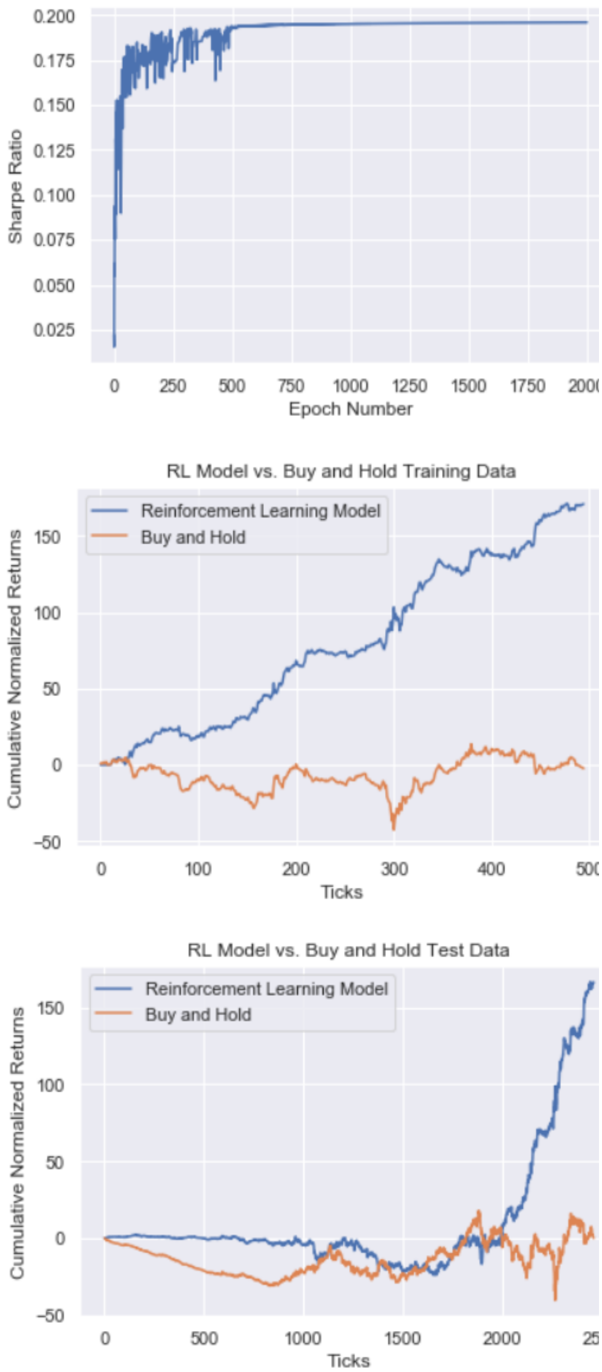


**Figure 6.37: RRL Output for Brittania**

### ITC

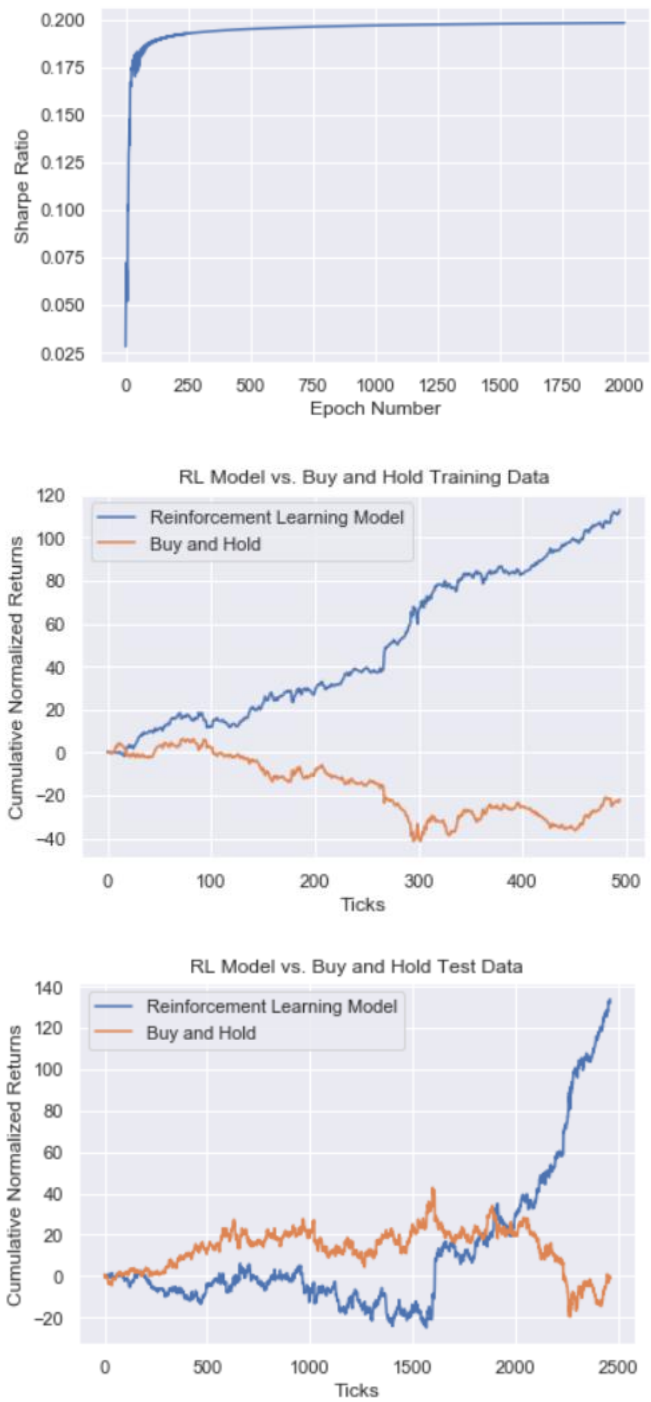


**Figure 6.38: RRL Output for ITC**

## IT

### Infosys


**Figure 6.39: RRL Output for Infosys**

### Tata Consultancy Services

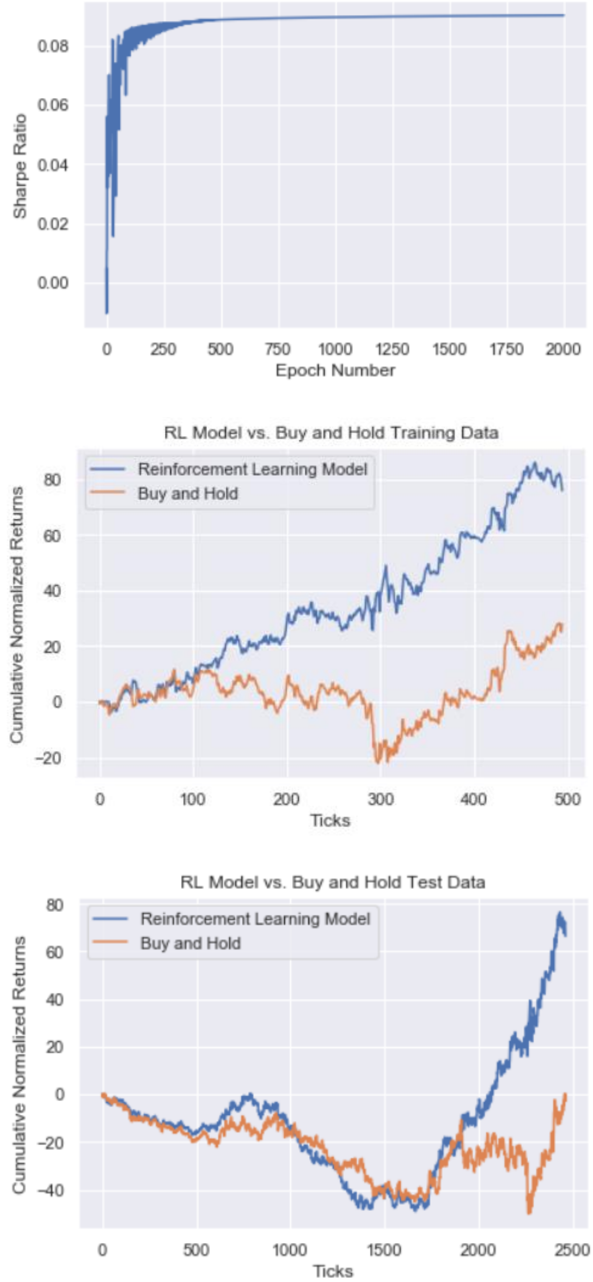


**Figure 6.40: RRL Output for Tata Consultancy Services**

## Comparative Study of Results

| | Finance | | Metals | | Consumer Goods | | IT | |
|---|---|---|---|---|---|---|---|---|
| | **HDFC Bank** | **Bajaj Finance** | **JSW Steel** | **Tata Steel** | **Brittania** | **ITC** | **Infosys** | **Tata Consultancy Services** |
| **Sharpe Ratio** | 0.271 | 0.192 | 0.187 | 0.242 | 0.195 | 0.202 | 0.178 | 0.092 |
| **Difference in Cumulative Normalized return for training set** | 231.74% | 155.85% | 129.81% | 127.40% | 173.51% | 134.76% | 96.70% | 47.99% |
| **Difference in Cumulative Normalized return for test set** | 233.73% | 177.30% | 52.94% | 139.28% | 166.41% | 133.91% | 108.68% | 66.30% |

**Table 6.10: Comparative Performance Analysis of RRL Model with Gradient Ascent vs. Buy& Hold Strategy Model Across Sectors**

Training Set – Weekly data from 01/01/2019 to 01/01/2021

Test Set- Weekly data from 01/01/2011 to 01/01/2021

We see that the Finance Sector companies have yielded the best results. In the training set, HDFC Bank yielded 231.74% more compared to 155.85% of Bajaj Finance. In the test set, the value for HDFC Bank was 1155.85% higher compared to 177.30% of Bajaj Finance

In the metals sector, although the difference in cumulative normalized returns are high, they are still lower than some of the other sectors. JSW steel performed slightly better with 129.81% more compared to the 127.4% of Tata Steel in the training. However, on the test set, the return for Tata Steel is 139.28% more which is significantly higher than the 52.94% of JSW steel. This could be due to the high volatility experienced in the stock prices of JSW Steel.

In the consumer goods segment, both Brittania and IC performed extremely well. In the training set, Brittania yielded 173.51% compared to the 134.6% of ITC. For the test set the value for Brittania was 166.4% more compared to the 133.91% of ITC.

The improvement in cumulative normalized return is least for the IT Sector. In the training set, Infosys yielded 96.70% compared to the 47.99% of TCS. For the test set the value for Infosys was 108.68% more compared to the 66.3% of TCS.

This would indicate that our model yields better results for the finance and Consumer goods sectors compared to the metals and IT sectors.

Based on the above results, it is evident that the cumulative normalized obtained through managing buy/sell decisions using Recurrent Reinforcement Learning(RRL) with gradient ascent is substantially higher than a buy and hold strategy for both the training and test set across the four sectors.

# Chapter 5: Conclusion

In this paper, we created a random forest model for predicting stock prices. The model was tested on eight companies belonging to four different sectors- finance, metals, consumer goods and IT. The accuracy of the results was compared with ARIMA and linear regression models. The performance of our RF model was substantially better than these models. While our RF model performed well across all sectors, ARIMA worked best for the consumer goods sector and worst for the metals sector, while LR worked well only for the IT stocks.

Traders often used forecasted stock prices to decide which stocks to buy and which ones to sell. However, timing the decision of buy or sell to maximise the profits is a complex task. In this paper, we have created a RRL with gradient ascent to help us make the buy or sell decisions. The model parameters which provided the best results were fixed after iterating through combinations of parameters.

To validate the performance of our model, we compared the results obtained for the eight previously chosen companies with a buy and hold strategy. Our model's cumulative normalized returns were significantly higher than those of the buy and hold strategy across different sectors. This indicates that our model's active buy/sell decisions yield better results than a buy and hold strategy. Therefore, it can be concluded that the usage of algorithmic strategies in stock market trading could improve the returns on investment.