# Data Understanding:

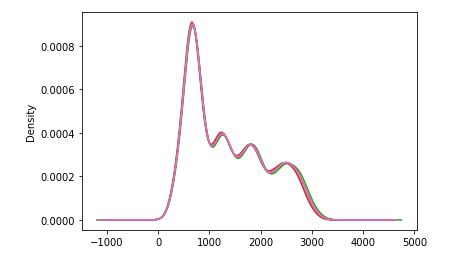


Table1: Data Distribution plot of Feature variables and Close price

The describe() function in pandas is very handy in getting various summary statistics. This function returns the count, mean, standard deviation, minimum and maximum values and the quantiles of the data.

1.The mean value is greater than median value of most of the feature variables which is represented by 50%(50th percentile) in index column meaning Data has positive skewed distribution.

2.There is notably a large difference between 75th %tile and max values of most of the feature variables. Thus observations 1 and 2 suggests that there are extreme values-Outliers in our data set.

# Data Preparation

## Classification Modelling on close price:

6-day consecutive closing price for the stock under consideration is being taken. These 6 days consecutive closing prices will be tabulated week on week for the entire dataset and will be utilized as 6 different feature variables for building the classification Model.

The difference between 7th and 8th day Closing price is determined. If the 8th day closing price is seen an increase from the 7th day by 0.7% or more, the direction of the closing price can be made as positive.

If the 8th day closing price is seen a decrease from the 7th day by -0.7% or less, the direction of the closing price can be made as negative. Between -0.7% and 0.7%, the direction of the closing price for the stock under consideration can be treated as sideways.

For data within the 0.7% and -0.7% band, usually the advice to the investor will be to hold on to existing portfolios and wait for the direction of the closing price to show as either negative or positive change. If there is a negative change, usually the advice to the investor will be to not to invest in such a circumstance. If there is a positive change the investor will be suggested to invest.

It is to be determined how many times the positive changes are identified by predicting and how many times positive changes are there in the actual data. This will be utilized to evaluate how many times true positives were detected and how many times the false positives were predicted in the prediction. Similar process to be followed for detecting true negatives and false negatives. Similar process to be followed for detecting true neutrals and false neutrals. Based on prediction accuracy, it can be suggested whether to invest or not to invest to the prospective investor.

Computation is being done to evaluate whether it is positive change, negative change or no change between 7th and 8th day closing price. The rule is being set to determine as to what has to be seen as direction change.0.7% change,1% change and 1.5% change -these are different classes of direction for which rule is being set which is to be followed for computing the direction change as either positive change, negative change or no change.

once it is determined say for example 0.7% change has the best prediction accuracy among all different classes of direction namely 0.7% change,1% change and 1.5% change then the range of consecutive days to be utilized as feature variable is increased to 10 days. Therefore,10-day consecutive closing price for the stock under consideration is being taken. These 10 days consecutive closing prices will be tabulated week on week for the entire dataset and will be utilized as different feature variables for building the classification Model.

Similar process is again repeated for range of consecutive days to be utilized as feature variable increased to 14 days. The prediction accuracy is determined to confirm that say 0.7% change has the best prediction accuracy among all different classes of direction even when range of consecutive days to be utilized as feature variable is increased to 14 days.

## Technical Indicators

Similarly, all technical indicators can be utilized in Technical Analysis to build another sets of classification Models. All different types of technical indicators namely momentum indicators, trend indicators, volatility indicators, volume indicators can be utilized as feature variables based on the input dataset and different classification models can be built to determine their prediction accuracy.

Generally Open price, High price, low price, close price and volume for the stock under consideration will be utilized to derive feature variables from technical indicators. These derived feature variables will then be used as the feature variables to predict the direction of the close price. The Actual direction of the close price is estimated as percentage change of the close price between upper-band +0.5% and lower band -0.5% for all technical indicators-based classification Models.

# Data Evaluation

## Data Evaluation for HDFC Stock

## Direction Detection by 6,10,14 days consecutive closing prices split week on week:

**(0-Negative,1-Neutral,2-Positive)**

### **Logistic Regression Classifier**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Target Variable** | **precision** | **recall** | **f1-score** | **support** | **accuracy score** |
| percentage change between upper-band +0.7% and lower band -.07%- | | | | | |
| 0 | 0.36 | 0.24 | 0.29 | 528 | 0.35 |
| 1 | 0.34 | 0.21 | 0.26 | 584 |
| 2 | 0.35 | 0.60 | 0.44 | 559 |
| percentage change between upper-band +1.0% and lower band -.1.0% | | | | | |
| 0 | 0.00 | 0.00 | 0.00 | 432 | 0.44 |
| 1 | 0.44 | 1.00 | 0.61 | 741 |
| 2 | 0.00 | 0.00 | 0.00 | 498 |
| percentage change between upper-band +1.5% and lower band -.1.5% | | | | | |
| 0 | 0.00 | 0.00 | 0.00 | 294 | 0.61 |
| 1 | 0.61 | 1.00 | 0.76 | 1021 |
| 2 | 0.00 | 0.00 | 0.00 | 356 |

From Table, it can be observed that logistic regression modelling done for percentage change in close price between upper-band +1.5% and lower band -.1.5% has given the highest efficiency in prediction. However, it predicts only neutral direction with 0.61 precision but it has 0.00 precision for predicting upward or downward trend. Hence, Logistic Regression Modelling results can be ignored.

### **Decision Tree Classifier**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Target Variable** | **precision** | recall | f1-score | support | accuracy score |
| percentage change between upper-band +0.7% and lower band -.07% | | | | | |
| 0 | 0.33 | 0.15 | 0.21 | 536 | 0.37 |
| 1 | 0.38 | 0.38 | 0.38 | 543 |
| 2 | 0.38 | 0.56 | 0.45 | 590 |
| percentage change between upper-band +1.0% and lower band -.1.0% | | | | | |
| 0 | 0.75 | 0.01 | 0.01 | 425 | 0.46 |
| 1 | 0.45 | 1.00 | 0.62 | 759 |
| 2 | 0.00 | 0.00 | 0.00 | 487 |
| percentage change between upper-band +1.5% and lower band -.1.5% | | | | | |
| 0 | 0.00 | 0.00 | 0.00 | 234 | 0.69 |
| 1 | 0.69 | 1.00 | 0.82 | 1103 |
| 2 | 0.00 | 0.00 | 0.00 | 253 |

From Table, it can be observed that logistic regression modelling done for percentage change in close price between upper-band +1.5% and lower band -.1.5% has given the highest efficiency in prediction. However, it predicts only neutral direction with 0.69 precision but it has 0.00 precision for predicting upward or downward trend. Hence, Decision Tree Modelling results can be ignored.

### **Random Forest Classifier**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Target Variable** | **precision** | **recall** | **f1-score** | **support** | **accuracy score** |
| percentage change between upper-band +0.7% and lower band -.07% (6 days consecutive closing prices split week on week) | | | | | |
| **0** | **0.91** | **0.81** | **0.86** | **544** | **0.87** |
| **1** | **0.85** | **0.90** | **0.88** | **580** |
| **2** | **0.85** | **0.89** | **0.87** | **547** |
| percentage change between upper-band +0.7% and lower band -.07%  (10 days consecutive closing prices split week on week) | | | | | |
| **0** | **0.87** | **0.86** | **0.87** | **559** | **0.87** |
| **1** | **0.87** | **0.87** | **0.87** | **550** |
| **2** | **0.87** | **0.88** | **0.87** | **561** |
| percentage change between upper-band +0.7% and lower band -.07%  (14 days consecutive closing prices split week on week) | | | | | |
| **0** | **0.80** | **0.77** | **0.79** | **536** | **0.80** |
| **1** | **0.79** | **0.81** | **0.80** | **543** |
| **2** | **0.80** | **0.81** | **0.80** | **590** |
| percentage change between upper-band +1.0% and lower band -.1.0% | | | | | |
| 0 | 0.90 | 0.09 | 0.16 | 425 | 0.53 |
| 1 | 0.50 | 0.97 | 0.66 | 759 |
| 2 | 0.63 | 0.22 | 0.32 | 487 |
| percentage change between upper-band +1.5% and lower band -.1.5% | | | | | |
| 0 | 1.00 | 0.02 | 0.03 | 234 | 0.70 |
| 1 | 0.70 | 1.00 | 0.82 | 1103 |
| 2 | 0.90 | 0.04 | 0.07 | 253 |

From Table, it can be observed that random forest modelling done for percentage change in close price between upper-band +0.7% and lower band -.0.7% has given considerable efficiency in prediction.

### **K Neighbours Classifier**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Target Variable** | **precision** | **recall** | **f1-score** | **support** | **accuracy score** |
| percentage change between upper-band +0.7% and lower band -.07% | | | | | |
| 0 | 0.38 | 0.37 | 0.38 | 384 | 0.38 |
| 1 | 0.41 | 0.46 | 0.43 | 386 |
| 2 | 0.35 | 0.31 | 0.33 | 344 |
| percentage change between upper-band +1.0% and lower band -.1.0% | | | | | |
| 0 | 0.32 | 0.25 | 0.28 | 313 | 0.42 |
| 1 | 0.50 | 0.67 | 0.57 | 521 |
| 2 | 0.27 | 0.17 | 0.21 | 280 |
| percentage change between upper-band +1.5% and lower band -.1.5% | | | | | |
| 0 | 0.24 | 0.07 | 0.11 | 213 | 0.60 |
| 1 | 0.64 | 0.90 | 0.75 | 704 |
| 2 | 0.30 | 0.11 | 0.16 | 197 |

From Table, it can be observed that logistic regression modelling done for percentage change in close price between upper-band +1.5% and lower band -.1.5% has given the highest efficiency in prediction. However, it predicts only neutral direction with 0.64 precision but it has negligible precision for predicting upward or downward trend. Hence, K nearest Neighbour Modelling results can be ignored.

### **XG Boost Classifier**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Target Variable** | **precision** | **recall** | **f1-score** | **support** | **accuracy score** |
| percentage change between upper-band +0.7% and lower band -.07% | | | | | |
| 0 | 0.39 | 0.16 | 0.23 | 384 | 0.40 |
| 1 | 0.43 | 0.61 | 0.51 | 386 |
| 2 | 0.35 | 0.42 | 0.38 | 344 |
| percentage change between upper-band +1.0% and lower band -.1.0% | | | | | |
| 0 | 0.40 | 0.08 | 0.13 | 313 | 0.48 |
| 1 | 0.49 | 0.90 | 0.64 | 521 |
| 2 | 0.37 | 0.14 | 0.20 | 280 |
| percentage change between upper-band +1.5% and lower band -.1.5% | | | | | |
| 0 | 0.42 | 0.05 | 0.08 | 213 | 0.64 |
| 1 | 0.65 | 0.98 | 0.78 | 704 |
| 2 | 0.56 | 0.08 | 0.13 | 197 |

From Table, it can be observed that logistic regression modelling done for percentage change in close price between upper-band +1.5% and lower band -.1.5% has given the highest efficiency in prediction. However, it predicts only neutral direction with 0.65 precision but its precision for predicting upward or downward trend should have been still better. Hence, **XG Boost** Modelling results can be considered but with caution.

## **Go Long Direction Prediction using Technical Indicators**

**(0-Non positive,1-Positive)**

The direction of the close price is estimated as percentage change of the close price between upper-band +0.5% and lower band -0.5%-if the percentage change of the closing price is more than 0.5%, the direction of the closing price is treated as positive and suitable for long Trading in stock market. Otherwise, the direction of the close price is treated as non-positive and not suitable for long Trading in stock market.

### **Logistic Regression Classifier**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Target Variable** | **precision** | **recall** | **f1-score** | **support** | **accuracy score** | **Roc AUC score** |
| Volume Indicators as Feature Variables | | | | | |  |
| **0** | **0.90** | **0.99** | **0.94** | **658** | **0.92** | **0.91** |
| **1** | **0.98** | **0.83** | **0.90** | **452** |
| Momentum Indicators as Feature Variables | | | | | |  |
| **0** | **0.79** | **0.84** | **0.81** | **685** | **0.76** | **0.74** |
| **1** | **0.71** | **0.63** | **0.67** | **423** |
| Trend Indicators as Feature Variables | | | | | |  |
| 0 | 0.78 | 0.92 | 0.85 | 679 | 0.80 | 0.76 |
| 1 | 0.83 | 0.59 | 0.69 | 431 |
| volatility Indicators as Feature Variables | | | | | |  |
| 0 | 0.73 | 0.98 | 0.84 | 658 | 0.77 | 0.73 |
| 1 | 0.93 | 0.47 | 0.63 | 452 |

From Table, it can be observed that logistic regression modelling done for percentage change in close price between upper-band +0.5% and lower band -.0.5% has given considerably good accuracy score for all technical categories of indicators namely Volume, momentum, trend and volatility. Precision and f1-score are also satisfactory. Recall can be improved further for trend indicators. ROC AUC score has been considerably satisfactory for all technical indicators.

### **Decision Tree Classifier**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Target Variable** | **precision** | **recall** | **f1-score** | **support** | **accuracy score** | **Roc AUC score** |
| Volume Indicators as Feature Variables | | | | | |  |
| 0 | 0.75 | 0.87 | 0.81 | 658 | 0.75 | 0.73 |
| 1 | 0.75 | 0.59 | 0.66 | 452 |
| Momentum Indicators as Feature Variables | | | | | |  |
| 0 | 0.75 | 0.82 | 0.78 | 685 | 0.72 | 0.69 |
| 1 | 0.66 | 0.55 | 0.60 | 423 |
| Trend Indicators as Feature Variables | | | | | |  |
| 0 | 0.72 | 0.75 | 0.73 | 679 | 0.66 | 0.64 |
| 1 | 0.57 | 0.53 | 0.55 | 431 |
| volatility Indicators as Feature Variables | | | | | |  |
| 0 | 0.70 | 0.80 | 0.75 | 658 | 0.68 | 0.65 |
| 1 | 0.63 | 0.51 | 0.56 | 452 |

From Table, it can be observed that Decision Tree modelling done for percentage change in close price between upper-band +0.5% and lower band -.0.5% has given considerably good accuracy score Volume indicators. Recall and accuracy can be improved further for trend and volatility indicators. ROC AUC score has been more than 50% for all technical indicators.

### **Random Forest Classifier**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Target Variable** | **precision** | **recall** | **f1-score** | **support** | **accuracy score** | **Roc AUC score** |
| Volume Indicators as Feature Variables | | | | | |  |
| 0 | 0.82 | 0.96 | 0.89 | 658 | 0.85 | 0.83 |
| 1 | 0.93 | 0.69 | 0.79 | 452 |
| Momentum Indicators as Feature Variables | | | | | |  |
| 0 | 0.75 | 0.90 | 0.82 | 685 | 0.75 | 0.70 |
| 1 | 0.76 | 0.51 | 0.61 | 423 |
| Trend Indicators as Feature Variables | | | | | |  |
| 0 | 0.77 | 0.95 | 0.85 | 679 | 0.80 | 0.75 |
| 1 | 0.87 | 0.56 | 0.68 | 431 |
| volatility Indicators as Feature Variables | | | | | |  |
| 0 | 0.75 | 0.97 | 0.84 | 658 | 0.79 | 0.75 |
| 1 | 0.92 | 0.53 | 0.67 | 452 |

From Table, it can be observed that Random Forest modelling done for percentage change in close price between upper-band +0.5% and lower band -.0.5% has given considerably good accuracy score for all technical indicators. Recall and accuracy can be improved further for all especially for predicting upward direction trend. ROC AUC score has been considerably satisfactory for all technical indicators.

### **K Neighbours Classifier**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Target Variable** | **precision** | **recall** | **f1-score** | **support** | **accuracy score** | **Roc AUC score** |
| Volume Indicators as Feature Variables | | | | | |  |
| 0 | 0.61 | 0.89 | 0.72 | 658 | 0.60 | 0.83 |
| 1 | 0.51 | 0.17 | 0.26 | 452 |
| Momentum Indicators as Feature Variables | | | | | |  |
| 0 | 0.68 | 0.87 | 0.76 | 685 | 0.67 | 0.70 |
| 1 | 0.62 | 0.34 | 0.43 | 423 |
| Trend Indicators as Feature Variables | | | | | |  |
| 0 | 0.62 | 0.87 | 0.73 | 679 | 0.60 | 0.75 |
| 1 | 0.45 | 0.16 | 0.24 | 431 |
| volatility Indicators as Feature Variables | | | | | |  |
| 0 | 0.60 | 0.88 | 0.71 | 658 | 0.59 | 0.75 |
| 1 | 0.47 | 0.16 | 0.24 | 452 |

From Table, it can be observed that K nearest neighbour modelling done for percentage change in close price between upper-band +0.5% and lower band -.0.5% can be improved further for accuracy score for all technical indicators. ROC AUC score has been considerably satisfactory for all technical indicators.

### **XG Boost Classifier**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Target Variable** | **precision** | **recall** | **f1-score** | **support** | **accuracy score** | **Roc AUC score** |
| Volume Indicators as Feature Variables | | | | | |  |
| 0 | 0.84 | 0.95 | 0.89 | 658 | 0.86 | 0.83 |
| 1 | 0.90 | 0.73 | 0.81 | 452 |
| Momentum Indicators as Feature Variables | | | | | |  |
| 0 | 0.78 | 0.83 | 0.80 | 685 | 0.75 | 0.70 |
| 1 | 0.70 | 0.61 | 0.65 | 423 |
| Trend Indicators as Feature Variables | | | | | |  |
| **0** | **0.81** | **0.92** | **0.86** | **679** | **0.82** | **0.75** |
| **1** | **0.85** | **0.65** | **0.74** | **431** |
| volatility Indicators as Feature Variables | | | | | |  |
| **0** | **0.81** | **0.91** | **0.86** | **658** | **0.82** | **0.75** |
| **1** | **0.84** | **0.69** | **0.76** | **452** |

From Table, it can be observed that XG Boost modelling done for percentage change in close price between upper-band +0.5% and lower band -.0.5% has given considerably good accuracy score for all technical categories of indicators namely Volume, momentum, trend and volatility. Precision and f1-score are also satisfactory. Recall can be improved further for trend indicators. ROC AUC score has been considerably satisfactory for all technical indicators.

## **Go Short Direction Prediction using Technical Indicators**

**(0-Negative,1-non-Negative)**

The direction of the close price is estimated as percentage change of the close price between upper-band +0.5% and lower band -0.5%-if the percentage change of the closing price is less than -0.5%, the direction of the closing price is treated as Negative and suitable for Short Trading in stock market. Otherwise, the direction of the close price is treated as non-negative and not suitable for Short Trading in stock market.

### **Logistic Regression Classifier**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Target Variable** | **precision** | **recall** | **f1-score** | **support** | **accuracy score** | **Roc AUC score** |
| Volume Indicators as Feature Variables | | | | | |  |
| **0** | **0.97** | **0.83** | **0.90** | **399** | **0.93** | **0.91** |
| **1** | **0.91** | **0.99** | **0.95** | **711** |
| Momentum Indicators as Feature Variables | | | | | |  |
| 0 | 0.70 | 0.59 | 0.64 | 394 | 0.76 | 0.73 |
| 1 | 0.79 | 0.86 | 0.82 | 714 |
| Trend Indicators as Feature Variables | | | | | |  |
| **0** | **0.91** | **0.56** | **0.69** | **414** | **0.81** | **0.76** |
| **1** | **0.79** | **0.97** | **0.87** | **696** |
| volatility Indicators as Feature Variables | | | | | |  |
| 0 | 0.89 | 0.44 | 0.59 | 399 | 0.78 | 0.70 |
| 1 | 0.75 | 0.97 | 0.85 | 711 |

From Table, it can be observed that logistic regression modelling done for percentage change in close price between upper-band +0.5% and lower band -.0.5% has given considerably good accuracy score for all technical categories of indicators namely Volume, momentum, trend and volatility. Precision and f1-score are also satisfactory. Recall can be improved further for trend indicators. ROC AUC score has been considerably satisfactory for all technical indicators.

### **Decision Tree Classifier**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Target Variable** | **precision** | **recall** | **f1-score** | **support** | **accuracy score** | **Roc AUC score** |
| Volume Indicators as Feature Variables | | | | | |  |
| 0 | 0.66 | 0.67 | 0.67 | 399 | 0.76 | 0.74 |
| 1 | 0.81 | 0.81 | 0.81 | 711 |
| Momentum Indicators as Feature Variables | | | | | |  |
| 0 | 0.58 | 0.55 | 0.56 | 394 | 0.70 | 0.66 |
| 1 | 0.76 | 0.78 | 0.77 | 714 |
| Trend Indicators as Feature Variables | | | | | |  |
| 0 | 0.55 | 0.44 | 0.49 | 414 | 0.66 | 0.61 |
| 1 | 0.70 | 0.79 | 0.74 | 696 |
| volatility Indicators as Feature Variables | | | | | |  |
| 0 | 0.56 | 0.43 | 0.49 | 399 | 0.67 | 0.62 |
| 1 | 0.72 | 0.81 | 0.76 | 711 |

From Table, it can be observed that Decision Tree modelling done for percentage change in close price between upper-band +0.5% and lower band -.0.5% has given considerably good accuracy score for volume and momentum indicators. Precision for predicting downward trend can be further improved. ROC AUC score has been more than 50% for all technical indicators.

### **Random Forest Classifier**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Target Variable** | **precision** | **recall** | **f1-score** | **support** | **accuracy score** | **Roc AUC score** |
| Volume Indicators as Feature Variables | | | | | |  |
| 0 | 0.87 | 0.71 | 0.78 | 399 | 0.85 | 0.82 |
| 1 | 0.85 | 0.94 | 0.89 | 711 |
| Momentum Indicators as Feature Variables | | | | | |  |
| 0 | 0.72 | 0.50 | 0.59 | 394 | 0.75 | 0.70 |
| 1 | 0.76 | 0.89 | 0.82 | 714 |
| Trend Indicators as Feature Variables | | | | | |  |
| 0 | 0.87 | 0.46 | 0.60 | 414 | 0.77 | 0.71 |
| 1 | 0.75 | 0.96 | 0.84 | 696 |
| volatility Indicators as Feature Variables | | | | | |  |
| **0** | **0.88** | **0.55** | **0.68** | **399** | **0.81** | **0.76** |
| **1** | **0.79** | **0.96** | **0.87** | **711** |

From Table, it can be observed that Random Forest modelling done for percentage change in close price between upper-band +0.5% and lower band -.0.5% has given considerably good accuracy score for all technical indicators. Recall and accuracy can be improved further for all especially for recalling downward direction trend. ROC AUC score has been considerably satisfactory for all technical indicators.

### **K Neighbours Classifier**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Target Variable** | **precision** | **recall** | **f1-score** | **support** | **accuracy score** | **Roc AUC score** |
| Volume Indicators as Feature Variables | | | | | |  |
| 0 | 0.45 | 0.42 | 0.44 | 399 | 0.61 | 0.82 |
| 1 | 0.69 | 0.71 | 0.70 | 711 |
| Momentum Indicators as Feature Variables | | | | | |  |
| 0 | 0.53 | 0.54 | 0.53 | 394 | 0.66 | 0.70 |
| 1 | 0.74 | 0.73 | 0.74 | 714 |
| Trend Indicators as Feature Variables | | | | | |  |
| 0 | 0.44 | 0.37 | 0.40 | 414 | 0.59 | 0.71 |
| 1 | 0.66 | 0.72 | 0.69 | 696 |
| volatility Indicators as Feature Variables | | | | | |  |
| 0 | 0.44 | 0.43 | 0.43 | 399 | 0.60 | 0.76 |
| 1 | 0.68 | 0.69 | 0.69 | 711 |

From Table, it can be observed that K nearest neighbour modelling done for percentage change in close price between upper-band +0.5% and lower band -.0.5% can be improved further for accuracy score for all technical indicators. ROC AUC score has been considerably satisfactory for all technical indicators.

### **XG Boost Classifier**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Target Variable** | **precision** | **recall** | **f1-score** | **support** | **accuracy score** | **Roc AUC score** |
| Volume Indicators as Feature Variables | | | | | |  |
| 0 | 0.86 | 0.79 | 0.82 | 399 | 0.88 | 0.82 |
| 1 | 0.89 | 0.93 | 0.91 | 711 |
| Momentum Indicators as Feature Variables | | | | | |  |
| **0** | **0.72** | **0.59** | **0.64** | **394** | **0.77** | **0.70** |
| **1** | **0.79** | **0.87** | **0.83** | **714** |
| Trend Indicators as Feature Variables | | | | | |  |
| 0 | 0.84 | 0.60 | 0.70 | 414 | 0.81 | 0.71 |
| 1 | 0.80 | 0.93 | 0.86 | 696 |
| volatility Indicators as Feature Variables | | | | | |  |
| 0 | 0.79 | 0.64 | 0.71 | 399 | 0.81 | 0.76 |
| 1 | 0.82 | 0.91 | 0.86 | 711 |

From Table, it can be observed that XG Boost modelling done for percentage change in close price between upper-band +0.5% and lower band -.0.5% has given considerably good accuracy score for all technical categories of indicators namely Volume, momentum, trend and volatility. Precision and f1-score are also satisfactory. Recall can be improved further for recalling downward trend direction. ROC AUC score has been considerably satisfactory for all technical indicators.

## Data Evaluation for KOTAK Stock

## Direction Detection by 6,10,14 days consecutive closing prices split week on week:

**(0-Negative,1-Neutral,2-Positive)**

### **Logistic Regression Classifier**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Target Variable** | **precision** | recall | f1-score | support | accuracy score |
| percentage change between upper-band +0.7% and lower band -.07%- | | | | | |
| 0 | 0.00 | 0.00 | 0.00 | 526 | 0.36 |
| 1 | 0.35 | 0.33 | 0.34 | 493 |
| 2 | 0.37 | 0.74 | 0.49 | 556 |
| percentage change between upper-band +1.0% and lower band -.1.0% | | | | | |
| 0 | 0.00 | 0.00 | 0.00 | 420 | 0.44 |
| 1 | 0.44 | 1.00 | 0.61 | 687 |
| 2 | 0.00 | 0.00 | 0.00 | 468 |
| percentage change between upper-band +1.5% and lower band -.1.5% | | | | | |
| 0 | 0.00 | 0.00 | 0.00 | 327 | 0.58 |
| 1 | 0.58 | 1.00 | 0.74 | 919 |
| 2 | 0.00 | 0.00 | 0.00 | 329 |

### **Decision Tree Classifier**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Target Variable** | **precision** | recall | f1-score | support | accuracy score |
| percentage change between upper-band +0.7% and lower band -.07% | | | | | |
| 0 | 0.37 | 0.20 | 0.26 | 506 | 0.37 |
| 1 | 0.40 | 0.47 | 0.43 | 535 |
| 2 | 0.35 | 0.44 | 0.39 | 532 |
| percentage change between upper-band +1.0% and lower band -.1.0% | | | | | |
| 0 | 0.30 | 0.21 | 0.25 | 408 | 0.43 |
| 1 | 0.45 | 0.83 | 0.59 | 681 |
| 2 | 0.44 | 0.05 | 0.09 | 486 |
| percentage change between upper-band +1.5% and lower band -.1.5% | | | | | |
| 0 | 0.29 | 0.01 | 0.01 | 314 | 0.59 |
| 1 | 0.60 | 0.98 | 0.75 | 937 |
| 2 | 0.29 | 0.04 | 0.07 | 324 |

### **Random Forest Classifier**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Target Variable** | **precision** | recall | f1-score | support | accuracy score |
| percentage change between upper-band +0.7% and lower band -.07%  (6 days consecutive closing prices split week on week) | | | | | |
| **0** | **0.77** | **0.70** | **0.73** | **517** | **0.74** |
| **1** | **0.75** | **0.74** | **0.75** | **532** |
| **2** | **0.71** | **0.79** | **0.75** | **526** |
| percentage change between upper-band +0.7% and lower band -.07%  (10 days consecutive closing prices split week on week) | | | | | |
| **0** | **0.86** | **0.81** | **0.83** | **533** | **0.81** |
| **1** | **0.76** | **0.79** | **0.78** | **491** |
| **2** | **0.80** | **0.82** | **0.81** | **550** |
| percentage change between upper-band +0.7% and lower band -.07%  (14 days consecutive closing prices split week on week) | | | | | |
| **0** | **0.88** | **0.85** | **0.86** | **506** | **0.86** |
| **1** | **0.87** | **0.87** | **0.87** | **535** |
| **2** | **0.84** | **0.87** | **0.86** | **532** |
| percentage change between upper-band +1.0% and lower band -.1.0% | | | | | |
| 0 | 0.84 | 0.54 | 0.66 | 408 | 0.73 |
| 1 | 0.67 | 0.93 | 0.78 | 681 |
| 2 | 0.83 | 0.62 | 0.71 | 486 |
| percentage change between upper-band +1.5% and lower band -.1.5% | | | | | |
| 0 | 0.82 | 0.18 | 0.30 | 314 | 0.67 |
| 1 | 0.66 | 0.98 | 0.79 | 937 |
| 2 | 0.66 | 0.23 | 0.35 | 324 |

### **K Neighbours Classifier**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Target Variable** | **precision** | recall | f1-score | support | accuracy score |
| percentage change between upper-band +0.7% and lower band -.07% | | | | | |
| 0 | 0.39 | 0.38 | 0.38 | 363 | 0.37 |
| 1 | 0.37 | 0.47 | 0.41 | 330 |
| 2 | 0.36 | 0.27 | 0.31 | 357 |
| percentage change between upper-band +1.0% and lower band -.1.0% | | | | | |
| 0 | 0.35 | 0.29 | 0.32 | 303 | 0.44 |
| 1 | 0.50 | 0.70 | 0.58 | 451 |
| 2 | 0.33 | 0.18 | 0.24 | 296 |
| percentage change between upper-band +1.5% and lower band -.1.5% | | | | | |
| 0 | 0.36 | 0.15 | 0.21 | 231 | 0.56 |
| 1 | 0.61 | 0.90 | 0.73 | 598 |
| 2 | 0.25 | 0.08 | 0.12 | 221 |

### **XG Boost Classifier**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Target Variable** | **precision** | recall | f1-score | support | accuracy score |
| percentage change between upper-band +0.7% and lower band -.07% | | | | | |
| 0 | 0.41 | 0.29 | 0.34 | 363 | 0.40 |
| 1 | 0.40 | 0.50 | 0.44 | 330 |
| 2 | 0.38 | 0.41 | 0.40 | 357 |
| percentage change between upper-band +1.0% and lower band -.1.0% | | | | | |
| 0 | 0.36 | 0.16 | 0.22 | 303 | 0.44 |
| 1 | 0.46 | 0.81 | 0.59 | 451 |
| 2 | 0.36 | 0.16 | 0.22 | 296 |
| percentage change between upper-band +1.5% and lower band -.1.5% | | | | | |
| 0 | 0.38 | 0.06 | 0.10 | 231 | 0.57 |
| 1 | 0.58 | 0.96 | 0.73 | 598 |
| 2 | 0.33 | 0.05 | 0.09 | 221 |

## **Go Long Direction Prediction using Technical Indicators**

**(0-Non positive,1-Positive)**

The direction of the close price is estimated as percentage change of the close price between upper-band +0.5% and lower band -0.5%-if the percentage change of the closing price is more than 0.5%, the direction of the closing price is treated as positive and suitable for long Trading in stock market. Otherwise, the direction of the close price is treated as non-positive and not suitable for long Trading in stock market.

### **Logistic Regression Classifier**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Target Variable** | **precision** | recall | f1-score | support | accuracy score | Roc AUC score |
| Volume Indicators as Feature Variables | | | | | |  |
| **0** | **0.96** | **0.99** | **0.98** | **632** | **0.97** | **0.96** |
| **1** | **0.99** | **0.93** | **0.96** | **414** |
| Momentum Indicators as Feature Variables | | | | | |  |
| **0** | **0.77** | **0.85** | **0.81** | **630** | **0.75** | **0.73** |
| **1** | **0.73** | **0.61** | **0.66** | **414** |
| Trend Indicators as Feature Variables | | | | | |  |
| 0 | 0.71 | 0.89 | 0.79 | 619 | 0.72 | 0.69 |
| 1 | 0.76 | 0.48 | 0.59 | 427 |
| volatility Indicators as Feature Variables | | | | | |  |
| 0 | 0.71 | 0.97 | 0.82 | 632 | 0.74 | 0.68 |
| 1 | 0.90 | 0.40 | 0.55 | 414 |

### **Decision Tree Classifier**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Target Variable** | **precision** | recall | f1-score | support | accuracy score | Roc AUC score |
| Volume Indicators as Feature Variables | | | | | |  |
| 0 | 0.85 | 0.93 | 0.89 | 632 | 0.85 | 0.84 |
| 1 | 0.87 | 0.75 | 0.80 | 414 |
| Momentum Indicators as Feature Variables | | | | | |  |
| 0 | 0.70 | 0.92 | 0.79 | 630 | 0.71 | 0.65 |
| 1 | 0.75 | 0.39 | 0.51 | 414 |
| Trend Indicators as Feature Variables | | | | | |  |
| 0 | 0.66 | 0.76 | 0.71 | 619 | 0.63 | 0.60 |
| 1 | 0.55 | 0.44 | 0.49 | 427 |
| volatility Indicators as Feature Variables | | | | | |  |
| 0 | 0.70 | 0.74 | 0.72 | 632 | 0.65 | 0.63 |
| 1 | 0.56 | 0.52 | 0.54 | 414 |

### **Random Forest Classifier**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Target Variable** | **precision** | recall | f1-score | support | accuracy score | Roc AUC score |
| Volume Indicators as Feature Variables | | | | | |  |
| 0 | 0.87 | 0.96 | 0.91 | 632 | 0.89 | 0.87 |
| 1 | 0.92 | 0.79 | 0.85 | 414 |
| Momentum Indicators as Feature Variables | | | | | |  |
| 0 | 0.73 | 0.91 | 0.81 | 630 | 0.75 | 0.70 |
| 1 | 0.78 | 0.50 | 0.61 | 414 |
| Trend Indicators as Feature Variables | | | | | |  |
| 0 | 0.71 | 0.95 | 0.81 | 619 | 0.74 | 0.69 |
| 1 | 0.85 | 0.44 | 0.58 | 427 |
| volatility Indicators as Feature Variables | | | | | |  |
| 0 | 0.75 | 0.96 | 0.84 | 632 | 0.78 | 0.73 |
| 1 | 0.89 | 0.50 | 0.64 | 414 |

### **K Neighbours Classifier**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Target Variable** | **precision** | recall | f1-score | support | accuracy score | Roc AUC score |
| Volume Indicators as Feature Variables | | | | | |  |
| 0 | 0.64 | 0.88 | 0.74 | 632 | 0.63 | 0.87 |
| 1 | 0.59 | 0.25 | 0.35 | 414 |
| Momentum Indicators as Feature Variables | | | | | |  |
| 0 | 0.68 | 0.89 | 0.77 | 630 | 0.68 | 0.70 |
| 1 | 0.68 | 0.36 | 0.47 | 414 |
| Trend Indicators as Feature Variables | | | | | |  |
| 0 | 0.61 | 0.85 | 0.71 | 619 | 0.58 | 0.69 |
| 1 | 0.47 | 0.20 | 0.28 | 427 |
| volatility Indicators as Feature Variables | | | | | |  |
| 0 | 0.62 | 0.84 | 0.71 | 632 | 0.58 | 0.73 |
| 1 | 0.45 | 0.20 | 0.28 | 414 |

### **XG Boost Classifier**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Target Variable** | **precision** | recall | f1-score | support | accuracy score | Roc AUC score |
| Volume Indicators as Feature Variables | | | | | |  |
| 0 | 0.92 | 0.95 | 0.94 | 632 | 0.92 | 0.87 |
| 1 | 0.92 | 0.87 | 0.90 | 414 |
| Momentum Indicators as Feature Variables | | | | | |  |
| 0 | 0.76 | 0.86 | 0.81 | 630 | 0.75 | 0.70 |
| 1 | 0.74 | 0.59 | 0.66 | 414 |
| Trend Indicators as Feature Variables | | | | | |  |
| **0** | **0.77** | **0.91** | **0.84** | **619** | **0.79** | **0.69** |
| **1** | **0.82** | **0.61** | **0.70** | **427** |
| volatility Indicators as Feature Variables | | | | | |  |
| **0** | **0.79** | **0.91** | **0.84** | **632** | **0.79** | **0.73** |
| **1** | **0.81** | **0.63** | **0.71** | **414** |

## **Go Short Direction Prediction using Technical Indicators**

**(0-Negative,1-non-Negative)**

The direction of the close price is estimated as percentage change of the close price between upper-band +0.5% and lower band -0.5%-if the percentage change of the closing price is less than -0.5%, the direction of the closing price is treated as Negative and suitable for Short Trading in stock market. Otherwise, the direction of the close price is treated as non-negative and not suitable for Short Trading in stock market.

### **Logistic Regression Classifier**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Target Variable** | **precision** | recall | f1-score | support | accuracy score | Roc AUC score |
| Volume Indicators as Feature Variables | | | | | |  |
| 0 | 0.99 | 0.87 | 0.92 | 397 | 0.95 | 0.93 |
| 1 | 0.92 | 1.00 | 0.96 | 649 |
| Momentum Indicators as Feature Variables | | | | | |  |
| 0 | 0.66 | 0.57 | 0.61 | 368 | 0.74 | 0.70 |
| 1 | 0.78 | 0.84 | 0.81 | 676 |
| Trend Indicators as Feature Variables | | | | | |  |
| 0 | 0.81 | 0.41 | 0.54 | 394 | 0.74 | 0.68 |
| 1 | 0.73 | 0.94 | 0.82 | 652 |
| volatility Indicators as Feature Variables | | | | | |  |
| 0 | 0.86 | 0.30 | 0.44 | 397 | 0.72 | 0.63 |
| 1 | 0.69 | 0.97 | 0.81 | 649 |

### **Decision Tree Classifier**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Target Variable** | **precision** | recall | f1-score | support | accuracy score | Roc AUC score |
| Volume Indicators as Feature Variables | | | | | |  |
| 0 | 0.81 | 0.78 | 0.80 | 397 | 0.85 | 0.83 |
| 1 | 0.87 | 0.89 | 0.88 | 649 |
| Momentum Indicators as Feature Variables | | | | | |  |
| 0 | 0.56 | 0.51 | 0.53 | 368 | 0.68 | 0.64 |
| 1 | 0.74 | 0.78 | 0.76 | 676 |
| Trend Indicators as Feature Variables | | | | | |  |
| 0 | 0.56 | 0.59 | 0.57 | 394 | 0.67 | 0.65 |
| 1 | 0.74 | 0.72 | 0.73 | 652 |
| volatility Indicators as Feature Variables | | | | | |  |
| 0 | 0.55 | 0.53 | 0.54 | 397 | 0.65 | 0.63 |
| 1 | 0.72 | 0.73 | 0.72 | 649 |

### **Random Forest Classifier**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Target Variable** | **precision** | recall | f1-score | support | accuracy score | Roc AUC score |
| Volume Indicators as Feature Variables | | | | | |  |
| 0 | 0.92 | 0.81 | 0.86 | 397 | 0.90 | 0.88 |
| 1 | 0.89 | 0.96 | 0.92 | 649 |
| Momentum Indicators as Feature Variables | | | | | |  |
| 0 | 0.68 | 0.45 | 0.54 | 368 | 0.73 | 0.67 |
| 1 | 0.75 | 0.89 | 0.81 | 676 |
| Trend Indicators as Feature Variables | | | | | |  |
| 0 | 0.91 | 0.43 | 0.59 | 394 | 0.77 | 0.70 |
| 1 | 0.74 | 0.97 | 0.84 | 652 |
| volatility Indicators as Feature Variables | | | | | |  |
| 0 | 0.93 | 0.42 | 0.58 | 397 | 0.77 | 0.70 |
| 1 | 0.73 | 0.98 | 0.84 | 649 |

### **K Neighbours Classifier**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Target Variable** | **precision** | recall | f1-score | support | accuracy score | Roc AUC score |
| Volume Indicators as Feature Variables | | | | | |  |
| 0 | 0.51 | 0.41 | 0.45 | 397 | 0.63 | 0.88 |
| 1 | 0.68 | 0.76 | 0.72 | 649 |
| Momentum Indicators as Feature Variables | | | | | |  |
| 0 | 0.51 | 0.54 | 0.52 | 368 | 0.66 | 0.67 |
| 1 | 0.74 | 0.72 | 0.73 | 676 |
| Trend Indicators as Feature Variables | | | | | |  |
| 0 | 0.43 | 0.32 | 0.37 | 394 | 0.58 | 0.70 |
| 1 | 0.64 | 0.74 | 0.69 | 652 |
| volatility Indicators as Feature Variables | | | | | |  |
| 0 | 0.46 | 0.37 | 0.41 | 397 | 0.59 | 0.70 |
| 1 | 0.65 | 0.73 | 0.69 | 649 |

### **XG Boost Classifier**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Target Variable** | **precision** | recall | f1-score | support | accuracy score | Roc AUC score |
| Volume Indicators as Feature Variables | | | | | |  |
| **0** | **0.92** | **0.85** | **0.88** | **397** | **0.92** | **0.88** |
| **1** | **0.91** | **0.96** | **0.93** | **649** |
| Momentum Indicators as Feature Variables | | | | | |  |
| **0** | **0.67** | **0.55** | **0.60** | **368** | **0.74** | **0.67** |
| **1** | **0.78** | **0.85** | **0.81** | **676** |
| Trend Indicators as Feature Variables | | | | | |  |
| **0** | **0.84** | **0.54** | **0.66** | **394** | **0.79** | **0.70** |
| **1** | **0.77** | **0.94** | **0.85** | **652** |
| volatility Indicators as Feature Variables | | | | | |  |
| **0** | **0.85** | **0.53** | **0.65** | **397** | **0.78** | **0.70** |
| **1** | **0.77** | **0.94** | **0.84** | **649** |

## Data Evaluation for SBI Stock

## Direction Detection by 6,10,14 days consecutive closing prices split week on week:

### **Logistic Regression Classifier**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Target Variable** | **precision** | recall | f1-score | support | accuracy score |
| percentage change between upper-band +0.7% and lower band -.07%- | | | | | |
| 0 | 0.50 | 0.00 | 0.00 | 558 | 0.36 |
| 1 | 0.00 | 0.00 | 0.00 | 506 |
| 2 | 0.36 | 1.00 | 0.53 | 607 |
| percentage change between upper-band +1.0% and lower band -.1.0% | | | | | |
| 0 | 0.00 | 0.00 | 0.00 | 476 | 0.41 |
| 1 | 0.41 | 1.00 | 0.59 | 694 |
| 2 | 0.00 | 0.00 | 0.00 | 501 |
| percentage change between upper-band +1.5% and lower band -.1.5% | | | | | |
| 0 | 0.00 | 0.00 | 0.00 | 345 | 0.57 |
| 1 | 0.57 | 1.00 | 0.72 | 947 |
| 2 | 0.00 | 0.00 | 0.00 | 379 |

### **Decision Tree Classifier**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Target Variable** | **precision** |  | recall | f1-score | support | accuracy score |
|  | percentage change between upper-band +0.7% and lower band -.07% | | | | | |
| 0 | 0.40 |  | 0.21 | 0.27 | 571 | 0.38 |
| 1 | 0.35 |  | 0.34 | 0.34 | 508 |
| 2 | 0.38 |  | 0.57 | 0.46 | 591 |
|  | percentage change between upper-band +1.0% and lower band -.1.0% | | | | | |
| 0 | 0.00 |  | 0.00 | 0.00 | 484 | 0.41 |
| 1 | 0.45 |  | 0.77 | 0.56 | 687 |
| 2 | 0.34 |  | 0.33 | 0.33 | 500 |
|  | percentage change between upper-band +1.5% and lower band -.1.5% | | | | | |
| 0 | 0.00 |  | 0.00 | 0.00 | 351 | 0.57 |
| 1 | 0.57 |  | 1.00 | 0.73 | 958 |
| 2 | 0.00 |  | 0.00 | 0.00 | 362 |

### **Random Forest Classifier**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Target Variable** | **precision** | recall | f1-score | support | accuracy score |
| percentage change between upper-band +0.7% and lower band -.07%  (6 days consecutive closing prices split week on week) | | | | | |
| **0** | **0.86** | **0.85** | **0.86** | **597** | **0.85** |
| **1** | **0.85** | **0.80** | **0.83** | **471** |
| **2** | **0.83** | **0.88** | **0.86** | **603** |
| percentage change between upper-band +0.7% and lower band -.07%  (10 days consecutive closing prices split week on week) | | | | | |
| **0** | **0.72** | **0.72** | **0.72** | **571** | **0.71** |
| **1** | **0.73** | **0.59** | **0.65** | **508** |
| **2** | **0.69** | **0.80** | **0.74** | **591** |
| percentage change between upper-band +0.7% and lower band -.07%  (14 days consecutive closing prices split week on week) | | | | | |
| **0** | **0.73** | **0.76** | **0.75** | **528** | **0.73** |
| **1** | **0.75** | **0.61** | **0.67** | **534** |
| **2** | **0.72** | **0.82** | **0.77** | **607** |
| percentage change between upper-band +1.0% and lower band -.1.0% | | | | | |
| 0 | 0.81 | 0.47 | 0.59 | 484 | 0.66 |
| 1 | 0.60 | 0.93 | 0.73 | 687 |
| 2 | 0.76 | 0.49 | 0.60 | 500 |
| percentage change between upper-band +1.5% and lower band -.1.5% | | | | | |
| 0 | 0.75 | 0.11 | 0.19 | 351 | 0.62 |
| 1 | 0.61 | 0.99 | 0.76 | 958 |
| 2 | 0.80 | 0.16 | 0.27 | 362 |

### **K Neighbours Classifier**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Target Variable** | **precision** | recall | f1-score | support | accuracy score |
| percentage change between upper-band +0.7% and lower band -.07% | | | | | |
| 0 | 0.36 | 0.45 | 0.40 | 384 | 0.35 |
| 1 | 0.32 | 0.36 | 0.34 | 336 |
| 2 | 0.39 | 0.26 | 0.31 | 394 |
| percentage change between upper-band +1.0% and lower band -.1.0% | | | | | |
| 0 | 0.32 | 0.32 | 0.32 | 316 | 0.40 |
| 1 | 0.45 | 0.59 | 0.51 | 467 |
| 2 | 0.38 | 0.20 | 0.26 | 331 |
| percentage change between upper-band +1.5% and lower band -.1.5% | | | | | |
| 0 | 0.28 | 0.19 | 0.23 | 227 | 0.53 |
| 1 | 0.58 | 0.83 | 0.69 | 627 |
| 2 | 0.38 | 0.09 | 0.15 | 260 |

### **XG Boost Classifier**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Target Variable** | **precision** | recall | f1-score | support | accuracy score |
| percentage change between upper-band +0.7% and lower band -.07% | | | | | |
| 0 | 0.37 | 0.33 | 0.35 | 384 | 0.37 |
| 1 | 0.36 | 0.31 | 0.33 | 336 |
| 2 | 0.38 | 0.47 | 0.42 | 394 |
| percentage change between upper-band +1.0% and lower band -.1.0% | | | | | |
| 0 | 0.36 | 0.13 | 0.19 | 316 | 0.43 |
| 1 | 0.44 | 0.83 | 0.58 | 467 |
| 2 | 0.40 | 0.15 | 0.22 | 331 |
| percentage change between upper-band +1.5% and lower band -.1.5% | | | | | |
| 0 | 0.43 | 0.03 | 0.05 | 227 | 0.57 |
| 1 | 0.57 | 0.97 | 0.72 | 627 |
| 2 | 0.49 | 0.07 | 0.12 | 260 |

## **Go Long Direction Prediction using Technical Indicators**

**(0-Non positive,1-Positive)**

The direction of the close price is estimated as percentage change of the close price between upper-band +0.5% and lower band -0.5%-if the percentage change of the closing price is more than 0.5%, the direction of the closing price is treated as positive and suitable for long Trading in stock market. Otherwise, the direction of the close price is treated as non-positive and not suitable for long Trading in stock market.

### **Logistic Regression Classifier**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Target Variable** | **precision** | recall | f1-score | support | accuracy score | Roc AUC score |
| Volume Indicators as Feature Variables | | | | | |  |
| **0** | **0.88** | **0.96** | **0.92** | **685** | **0.90** | **0.88** |
| **1** | **0.92** | **0.80** | **0.85** | **425** |
| Momentum Indicators as Feature Variables | | | | | |  |
| **0** | **0.76** | **0.82** | **0.79** | **669** | **0.74** | **0.72** |
| **1** | **0.69** | **0.62** | **0.65** | **439** |
| Trend Indicators as Feature Variables | | | | | |  |
| 0 | 0.72 | 0.91 | 0.80 | 659 | 0.74 | 0.70 |
| 1 | 0.78 | 0.49 | 0.61 | 451 |
| volatility Indicators as Feature Variables | | | | | |  |
| 0 | 0.69 | 0.95 | 0.80 | 685 | 0.70 | 0.63 |
| 1 | 0.81 | 0.30 | 0.44 | 425 |

### **Decision Tree Classifier**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Target Variable** | **precision** | recall | f1-score | support | accuracy score | Roc AUC score |
| Volume Indicators as Feature Variables | | | | | |  |
| 0 | 0.83 | 0.85 | 0.84 | 685 | 0.80 | 0.79 |
| 1 | 0.75 | 0.72 | 0.74 | 425 |
| Momentum Indicators as Feature Variables | | | | | |  |
| 0 | 0.74 | 0.80 | 0.77 | 669 | 0.71 | 0.69 |
| 1 | 0.65 | 0.57 | 0.61 | 439 |
| Trend Indicators as Feature Variables | | | | | |  |
| 0 | 0.68 | 0.81 | 0.74 | 659 | 0.66 | 0.62 |
| 1 | 0.61 | 0.44 | 0.51 | 451 |
| volatility Indicators as Feature Variables | | | | | |  |
| 0 | 0.73 | 0.72 | 0.72 | 685 | 0.66 | 0.64 |
| 1 | 0.55 | 0.56 | 0.56 | 425 |

### **Random Forest Classifier**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Target Variable** | **precision** | recall | f1-score | support | accuracy score | Roc AUC score |
| Volume Indicators as Feature Variables | | | | | |  |
| 0 | 0.85 | 0.95 | 0.90 | 685 | 0.86 | 0.84 |
| 1 | 0.90 | 0.73 | 0.80 | 425 |
| Momentum Indicators as Feature Variables | | | | | |  |
| 0 | 0.75 | 0.86 | 0.80 | 669 | 0.74 | 0.71 |
| 1 | 0.72 | 0.55 | 0.63 | 439 |
| Trend Indicators as Feature Variables | | | | | |  |
| 0 | 0.76 | 0.92 | 0.83 | 659 | 0.78 | 0.74 |
| 1 | 0.83 | 0.57 | 0.67 | 451 |
| volatility Indicators as Feature Variables | | | | | |  |
| 0 | 0.79 | 0.93 | 0.85 | 685 | 0.80 | 0.77 |
| 1 | 0.83 | 0.61 | 0.70 | 425 |

### **K Neighbours Classifier**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Target Variable** | **precision** | recall | f1-score | support | accuracy score | Roc AUC score |
| Volume Indicators as Feature Variables | | | | | |  |
| 0 | 0.64 | 0.85 | 0.73 | 685 | 0.61 | 0.84 |
| 1 | 0.48 | 0.23 | 0.31 | 425 |
| Momentum Indicators as Feature Variables | | | | | |  |
| 0 | 0.68 | 0.87 | 0.77 | 669 | 0.68 | 0.71 |
| 1 | 0.66 | 0.38 | 0.48 | 439 |
| Trend Indicators as Feature Variables | | | | | |  |
| 0 | 0.60 | 0.86 | 0.71 | 659 | 0.58 | 0.74 |
| 1 | 0.45 | 0.17 | 0.25 | 451 |
| volatility Indicators as Feature Variables | | | | | |  |
| 0 | 0.62 | 0.81 | 0.71 | 685 | 0.58 | 0.77 |
| 1 | 0.41 | 0.20 | 0.27 | 425 |

### **XG Boost Classifier**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Target Variable** | **precision** | recall | f1-score | support | accuracy score | Roc AUC score |
| Volume Indicators as Feature Variables | | | | | |  |
| 0 | 0.89 | 0.93 | 0.91 | 685 | 0.89 | 0.84 |
| 1 | 0.88 | 0.82 | 0.85 | 425 |
| Momentum Indicators as Feature Variables | | | | | |  |
| 0 | 0.76 | 0.83 | 0.79 | 669 | 0.74 | 0.71 |
| 1 | 0.70 | 0.59 | 0.64 | 439 |
| Trend Indicators as Feature Variables | | | | | |  |
| **0** | **0.80** | **0.91** | **0.85** | **659** | **0.81** | **0.74** |
| **1** | **0.83** | **0.67** | **0.74** | **451** |
| volatility Indicators as Feature Variables | | | | | |  |
| **0** | **0.81** | **0.90** | **0.85** | **685** | **0.81** | **0.77** |
| **1** | **0.80** | **0.67** | **0.73** | **425** |

## **Go Short Direction Prediction using Technical Indicators**

**(0-Negative,1-non-Negative)**

The direction of the close price is estimated as percentage change of the close price between upper-band +0.5% and lower band -0.5%-if the percentage change of the closing price is less than -0.5%, the direction of the closing price is treated as Negative and suitable for Short Trading in stock market. Otherwise, the direction of the close price is treated as non-negative and not suitable for Short Trading in stock market.

### **Logistic Regression Classifier**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Target Variable** | **precision** | recall | f1-score | support | accuracy score | Roc AUC score |
| Volume Indicators as Feature Variables | | | | | |  |
| **0** | **0.94** | **0.79** | **0.86** | **413** | **0.90** | **0.88** |
| **1** | **0.89** | **0.97** | **0.93** | **697** |
| Momentum Indicators as Feature Variables | | | | | |  |
| **0** | **0.68** | **0.61** | **0.64** | **417** | **0.75** | **0.72** |
| **1** | **0.78** | **0.83** | **0.80** | **691** |
| Trend Indicators as Feature Variables | | | | | |  |
| 0 | 0.79 | 0.48 | 0.60 | 416 | 0.76 | 0.70 |
| 1 | 0.75 | 0.93 | 0.83 | 694 |
| volatility Indicators as Feature Variables | | | | | |  |
| 0 | 0.81 | 0.27 | 0.41 | 413 | 0.71 | 0.62 |
| 1 | 0.69 | 0.96 | 0.80 | 697 |

### **Decision Tree Classifier**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Target Variable** | **precision** | recall | f1-score | support | accuracy score | Roc AUC score |
| Volume Indicators as Feature Variables | | | | | |  |
| 0 | 0.76 | 0.63 | 0.69 | 413 | 0.79 | 0.76 |
| 1 | 0.80 | 0.88 | 0.84 | 697 |
| Momentum Indicators as Feature Variables | | | | | |  |
| 0 | 0.62 | 0.56 | 0.59 | 417 | 0.70 | 0.68 |
| 1 | 0.75 | 0.79 | 0.77 | 691 |
| Trend Indicators as Feature Variables | | | | | |  |
| 0 | 0.57 | 0.43 | 0.49 | 416 | 0.67 | 0.62 |
| 1 | 0.70 | 0.81 | 0.75 | 694 |
| volatility Indicators as Feature Variables | | | | | |  |
| 0 | 0.60 | 0.53 | 0.57 | 413 | 0.70 | 0.66 |
| 1 | 0.74 | 0.79 | 0.77 | 697 |

### **Random Forest Classifier**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Target Variable** | **precision** | recall | f1-score | support | accuracy score | Roc AUC score |
| Volume Indicators as Feature Variables | | | | | |  |
| 0 | 0.89 | 0.71 | 0.79 | 413 | 0.86 | 0.83 |
| 1 | 0.85 | 0.95 | 0.90 | 697 |
| Momentum Indicators as Feature Variables | | | | | |  |
| 0 | 0.71 | 0.52 | 0.60 | 417 | 0.74 | 0.69 |
| 1 | 0.75 | 0.87 | 0.81 | 691 |
| Trend Indicators as Feature Variables | | | | | |  |
| 0 | 0.89 | 0.48 | 0.63 | 416 | 0.78 | 0.72 |
| 1 | 0.76 | 0.97 | 0.85 | 694 |
| volatility Indicators as Feature Variables | | | | | |  |
| 0 | 0.87 | 0.53 | 0.66 | 413 | 0.79 | 0.74 |
| 1 | 0.77 | 0.95 | 0.85 | 697 |

### **K Neighbours Classifier**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Target Variable** | **precision** | recall | f1-score | support | accuracy score | Roc AUC score |
| Volume Indicators as Feature Variables | | | | | |  |
| 0 | 0.48 | 0.47 | 0.48 | 413 | 0.61 | 0.83 |
| 1 | 0.69 | 0.69 | 0.69 | 697 |
| Momentum Indicators as Feature Variables | | | | | |  |
| 0 | 0.58 | 0.58 | 0.58 | 417 | 0.68 | 0.69 |
| 1 | 0.74 | 0.75 | 0.74 | 691 |
| Trend Indicators as Feature Variables | | | | | |  |
| 0 | 0.45 | 0.45 | 0.45 | 416 | 0.59 | 0.72 |
| 1 | 0.67 | 0.67 | 0.67 | 694 |
| volatility Indicators as Feature Variables | | | | | |  |
| 0 | 0.43 | 0.41 | 0.42 | 413 | 0.58 | 0.74 |
| 1 | 0.66 | 0.68 | 0.67 | 697 |

### **XG Boost Classifier**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Target Variable** | **precision** | recall | f1-score | support | accuracy score | Roc AUC score |
| Volume Indicators as Feature Variables | | | | | |  |
| 0 | 0.86 | 0.80 | 0.83 | 413 | 0.88 | 0.82 |
| 1 | 0.89 | 0.92 | 0.90 | 697 |
| Momentum Indicators as Feature Variables | | | | | |  |
| 0 | 0.70 | 0.61 | 0.65 | 417 | 0.75 | 0.69 |
| 1 | 0.78 | 0.84 | 0.81 | 691 |
| Trend Indicators as Feature Variables | | | | | |  |
| **0** | **0.84** | **0.60** | **0.70** | **416** | **0.81** | **0.72** |
| **1** | **0.80** | **0.93** | **0.86** | **694** |
| volatility Indicators as Feature Variables | | | | | |  |
| **0** | **0.79** | **0.61** | **0.69** | **413** | **0.80** | **0.74** |
| **1** | **0.80** | **0.91** | **0.85** | **697** |

# Analysis and Results

## Analysis for HDFC Stock

## Direction Detection by 6,10,14 days consecutive closing prices split week on week:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Target Variable** | **precision** | **recall** | **f1-score** | **support** | **accuracy score** |
| percentage change between upper-band +0.7% and lower band -.07% (6 days consecutive closing prices split week on week) | | | | | |
| **0** | **0.91** | **0.81** | **0.86** | **544** | **0.87** |
| **1** | **0.85** | **0.90** | **0.88** | **580** |
| **2** | **0.85** | **0.89** | **0.87** | **547** |
| percentage change between upper-band +0.7% and lower band -.07%(10 days consecutive closing prices split week on week) | | | | | |
| **0** | **0.87** | **0.86** | **0.87** | **559** | **0.87** |
| **1** | **0.87** | **0.87** | **0.87** | **550** |
| **2** | **0.87** | **0.88** | **0.87** | **561** |
| percentage change between upper-band +0.7% and lower band -.07%(14 days consecutive closing prices split week on week) | | | | | |
| **0** | **0.80** | **0.77** | **0.79** | **536** | **0.80** |
| **1** | **0.79** | **0.81** | **0.80** | **543** |
| **2** | **0.80** | **0.81** | **0.80** | **590** |

From Table, it can be observed that random forest modelling done for percentage change in close price between upper-band +0.7% and lower band -.0.7% has given the highest efficiency in prediction among all Modelling techniques namely logistic regression, decision tree, random forest, k nearest neighbours and XG Boost Modelling. It predicts upward, neutral and downward trend direction with reasonably good precision. F1-score combining the precision and recall of a classifier into a single metric is also reasonably good. This has been tested and proven with 6,10- and 14-days consecutive closing prices split week on week as 6,10 and 14 feature variables. Hence, Random Forest Modelling provides a reasonably good modelling technique to be able to provide optimal prediction performance.

## **Go Long Direction Prediction using Technical Indicators**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Target Variable** | **precision** | **recall** | **f1-score** | **support** | **accuracy score** | **Roc AUC score** |
| Volume Indicators as Feature Variables for Logistic Regression Classifier | | | | | |  |
| **0** | **0.90** | **0.99** | **0.94** | **658** | **0.92** | **0.91** |
| **1** | **0.98** | **0.83** | **0.90** | **452** |
| Momentum Indicators as Feature Variables for Logistic Regression Classifier | | | | | |  |
| **0** | **0.79** | **0.84** | **0.81** | **685** | **0.76** | **0.74** |
| **1** | **0.71** | **0.63** | **0.67** | **423** |
| Trend Indicators as Feature Variables for XG Boost Classifier | | | | | |  |
| **0** | **0.81** | **0.92** | **0.86** | **679** | **0.82** | **0.75** |
| **1** | **0.85** | **0.65** | **0.74** | **431** |
| volatility Indicators as Feature Variables for XG Boost Classifier | | | | | |  |
| **0** | **0.81** | **0.91** | **0.86** | **658** | **0.82** | **0.75** |
| **1** | **0.84** | **0.69** | **0.76** | **452** |

From Table, it can be observed that logistic regression modelling done for percentage change in close price between upper-band +0.5% and lower band -.0.5% has given highest precision, recall, f1-score and accuracy score for volume and momentum indicators whereas XG Boost Classifier provided best prediction performance for trend and volatility indicators.

## **Go Short Direction Prediction using Technical Indicators**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Target Variable** | **precision** | **recall** | **f1-score** | **support** | **accuracy score** | **Roc AUC score** |
| Volume Indicators as Feature Variables for Logistic Regression Classifier | | | | | |  |
| **0** | **0.97** | **0.83** | **0.90** | **399** | **0.93** | **0.91** |
| **1** | **0.91** | **0.99** | **0.95** | **711** |
| Momentum Indicators as Feature Variables for XG Boost Classifier | | | | | |  |
| **0** | **0.72** | **0.59** | **0.64** | **394** | **0.77** | **0.70** |
| **1** | **0.79** | **0.87** | **0.83** | **714** |
| Trend Indicators as Feature Variables for Logistic Regression Classifier | | | | | |  |
| **0** | **0.91** | **0.56** | **0.69** | **414** | **0.81** | **0.76** |
| **1** | **0.79** | **0.97** | **0.87** | **696** |
| volatility Indicators as Feature Variables for Random Forest Classifier | | | | | |  |
| **0** | **0.88** | **0.55** | **0.68** | **399** | **0.81** | **0.76** |
| **1** | **0.79** | **0.96** | **0.87** | **711** |

From Table, it can be observed that logistic regression modelling done for percentage change in close price between upper-band +0.5% and lower band -.0.5% has given highest precision, recall, f1-score and accuracy score for volume and trend indicators whereas XG Boost Classifier provided best prediction performance for momentum indicators. Similarly Random Forest Classifier provided best predictions for volatility indicators.

## Analysis for KOTAK Stock

Similar analysis can be done for KOTAK stock.

## Analysis for SBI Stock

Similar analysis can be done for SBI stock.