Machine Learning Approach in Stock Market Prediction: A Case of Colombo Stock Exchange

Gayan Wickramarathna*, Hashan Ratnayake**

**Student, Rajarata Unversity of Sri Lanka, Mihintale, Sri Lanka*
***Leturer, Rajarata Unversity of Sri Lanka, Mihintale, Sri Lanka*

DOI: 10.29322/IJSRP.12.05.2022.p12519
http://dx.doi.org/10.29322/IJSRP.12.05.2022.p12519

Paper Received Date: 14th April 2022
Paper Acceptance Date: 29th April 2022
Paper Publication Date: 6th May 2022

Abstract- The profitability of the stock market investment depends on the investor's decisions and is based on a mix of dynamic environmental factors. Stock price trends are repeatedly forecasted to extract useful patterns and predict their movements. There are different approaches to stock price prediction and different forecasting methods are used by stock market analysts. As a replica of many scientific endeavors, several methods have been found to accurately predict stock prices. Most researchers have been used technical analysis to get more accurate results, while limited researchers used fundamental analysis. However, there is no evidence that compares these two analyzes. Ceylon Tobacco Company PLC, Dialog Axiata PLC, and John Keells Holdings PLC were selected as the sample for this study by employing a simple random sampling method and market capitalization. The technical analysis was performed using five different classifiers and the results were evaluated using the mean absolute error (MAE), root mean square error (RMSE), relative absolute error (RAE), and root-relative square error (RRSE). Sequential Minimum Optimization regression (SMOreg) has yielded more accurate results. The fundamental analysis was conducted by employing natural language processing mechanisms and the Random Trees classifier presented the best results. The outcome of this study shows that a comprehensive model can be built with a combination of technical and fundamental analysis. The findings facilitate to predict the share price of the Colombo Stock Exchange using machine learning techniques. This model was able to predict the stock price with 65% accuracy and would benefit all individual investors in the local stock market.

Index Terms*- Colombo Stock Exchange, Fundamental Analysis, Machine Learning, Random Trees, Sequential Minimum Optimization, Stock Price Prediction, Technical Analysis.*

I. INTRODUCTION

The Colombo Stock Exchange (CSE) is the leading stock exchange in Sri Lanka that offers investors buy and sell shares. It is one of the major exchanges providing an e-commerce platform in South Asia [1]. The stock market is the bone of fast-emerging economies like Sri Lanka. Therefore, the growth of our country is strongly associated with the performance of the Colombo Stock Exchange. Most of the investors and brokers nowadays use intelligent trading systems that help them to predict stock prices based on various situations and conditions, thereby helping them in making direct investment decisions. Many researchers believe that fundamental analysis is only good in the long term. However, it is not suitable for the medium to short term. Some other researchers used technical analysis, and they suppose history repeats itself. Similarly, there is a paucity of research on stock price forecasting using machine learning in the context of Sri Lanka. Therefore, it is important to research to measure the variance between the results of fundamental and technical analysis of stock price forecasting using machine learning in the context of Sri Lanka.

II. LITERATURE REVIEW

Predicting stock price volatility is a difficult task, as price movements are randomly walked and change over time. Since the last decade, stockbrokers and prospects have relied on a variety of intelligent systems to make trading decisions. Machine learning methods for stock price predictions are becoming popular. Through various experiments, it is possible to test machine-learning techniques and select the most suitable one for predicting stock prices[2].

One area of limited success in stock market forecasting is the use of text data and news articles for price forecasts. Chan, Chui, & Kwok [3], confirm the response to news articles. They have shown that economic news always has a positive or negative impact on the number of shares traded. In their research, Nagar & Hahsler[4] proposed an automated text-based approach to gathering news from various sources and forming a news corporation. Corpus is filtered into relevant sentences and analyzed using natural language processing (NLP) techniques.

ISSN 2250-3153
This research paper [5] studies how the results of financial forecasting can be improved when multiple levels of relevant news articles are used simultaneously for the target stock. They used multi-kernel learning methods to segment information extracted from five different news categories based on sectors, sub-sectors, industries, and so on.

On the other hand, Technical Analysis is used for price forecasting using historical stock prices such as open price, high price, low price, close price and volume. Many researchers have had successful results. In addition, many researchers have used this method rather than the Fundamental Analysis. However, this analysis is not suitable for long-term prediction.

The TA ultimately relies on human interpretation and, due to its subjective nature, technicians can often predict the same data and explain it in BE theory. While TA is generally a controversial target for scientific statements, some studies support it [6], [7], [8], others point out that less predictive problems are Ability to do Power [9] and [10].

Recently, data mining techniques and artificial intelligence (AI) systems such as decision trees and artificial neural networks (ANN) have been applied to this area [11]. Stock market data mining has been used to forecast trends and prices to achieve maximum profits [12]. Data mining techniques have been successfully shown to generate high predictive accuracy of the stock price movement [13].

III. GENERAL FRAMEWORK

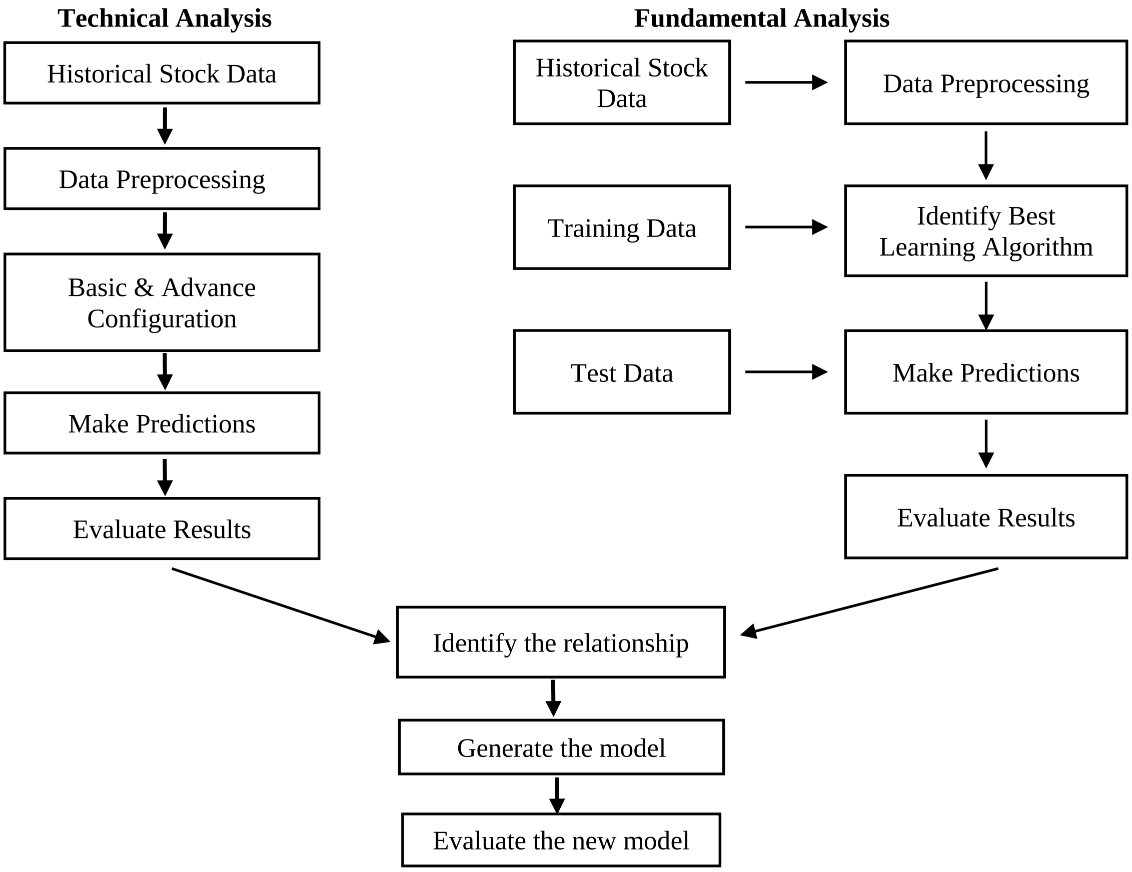

Fig1 General Framework

A. Technical analysis

Step 01: Historical Stock Data

Data collection is a very basic module and is the starting point of a project. It usually deals with gathering the right data set. Historical data on Open, Close, High, Low and Volume share prices have been collected by the researcher at the Colombo Stock Exchange. Each attribute has its own value and time. All values are numeric values.

Step 02: Data Preprocessing

Data processing can measure data quality in terms of accuracy, completeness, consistency, timeliness, reliability, and interpretation. Data preprocessing consists of data cleaning, data integration, data transformation, and data reduction.

Step 03: Basic & Advance Configuration

Researcher used the WEKA "Forecast" tab to predict closing price. The version 3.7.3 and above WEKA has a dedicated time series environment.

This environment can be used for developing, visualizing and evaluating forecasting models.

Step 04: Make Predictions

After basic and advanced configuration, the close price can be predicted. Other properties, such as open, high, or low price, and volume can be predicted if needed. Figure 3.8 shows the sample prediction.

Step 5: Evaluate Results

The accuracy of the model can be estimated by comparing the actual results with the predicted results. Similarly, the Mean Absolute Error (MAE), the Root Mean Square Error (RMSE), the Relative Absolute Error (RAE), and the Root-Relative Square Error (RRSE) can be evaluated and select the most relevant classification.

B. Fundamental analysis

Step 01: Historical Stock Data

Data in fundamental analysis both qualitative and quantitative data such as financial statements, newspaper articles, and announcements. In this research, only announcements published by the cse.lk website was considered. The total data for the past five years is collected.

Step 02: Data Preprocessing

Most of the data collected in cse.lk are scanned files and some are text files. Text files (.txt) can be directly converted to ARFF files. However, the scanned files (.pdf) cannot be converted to ARFF. Scanned files must first be converted to txt files and converted to ARFF files.

Step 03: Learning Algorithm (Learn Rules)

This is about the first step, called training data. It is assumed that each tuple/sample belongs to a predefined class as determined by the class label attribute. In the fundamental analysis, the class attribute was the "Negative", "Positive" or "Not Change". The model is represented as a classification rule, decision trees, or mathematical formulas. For the training data, the researcher used a classify tab of WEKA.

Step 04: Make Predictions

There are two steps to the classification process. The first is to build a classification model using training data. Second, the model generated in the previous step is tested using the test dataset. This step discusses how to predict values using the model. It assesses the accuracy of the model. The known label of the test sample is compared with the classification result of the model.

Step 5: Evaluate Results

The accuracy of the model can be estimated by comparing the actual results with the predicted results. Similarly, the Accuracy, Precision, Recall. F-Measure and ROC Area can be evaluated and select the most relevant classification.

IV. DATA ANALYSIS & PRESENTATION

A. Technical analysis

In processing the data, the researcher considered JKH, CTC and, DIAL daily data for five years. Minimum price, maximum price, mean, standard deviation, distinct and unique percentage were found. Table 1 and Figure 3 show Statistical values of CTC.

Analysis of results

The table 2 shows the comparison between the predicted results using linear regression and the actual results. The difference is expressed as a percentage and is the minimum. To compare the accuracy of the results predicted by the researcher, the actual price considered for 10 days in January 2020. The predicted values shown in four decimal places. However, the stock market considers only at one decimal place. For an example, 2.10, 2.20, 2.30 ..., etc. Therefore, the predicted values rounded to one decimal place.

The table 3 shows that the SMO regression classifier enables the prediction of share prices more accurately than other classifiers.

Table 2 Statistical Value of CTC

Attributes	Min	Max	Mean	Std Dev	Distinct	Unique (100%)
Open	800.2	1496.7	1064.204	168.745	434	24
High	808.0	1500.0	1069.663	168.859	395	21
Low	797.0	1500.0	1057.069	167.274	388	20
Close	800.2	1496.7	1063.622	168.113	481	27

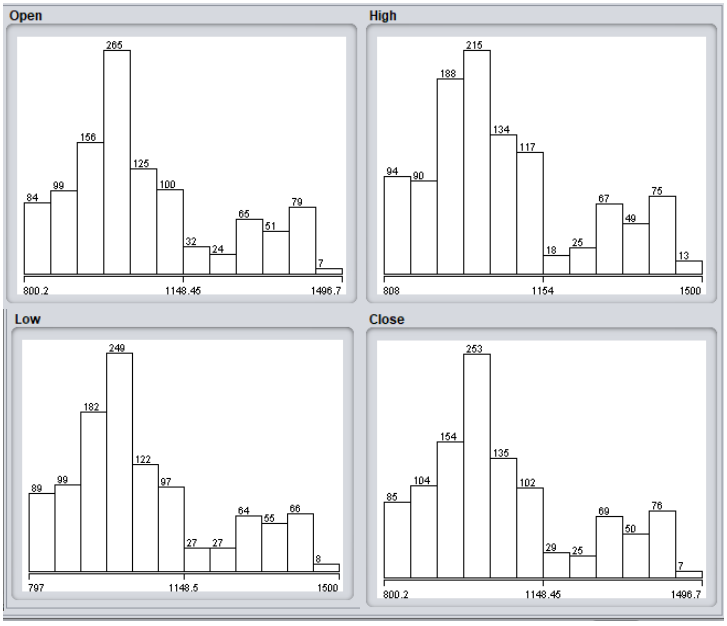

Fig 2 Statistical Value of CTC

Table 2 Comparison Forecast Values and Actual Values

Linear Regression – CTC												
Day	**Forecast**				**Actual**				**Difference in Percentage**			
	Open	**High**	**Low**	**Close**	**Open**	**High**	**Low**	**Close**	**Open**	**High**	**Low**	**Close**
1	1110.60	1110.60	1099.20	1096.10	1110.00	1120.00	1100.00	1100.00	0.05	-0.84	-0.07	-0.35
2	1099.90	1106.40	1082.10	1096.80	1110.00	1120.00	1095.00	1095.00	-0.91	-1.21	-1.18	0.16
3	1099.60	1110.40	1097.20	1097.90	1095.00	1095.00	1093.00	1095.00	0.42	1.41	0.38	0.26
4	1105.00	1111.10	1093.60	1091.80	1100.00	1100.00	1095.00	1099.80	0.45	1.01	-0.13	-0.73
5	1103.00	1107.30	1081.00	1085.70	1096.00	1096.00	1091.00	1095.00	0.64	1.03	-0.92	-0.85
6	1096.70	1100.60	1089.70	1088.30	1090.10	1100.00	1085.00	1100.00	0.61	0.05	0.43	-1.06
7	1100.20	1109.00	1089.00	1088.70	1080.00	1139.90	1080.00	1100.00	1.87	-2.71	0.83	-1.03

8	1094.50	1100.70	1083.00	1082.00	1090.00	1090.00	1070.20	1079.90	0.41	0.98	1.20	0.19
9	1089.20	1097.70	1078.20	1079.50	1070.20	1090.00	1070.00	1070.10	1.78	0.71	0.77	0.88
10	1089.10	1096.30	1075.70	1076.50	1089.90	1089.90	1060.00	1076.80	-0.07	0.59	1.48	-0.03

Table 3 Classifier Comparison – CTC

Measurement	Gaussian Processes	Linear Regression	Multilayer Perceptron	SMOreg	M5P
MAE	7.3477	4.4815	5.3450	4.3357	4.4399
RMSE	125.8979	99.3761	100.3510	99.6994	96.8065
RAE	176.1047	107.2517	127.9906	103.7592	106.2517
RRSE	11.3965	9.0023	9.0883	9.0314	8.7702

B. Fundamental analysis

The training dataset looks at the past five years and the test dataset looks at this year's announcements. Table 4 shows how to split the training and test dataset.

Table 4 Text Data Set

Company	Train	Test
CTC	43	3
DIAL	71	7
JHK	90	7

In the case of CTC, the total number of instances is 43, it contains 17 positive instances, 12 negative instances, and 14 not change instances. The 14 announcements do not affect the stock price. However, 29 announcements affect the company's share price. That is 67% of the total instances. In terms of DIAL, the total number of cases is 71, which includes 33 positives, 18 negatives, and 20 not change cases. The 20 announcements do not affect the stock price. However, 51 announcements affect the company's share price. That is 72% of the total chances. In the case of JKH, there are 90 positive cases, 42 positive, 44 negatives, and four not change. Announcements 4 do not affect the stock price. However, 86 announcements affect the company's share price. That is 95.6% of the total chances. This shows that investors should be very careful about announcements when buying and selling JKH shares.

Analysis of results

Table 5 Classifier Evaluation - CTC

Measurement	Naïve Bayes	SMO	J48	Random Forest	Random Tree
Accuracy (%)	90.7	100	83.72	100	100
Precision	0.919	1.000	0.835	1.000	1.000
Recall	0.907	1.000	0.837	1.000	1.000
ROC	0.975	1.000	0.939	1.000	1.000
F – Measure	0.902	1.000	0.834	1.000	1.000
AUC	0.975	1.000	0.940	1.000	1.000
MAE	0.0628	0.2222	0.1597	0.1524	0
RMSE	0.2464	0.2722	0.2826	0.1803	0
RAE (%)	14.27	50.48	36.28	34.62	0
RRSE (%)	52.55	58.03	60.25	38.45	0

When the look at the table 5 SMO, Random Forest, and Random Tree classifiers show 100% accuracy and 1.0 Precision, Recall, ROC, AUC, and F-measure. The random tree shows the best results. Its accuracy is 100% and the error is zero. Therefore, this is

ISSN 2250-3153
the best classification for stock price prediction using text analysis. The tables below (Table 6, Table 7, and Table 8) show a comparison between the predicted results and the actual results. All predictions conducted using "Random Tree".

Table 6 Results Comparison - CTC

Instance	Predicted Results	Actual Results
1	Positive	Not Change
2	Positive	Positive
3	Positive	Negative

Table 7 Results Comparison - DIAL

Instance	Predicted Results	Actual Results
1	Positive	Not Change
2	Positive	Negative
3	Positive	Positive
4	Positive	Positive
5	Positive	Positive
6	Positive	Not Change
7	Not Change	Not Change

Table 8 Results Comparison - JKH

Instance	Predicted Results	Actual Results
1	Negative	Positive
2	Negative	Positive
3	Negative	Negative
4	Negative	Negative
5	Negative	Negative
6	Positive	Negative
7	Negative	Positive

C. The relation between fundamental and technical analysis.

Most researchers have performed fundamental and technical analysis separately, but they have not analyzed the correlation between them. This study fills that gap in the literature. This relationship can be understood by studying the following tables (Table 9, Table 10, and Table 11).

Table 9 Evaluation Results - CTC

Date	Technical	Fundamental	Actual
27 Feb 2020	1119.30	Positive	1123
28 Feb 2020	1121.90	Positive	1110
14 Mar 2020	998.70	Positive	1000

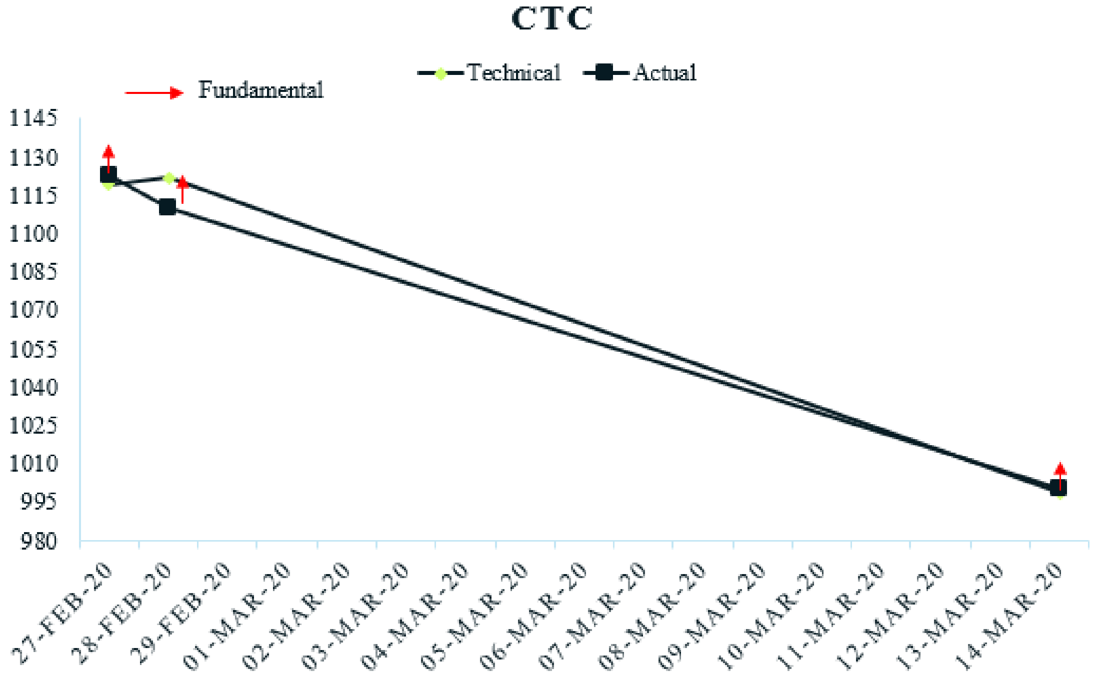

Fig 3 Evaluation Results – CTC

Table 10 Evaluation Results - DIAL

Date	Technical	Fundamental	Actual
16 Jan 2020	12.50	Positive	12.50
17 Feb 2020	12.60	Positive	12.70
20 Feb 2020	12.60	Positive	12.60
02 Mar 2020	11.80	Positive	11.80
02 Mar 2020	11.80	Not Change	11.80
11 Mar 2020	10.40	Positive	10.40
12 Mar 2020	10.40	Positive	10.20

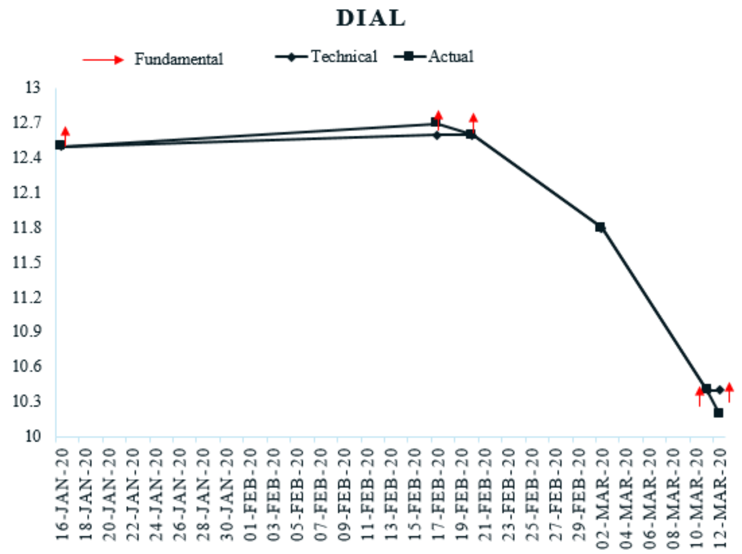

Fig 4 Evaluation Results - DAIL

Table 11 Evaluation Results - JKH

Date	Technical	Fundamental	Actual
09 Jan 2020	160.80	Negative	162.80
13 Jan 2020	163.20	Negative	163.50
16 Jan 2020	166.70	Negative	164.00
17 Jan 2020	163.90	Negative	163.10
17 Jan 2020	163.90	Positive	163.10
30 Jan 2020	160.70	Negative	162.10
11 Feb 2020	160.00	Negative	157.60

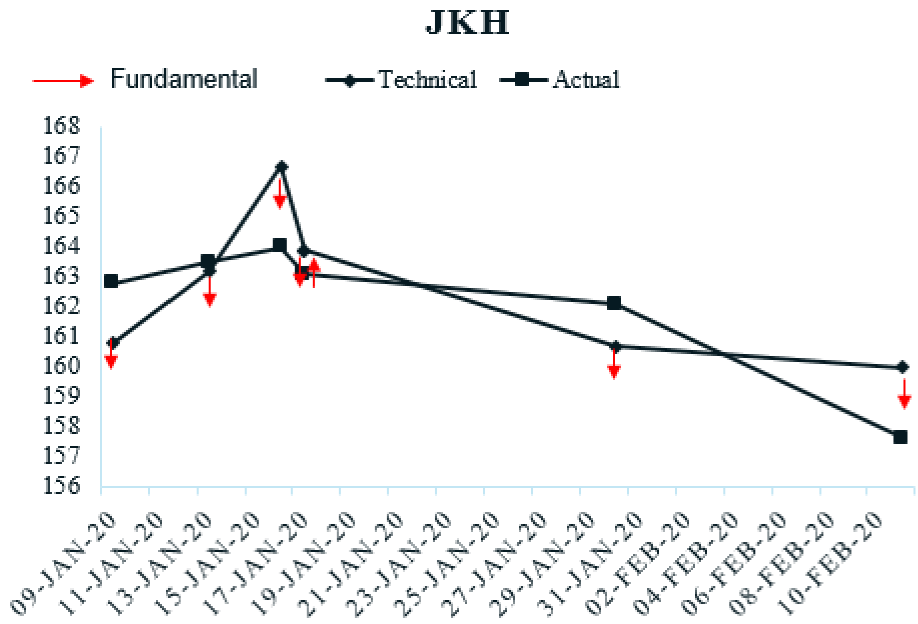

Fig 5 Evaluation Results - JKH

There are 17 instances in the above tables (Table 9, Table 10, and Table 11). There are eight instances, similar results compared to the actual results of the fundamental analysis. That is 47 percent. In technical analysis, five instances show the same values with actual values. That is 29 percent. In this case, the fundamental analysis results in better results than the technical analysis. However, there is a problem with the fundamental analysis; it does not predict numerical value.

The combination of the two analyzes appears to yield better results than the results of the fundamental and technical analyses performed separately. Accordingly, the researcher can make the following formula:

$$CP = VTA \pm RFA \quad (1)$$

Where,

CP – Closing Price

VTA – Value of Technical Analysis

RFA – Results of Fundamental Analysis

In this model, 11 cases show the correct results. It is 65 percent. However, there is an assumption that positive news will keep the close price running smoothly.

Technical analysis generates numeric value. However, predict news has a negative or positive effect. If there is a negative effect, it reduces the TA value. In addition, if it is positive it increases TA value.

V. CONCLUSION & RECOMMENDATIONS

In technical analysis, the SMO regression function provides the ability to predict CSE's share price more accurately than other functions such as Gaussian processes, linear regression, and multi-layer perception. In the fundamental analysis, its accuracy is more than 80% of all models. Similarly, SMO, Random Forest, and Random Tree classifiers show 100% accuracy and 1.0 Precision, Recall, ROC, and F-measure. The random tree shows the best results. Its accuracy is 100% and the error is zero. Therefore, this is the best classification for stock price prediction using text analysis.

Financial news affects the closing price of CTC 67%, Dial 72%, and JKH 95.6 percentage. It can be a positive or a negative impact. If it is, a positive impact closing price will be increased. Otherwise, is negative impact closing price will be decreased. Thus, the researcher was able to achieve all the objectives that he had hoped to achieve. Finally, the researcher suggests that technical and fundamental analysis should be combined, rather than separately. It gets more accuracy than other models.

In the fundamental analysis, the researcher considers only the announcements posted on the CSE website. Newspaper articles, Twitter, FB posts, and financial reporting can also be considered in the fundamental analysis. The most important thing for further research is to identify the value of fundamental analysis. That is to say how much the value of the shares will increase or decrease by the fundamental analysis.

REFERENCES

[1] "Colombo Stock Exchange - Wikipedia." https://en.wikipedia.org/wiki/Colombo_Stock_Exchange (accessed Jul. 11, 2020).

[2] V. H. Shah, "Machine Learning Techniques for Stock Prediction," pp. 1–19, 2007.

[3] Y. C. Chan, A. C. W. Chui, and C. C. Y. Kwok, "The impact of salient political and economic news on the trading activity," *Pacific Basin Financ. J.*, vol. 9, no. 3, pp. 195–217, Jun. 2001, doi: 10.1016/S0927-538X(01)00015-4.

[4] A. Nagar and M. Hahsler, "Using Text and Data Mining Techniques to extract Stock Market Sentiment from Live News Streams," *Int. Proc. Comput. …*, vol. 47, no. Iccts, pp. 91–95, 2012, doi: 10.7763/IPCSIT.2012.V47.18.

[5] Y. Shynkevich, T. M. McGinnity, S. Coleman, and A. Belatreche, "Predicting stock price movements based on different categories of news articles," *Proc. - 2015 IEEE Symp. Ser. Comput. Intell. SSCI 2015*, no. December, pp. 703–710, 2015, doi: 10.1109/SSCI.2015.107.

[6] D. R. Aronson, *Evidence-Based Technical Analysis: Applying the Scientific Method and Statistical Inference to Trading Signals*. John Wiley and Sons, 2011.

[7] University of Cambridge, "Finance Technical Analysis," *TA - B.*, pp. 103–105, 2011, [Online]. Available: http://www.mrao.cam.ac.uk/~mph/Technical_Analysis.pdf.

[8] C. H. Park and S. H. Irwin, "What do we know about the profitability of technical analysis?," *Journal of Economic Surveys*, vol. 21, no. 4. pp. 786–826, Sep. 2007, doi: 10.1111/j.1467-6419.2007.00519.x.

[9] G. A. W. Griffioen, "Technical Analysis in Financial Markets," *SSRN Electron. J.*, p. 322, 2003, doi: 10.2139/ssrn.566882.

[10] H. Yu, G. V. Nartea, C. Gan, and L. J. Yao, "Predictive ability and profitability of simple technical trading rules: Recent evidence from Southeast Asian stock markets," *Int. Rev. Econ. Financ.*, vol. 25, pp. 356–371, Jan. 2013, doi: 10.1016/j.iref.2012.07.016.

[11] Q. A. Al-Radaideh, A. A. Assaf, and E. Alnagi, "PREDICTING STOCK PRICES USING DATA MINING TECHNIQUES," *Int. Arab Conf. Inf. Technol.*, 2013.

[12] A. Chelawat, R. Agarwal, and A. Das, "Application of Data Mining in Equity Market," *Int. J. Eng. Tech.*, vol. 3, no. 3, pp. 42–49, 2017, [Online]. Available: http://www.ijetjournal.org.

[13] P. Ou and H. Wang, "Prediction of Stock Market Index Movement by Ten Data Mining Techniques," *Mod. Appl. Sci.*, vol. 3, no. 12, Nov. 2009, doi: 10.5539/mas.v3n12p28.

AUTHORS

First Author – Gayan Wickramarathna, Undergraduate Student, Rajarata University of Sri Lanka. lhmgiw@gmail.com

Second Author – Hashan Ratnayake, Leturer, Rajarata Unversity of Sri Lanka, hashanpdn@gmail.com

Correspondence Author – Gayan Wickramarathna, lhmgiw@gmail.com , l.h.m.gayan@gmail.com, +94712023174