PERFORMANCE ANALYSIS and PREDICTION of NEPAL STOCK MARKET (NEPSE) for INVESTMENT DECISION using MACHINE LEARNING TECHNIQUES

Er. Hari K.C.[1]

[1] Department of Electronics and Computer Engineering,
Paschimanchal Campus, IOE, Tribhuvan University
Phone: +9779846104162,
Email: harikc@wrc.edu.np

Abstract
Predicting behaviour of Stock Market is a challenging task. It is a hot topic in machine learning. It assists the investor to minimize the risk and fluctuation in Stock Market. Five years data from different Companies from different sectors stock information's are analyzed. The information such as stock close price, volume of shares, interest rate, stock sentiments are main attributes and features for this research. Regression Machine Learning is applied for predicting the behaviour of Nepal Stock Exchange, NEPSE index which is totally a new concept in Nepal Stock Exchange. NEPSE is one and only stock exchange in Nepal to trade equity and stocks. There are at least 231 public companies currently listed in NEPSE index.

In this research, the performance of Support Vector Regression Kernels (linear, polynomial and RBF) is used to evaluate predicting ability. The various performance parameters such as Moving Average, Interest Rate Correlation, and Stock news sentiment analysis are evaluated. Validation is performed by taking the different timeframe behavior of the stock and comparing the actual and predicted stock prices. The investment decision is made based on the investment strategy developed to allow minimum risk investment for investors. Machine learning predictions for stock market are rare in developing country like Nepal. Different attributes and parameter are taken as suitable to Nepal stock market which provides better prediction result. It ensures that investors invest the money with little or no risk and able to make profit.

Key words: Stock exchange, Big data, Regression, Machine Learning, Financial analysis, NEPSE;

1. INTRODUCTION

1.1. Background

Predicting future is a challenging task when it involves money and risk. The evolution of web 2.0 emphasizes the user generated information in varieties form with no boundaries. This dynamic information can be analyzed and used for prediction. The series of historical data can be analyzed in a suitable manner and further used for predicting the market value. Market performance evaluation is a difficult task where different stock values are determined and analyzed. Numerous Machine learning algorithms and data mining techniques are used to predict the performance of stock market and individual stock prices. There are numbers of stock market in different parts of the world [1]. Some of the popular stock exchanges are NYSE, NASDAQ, KSE, LSE and SSE. Stock exchanges are the financial institutions for transferability of monetary values. The value of a stock is given by its entry on the stock exchange and the volume of its transactions. Share is an indivisible unit of capital or a unit of ownership that represent the equal proportion of a company capital. The more a share is transacted, the more it is valuable, and conversely, if a share is put into transaction in a low volume, it is not so important for some traders and by default its value decreases.

NASDAQ, "National Association of Securities Dealers Automation Quotations" is the largest American stock exchange in the world by market capitalization. It is owned by NASDAQ incorporation founded in February 04, 1971 A.D. It list the various public companies and incorporation with their stock prices, volume of shares, dividend, closing Price, opening Price, maximum Price , minimum Price and so on. Some of the popular Public Companies listed in NASDAQ stock exchange are Apple Incorporation, Facebook Incorporation, and Amazon Incorporation. These companies influence and affect the market behavior whole over the world. Similarly, the only stock exchange in Nepal is Nepal stock exchange indexed as "NEPSE". NEPSE function on NATS,

NEPSE Automated Trading system, a total screen based trading that adopts the basics of an order driven demand. Buying and selling of physical and dematerialized securities is done through NATS. NEPSE was founded in 13th January, 1993 A.D under the company Act, operating under Securities Exchange Act. NEPSE list various sub-indexes including Commercial Bank, Hydropower's, Insurance, Hotels and Manufacturing Companies. The popular public companies listed in NEPSE are Agriculture Development Bank , Arun Hydropower, Everest Bank Limited, Nepal life insurance, Bottlers Nepal Limited etc. The main goal of stock exchanges is to facilitate and promote the growth of capital markets. Stock exchanges provide free marketability and liquidity to the government and corporate securities by facilitating transaction in its addressing floor through members and market mediator such as Agent and Market Maker. There are various factors such as government policy, market sentiment, rise of dollar price, interest rates may have impact on Stock Exchange. NEPSE stock exchange listed 213 companies up to now. Market is closed before 5 pm every day except holidays, Friday and Saturday. During Friday, Saturday and Holidays, there will be no stock openings. Members of Nepal Stock Exchange are given permission to act as mediator in buying and selling of government bonds. Till now, there are fifty member brokers and two market makers, who operate on the trading floor as per the Securities Exchange Act, 1983, rules and bye-laws. Besides this, NEPSE has also provided membership to issue and sales manager securities trader. Issue and sales manager manage the issue and underwriter for public issue of securities whereas securities trader works as individual portfolio manager. The time period of the membership is one year.

Prediction involves use of new and historical data to forecast future values and trends in stock market. Linear regression is the most basic type of regression and commonly used predictive analysis. Regression analysis predicts trends and future values. The regression analysis can be used to get point estimates. Simple linear regression is a technique to summarize and study relationships between two continuous (quantitative) variables .In Linear Regression, machine learns from the features of stock market and shows the correct answer. After the machine is trained, the machine is tested on unseen stock values. Taking the continuous value, finding the best equation to fits the value to forecast the future prices. Regression considers the fluidity of price over time and attempt to forecast the next fluid price in future using the continuous dataset. Support Vector Regression is based on Support Vector Machine. Support Vector Regression gives the solution with lesser computation. It involves exponential, polynomial and sigmoid kernel function. SVM performs non-linear mapping. The kernel functions transform the data into a higher dimensional feature space to make it possible to perform the linear separation.

1.2. Objectives

The objectives for the study are:

- To analyze and predict the stock price for investment decision using regression techniques.
- To compare the accuracy of different Machine Learning regression techniques.
- To compare the actual and predicted stock price.

2. Literature Review

Many research groups are exploring stock market trend prediction using web data analytics. In [1], the researchers had predicted the stock market of Pakistan, Karachi Stock Exchange (KSE) using Naive Bayes, SVR and Decision Tree ML algorithms. Similarly In [2], the researchers uses neural network to predict the stock market of Bangladesh, Dhaka Stock Exchange (DSE). In [3], researchers tried to predict stock price using sentiment analysis and classification technique. They have used some semi-automatic and manual dictionaries like Harvard, Loughran and McDonald, cambria etc. They concluded that this method outperform bag of words model in validation and independent testing data sets. Researchers used Genetic Algorithm to predict stocks of Bombay stock Exchange (BSE) [4]. Researchers tried to compare cumulative emotional valence to average daily stock market returns of firms. They concluded that users with many followers have large impact on intraday returns in contrast with user with few followers have large impact on stock return in future few days. There exist few theories that are valid as well as oppose each other. The theory of random walk says price of a security can't be predicted using the historical data. It supports the argument that the difference between old price and current price of a security is completely independent. On the contrary, the chartist theories say there is some hidden information in the historical prices of a security that gives a clue to future price of that security.

In [5], the researchers have used historical data to predict the position of stock market and they proved that historical data has strong predictive ability. In [6], the research was performed on Asian markets to find out the factors that have concrete impact on market performance. In [7], the researchers have used Artificial Neural Networks (ANN) and statistical technique ARIMA on almost 3 year's data to predict KSE-100 index. The study presented in [8] has done a comprehensive analysis of the underlying relationship between macro-economic factors and KSE market. Similarly, few researches like [9] showed that person's mood plays the critical role in decision making. If the collective mood of the public is found using social media, then this can also help in

predicting the decision they will make about investing the money in the market and thus market performance. In [10], the researchers proposed the use of data collected from different global financial markets with machine learning algorithms to predict the stock index movements. The study [11] made the use of financial NEWS in order to predict market prices. Current and historical NEWS about companies, economic and political events can help in stock prices prediction. Similarly, different statistical techniques have also been used like in [12] the researchers have applied the Dynamic moving average (DMA) on data of Vietnamese Stock Market. A.E. Hassanien et al. [13] proposed a generic rough set predictive model using the data set consisting of daily movements of a stock traded by gulf bank of Kuwait. The objective of this paper is to modify the already existing stock market predictive model based on rough set approach and to construct a rough set data model that would significantly reduce the total number of generated decision rules as compared to [13], keeping the degree of dependency intact. The creation of information table usually consists of several market indicators like closing price, high price, low price, trade, value, average, momentum, disparity in 5 days, price oscillator, RSI (relative strength index) and ROC (rate of change). These indicator acts as the conditional attributes of decision table which have been clearly shown ill stock price movement decision table. Recently, the bankruptcy prediction model has become one of the most important topics in the financial risk management.

3. Research Methodology

3.1. System Model:

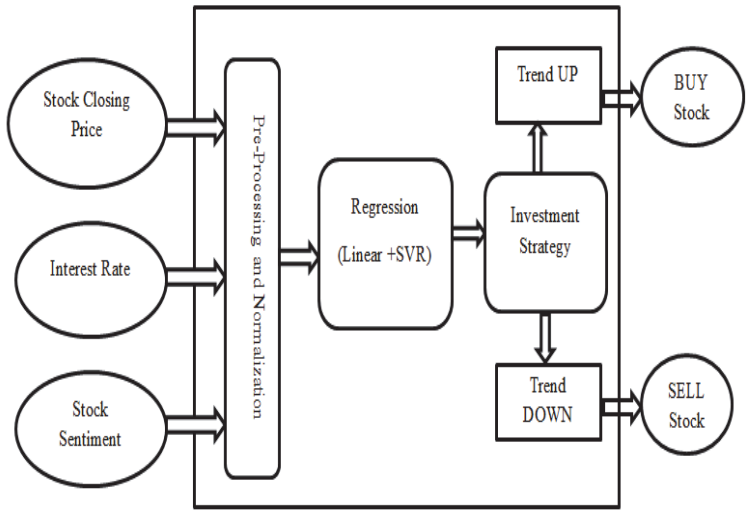

Figure 3.1: Stock Predictive Investment Decision Model

3.2. Main Process Description

The overall process follows the below algorithms and functions.

A) Main Process

1) Start

2) Input stock details, stock sentiment and Interest Rate

3) Perform Preprocessing and Analysis

i) Calculate Moving Average

ii) Calculate stock sentiment index and sentiment values

iii) Calculate interest rate analysis with correlation coefficient.

4) Remove outlier's data using rolling standard deviation.

5) Perform z-score normalization.

$$\text{Close\_price (t)} = ((\text{Close\_price (t)} - \text{Mean}) / (\text{StdDev}))$$

5) Generate Feature set vector.

6) Create Training dataset and testing dataset

7) Perform regression techniques (Linear/Polynomial/RBF) on given dataset

8) Obtain Accuracy of regression techniques

9) Perform Cross- Validation with forward validation testing.

10) Obtain Predicted stock Price

11) Calculate stock prediction error.

[Prediction Error = $\sum$((ASP – PSP)/ ASP) *100 %]

12) Make investment decision

i) Prediction error of time,'t+1' is less than that of time 't'.

a) Indicate stock trend UP

b) Compare the Company stock prices

c) Buy the company stocks

else

a) Indicate stock trend DOWN

b) Sell the Company stocks

13) Stop

B) Radial Basis Function SVR Process

1) Input Dataset.

2) Consider training set and testing set.(70/30 or 88/12)

3) Select center dataset randomly. Calculate width attribute of center dataset.

$w= 1/c\sum (C_j - C_i)^2$

4) Generate the weight factor, [Wt] from the numbers of centers selected

5) Calculate the Gaussian Function

$f(n) =\exp[- (X(n) – C_k)^2 / w]$, where w is width

C_k is center value of centerlist

X(n) is input set

6) Obtain the predicted output values.

$g(n) = Wt * f(n)$, where Wt is assigned weight

7) Calculate Average Mean square error of prediction.

C) Linear Regression Process

1) Input Stock dataset.

2) Check linearity of dataset.

3) Set stock closing price as independent parameter.

4) Estimate the parameter

$Slope=\sum[(Xi – X\ mean) (Yi - Ymean)] / \sum(Xi –Xmean)^2$

$Intercept = Ymean – Slope * Xmean$

5) Determine the output dependent parameter as predicted value.

$Y[ ]= C +M1*X[ ]$ where M1 is slope and C is intercept.

6) Determine Average Prediction Error

D) Polynomial Regression Process

1) Input Stock dataset

2) Check non-linearity of dataset

3) Set stock Closing Price as independent parameter(X)

4) Determine the Polynomial Regression Coefficients(ß) and Intercept.

5) Determine the dependent parameter(Y) as predicted price

$$Y = a + ß1*X + ß2* X^2 + ß3* X^3$$

6) Determine Average Prediction error.

3.3. Data Collection:

Basically the data is secondary data. It is collected through website using application program interface. Different information of public companies are collected which include stock volume, open price, close price, maximum price, minimum price over different series of years. Interest rate is obtained from Nepal Rastra Bank . Similarly, the stock news are obtained from twitter, and stock news website "sharesansar.com". Nepal stock exchange "nepalstock.com" provides the historical and present stock data needed for our system. They are the source of information for stock values of Public companies.

4. Analysis and Result

4.1. Correlation Analysis

The Correlation Analysis between Interest Rate and Closing Price of ADBL, ALICL and AHPC and TRH stock is as given below.

Table 4.1: Correlation between interest rate and stock price.

Company	Correlation Coefficient with Interest Rate
ADBL	-0.609354
ALICL	-0.859071
AHPC	0.620511

This table shows the Comparison of correlation between the various companies with respect to interest rate .In case of ADBL and ALICL , the correlation coefficient is negative which means that when interest rate increases , the closing price of ADBL, ALICL and AHPC decreases and vice-versa. This reflects the better scenario of stock market. But in case of AHPC, it doesn't have negative correlation coefficients which indicate that AHPC closing price doesn't depend on interest rate. This mean Investor doesn't look after the interest rate before investing in Hydropower such as AHPC.

4.2. Moving Average Analysis

A) Moving Average Analysis of ADBL

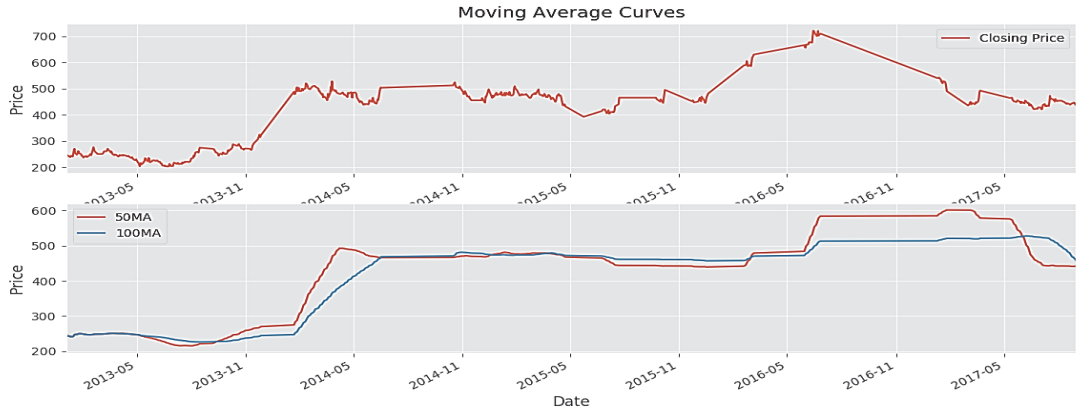

Figure 4.6: ADBL Moving Average Analysis

This graph shows the moving average analysis of ADBL. The upper portion of graph is about closing price of ADBL and lower portion graph shows 50days and 100 days moving average. During the period 2013-11 to 2013-05 and 2016-05 to 2017-05 ,the 50MA line crosses and above 100MA which means change in stock trend .These crossing can be used as trading signals that financial security is changing direction.

B) Moving Average Analysis of ALICL

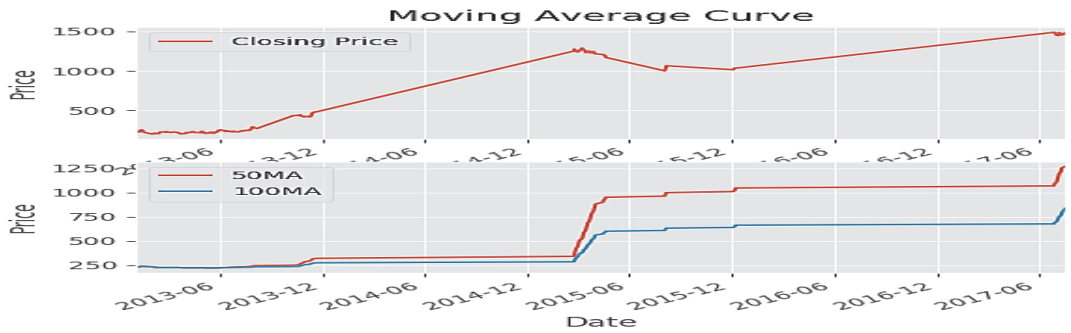

Figure 4.7: ALICL Moving Average Analysis Graph

C) Moving Average Analysis of AHPC

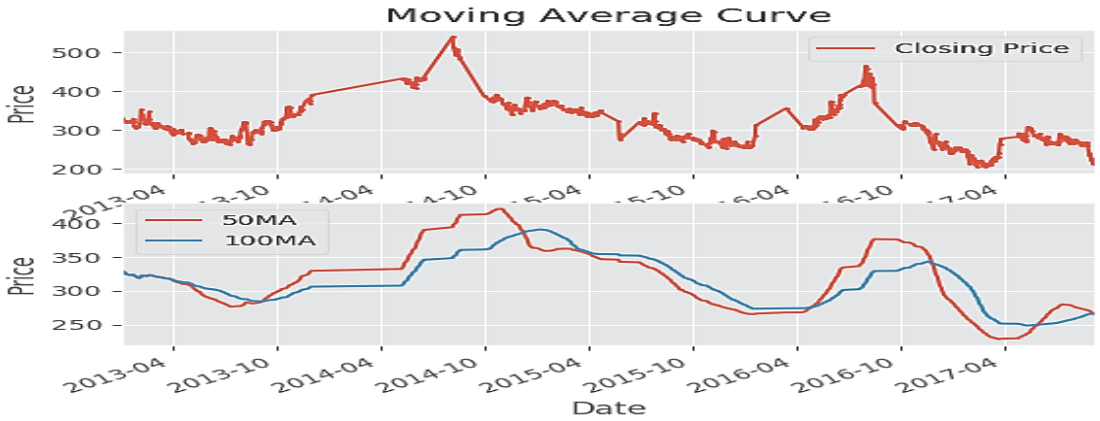

Figure 4.8: ALICL Moving Average Analysis Graph

This graph shows the lots of crossing between 50 days moving average and 100 days moving average. It means stock trend seems to changing greatly over the different time period. This also means the number of times the closing price of AHPC is decreasing and increasing. So, when the moving average of AHPC is compared with than that of ALICL and ADBL , it is seen that there are lots of fluctuations in AHPC stock closing price.

4.3. Stock Sentiment Analysis

A) Stock Sentiment Analysis of ADBL

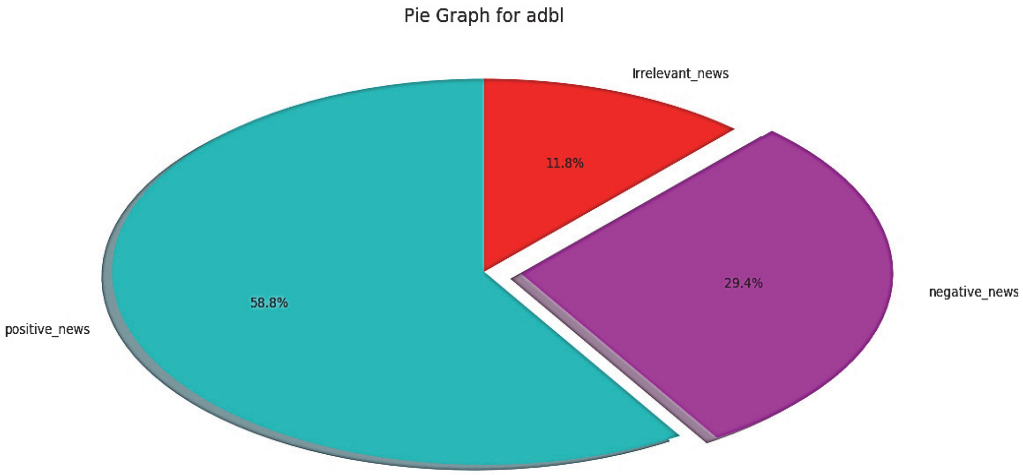

Figure 4.10: Sentiment Pie Chart for ADBL

The sentiment index for adbl is 64.70 percent. The positive news are 58.8 % , negative news are 29.4% and irrelevant news are 11.8%. So, the sentiment index is evaluated using formula of equation. The sentiment index for adbl is 64.70 percent which is really impressive. This shows that there is no negative impact of Company stock news in future.

B) Stock Sentiment Analysis of ALICL

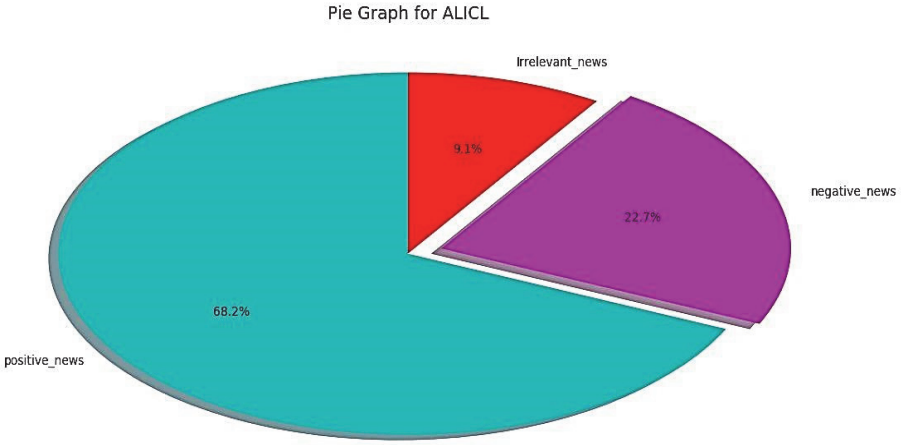

Figure 4.11: Sentiment Pie Chart for ALICL

The positive news are 68.2 % , negative news are 22.7 and irrelevant news are 9.1 so, the sentiment index is evaluated usin formula of equation .The sentiment index for alicl is 72.72 % which is really impressive. This shows that there is no negative impact of Company stock news in future.

C) Stock Sentiment Analysis of AHPC

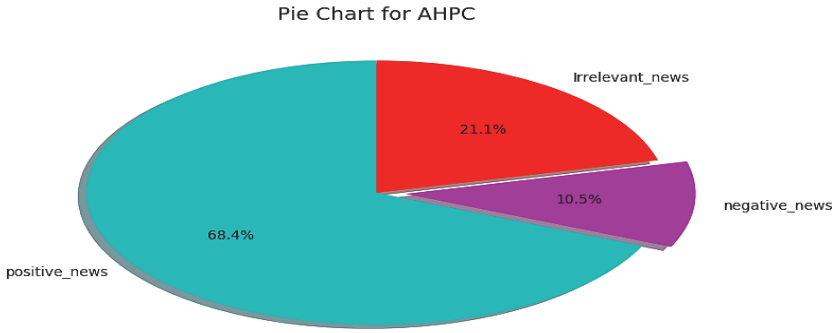

Figure 4.12: Sentiment Pie Chart for AHPC

The positive news are 68.4 % , negative news are 10.5 % and irrelevant news are 21.1 %. So, the sentiment index is evaluated using formula of equation. The sentiment index for AHPC is 78.94 percent which is really impressive. This shows that there is no negative impact of Company stock news in future.

4.4. Data Coverage of Different Regressions

A) ADBL data coverage.

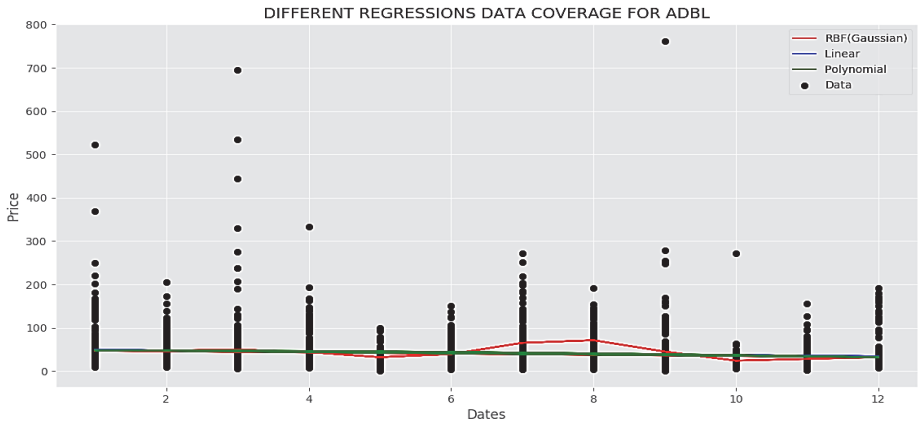

Figure 4.14: Data Coverage for ADBL

This figure shows data coverage. In this graph, it is clear that RBF has covered most of the data than other regression. Linear has least coverage. It means data possess non- linearity characteristics.

B) ALICL data Coverage

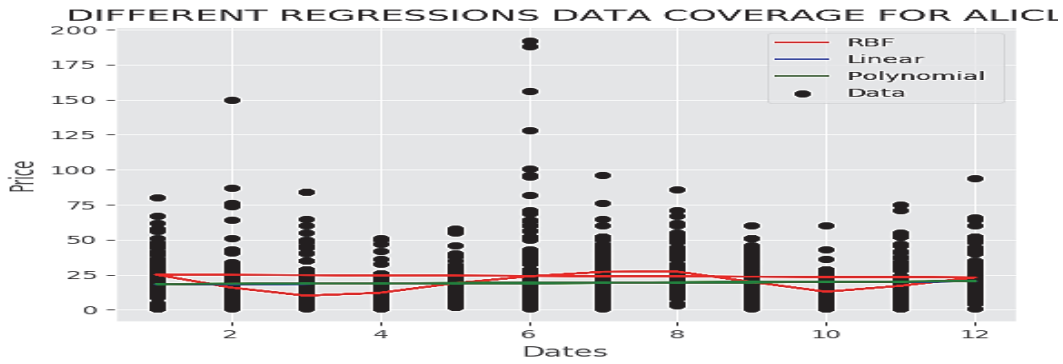

Figure 4.15: Data Coverage for ALICL

This figure shows data coverage. In this graph, it is clear that RBF has covered most of the data than other regression. It shows multiple Gaussian coverage or fitting. It means data possess non- linearity characteristics.

C) AHPC data Coverage:

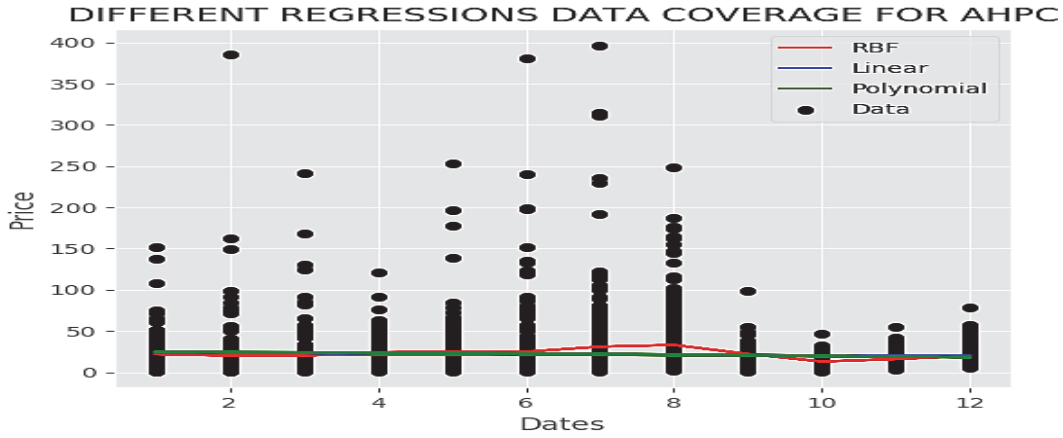

Figure 4.16: Data Coverage for AHPC

This figure shows data coverage of AHPC. In this graph, it is clear that RBF regression has covered most of the data than other regression. Linear regression has least data fitting.

4.5. Performance Comparison and Validation of Stock

4.5.1. Performance Comparison of Regression Techniques

Table 4.2: Prediction Performance of various Regression Techniques

NEPSE stock exchange				KSE-100	S &P 500
Regression/ Companies	Linear Accuracy	RBF Accuracy	Polynomial Accuracy	RBF Accuracy	RBF Accuracy
1) ADBL	13%	80%	41%	63%	66%
2) ALICL	5%	77%	47%		
3) AHPC	21%	85%	67%		

From the table, differences between the models are very significant. In addition, from the experiment it is also known that SVR RBF regression computes faster than linear and polynomial regression. So, RBF model have better learning ability and forecasting capability. Also the accuracy obtained between NEPSE and KSE-100 stock exchange using SVR techniques is little different. This difference may be due to the fact that in KSE-100 stock exchange, data is taken of only 3 months and include various parameter such as foreign exchange , Commodity prices, Investor sentiment which are not included in this research. Similarly, S & P 500 index RBF accuracy is similar to NEPSE but with low accuracy.

4.5.2. Actual Price and Predicted Price Comparison with Validation

Cross Validation is a validation to generalize dataset using forward validation testing. It is used to estimate the accuracy of prediction. Training dataset is known dataset and testing dataset is unknown dataset. It helps to test the prediction to limit overfitting. It involves partitioning the data into subsets, performing analysis and validating the testing set. Here, K- fold cross validation is used that divides the original dataset into K equal sized subdataset. Out of the K subdataset, a single dataset is used for validation for testing the prediction. Here K=30 is taken. For our experiment, at each iteration 70 percent of data used for training and 30 percent data are used for testing

A) Agricultural Development Bank Ltd

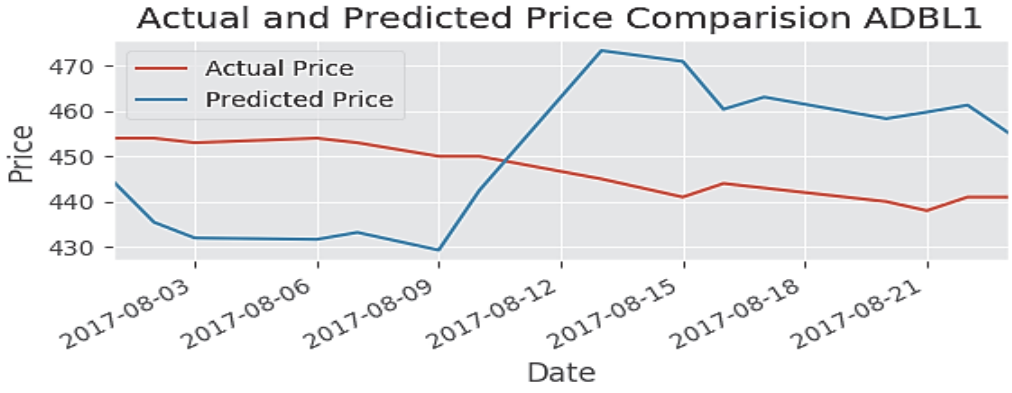

Figure 4.18: Comparison between actual and predicted price ADBL1

Average Prediction Error = $(|(Actual_{Average} - Predicted_{Average}) / Actual_{Average}|) * 100\%$

$=|(446.73-450.05)/ 446.73| * 100\%$

$=0.74\ \%$

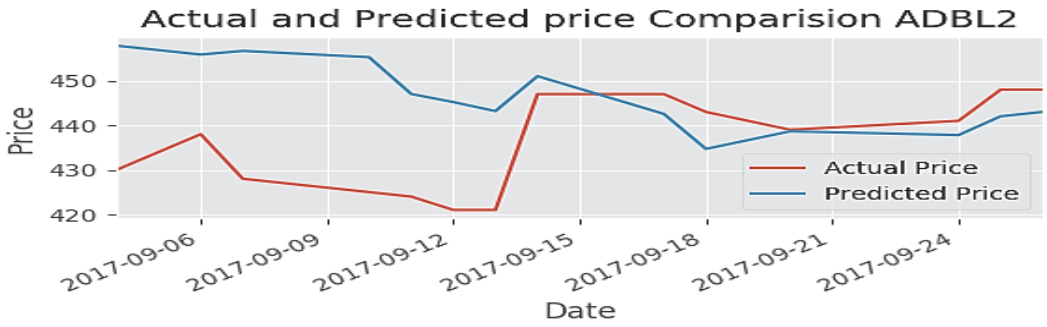

Figure 4.19: Comparison between actual and predicted price ADBL2

Average Prediction Error = $(|(Actual_{Average} - Predicted_{Average}) / Actual_{Average}|) * 100\%$

$= (|436.86 - 446.39| / 436.86) * 100\%$

$= 2.08\ \%$

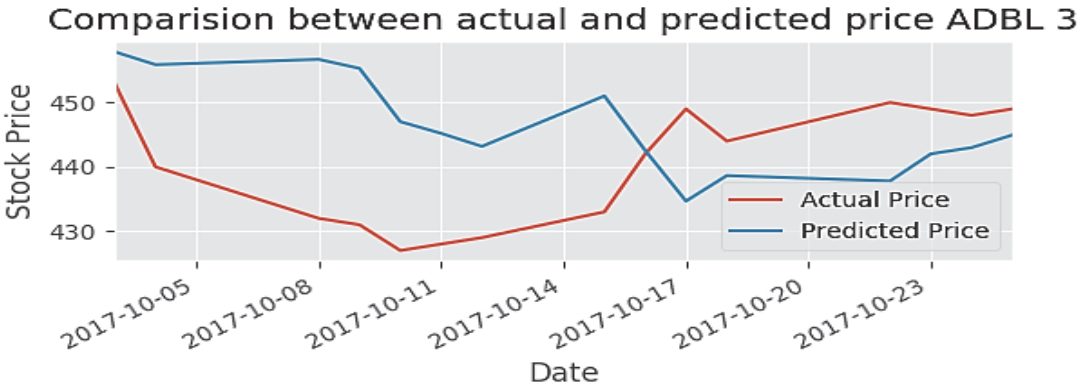

Figure 4.20: ADBL Actual and predicted price comparison

Average Prediction Error = $(|(Actual_{Average} - Predicted_{Average}) / Actual_{Average}|) * 100\%$

$= (|440.26 - 446.38| / 440.26) * 100\%$

$= 1.39\ \%$

B) Asian Life Insurance Company Ltd.

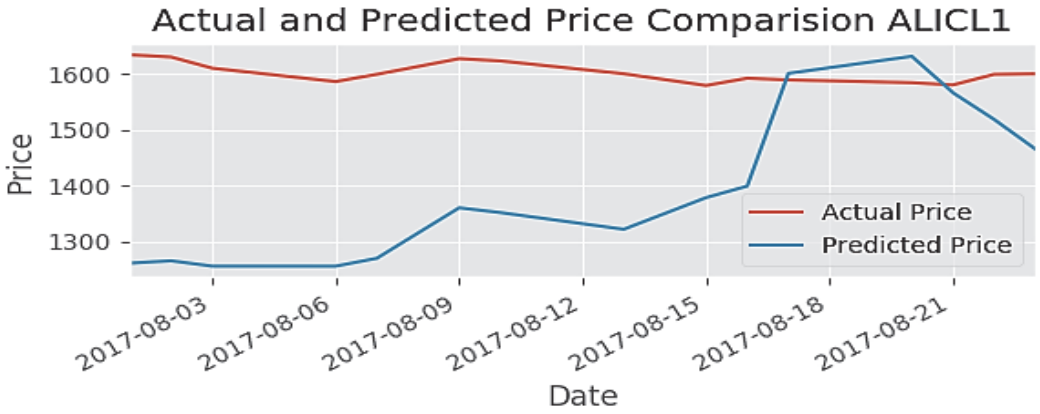

Figure 4.21: Comparison between actual and predicted price ALICL 1

Average Prediction Error =$(|1603.13 - 1394.22| / 1603.13) * 100\%$

$= 13.03\ \%$

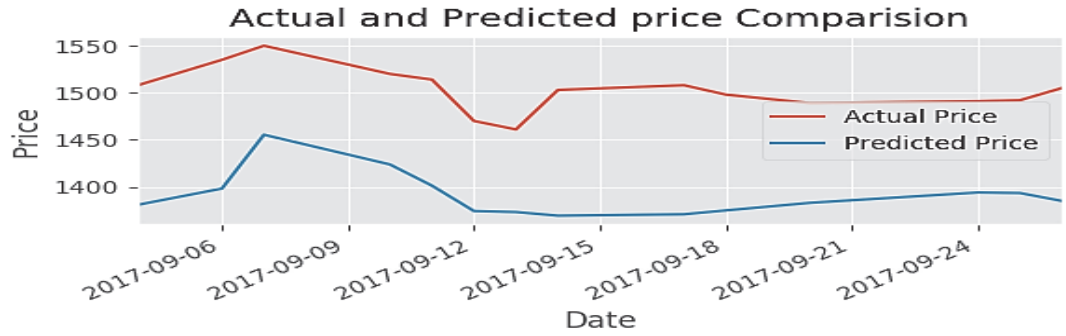

Figure 4.22: Comparison between actual and predicted price ALICL 2

Average Prediction Error = ((|1503.14 – 1391.22|) / 1503.14) *100 %

= 7.44 %

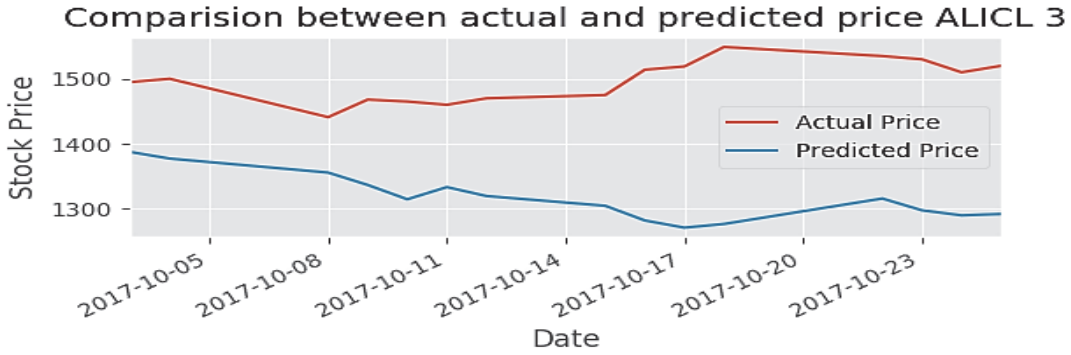

Figure 4.23: ALICL actual and predicted price comparison

Average Prediction Error = ((|1496.73 – 1316.91|) / 1316.91) *100 %

= 13.65 %

C) ArunValley Hydropower Company Ltd.

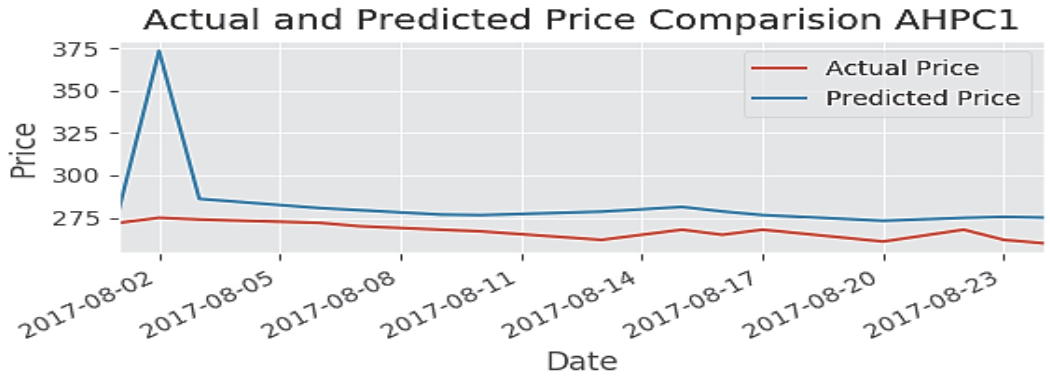

Figure 4.24: Comparison between actual and predicted price for AHPC 1

Average Prediction Error = ((|267.46- 284.46|)/ 267.46) * 100 %

=6.35 %

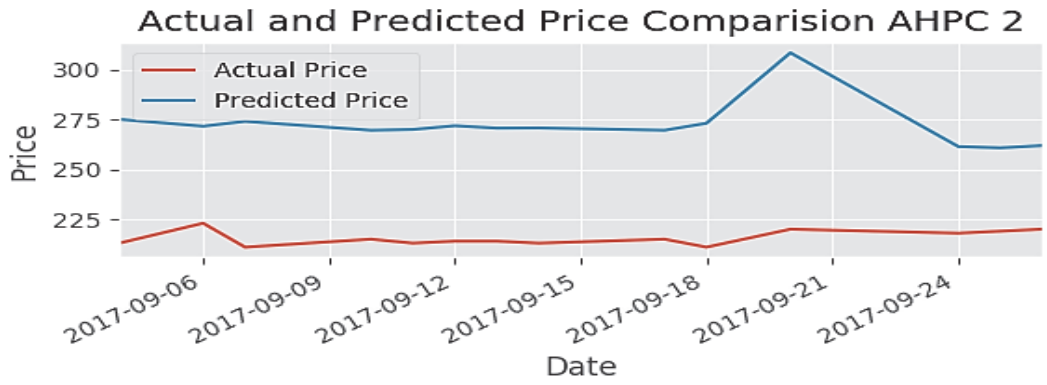

Figure 4.25: Comparison between actual and predicted price for AHPC 2

Average Prediction Error = ((|215.64- 272.08|)/ 215.64) * 100 %

= 26.17 %

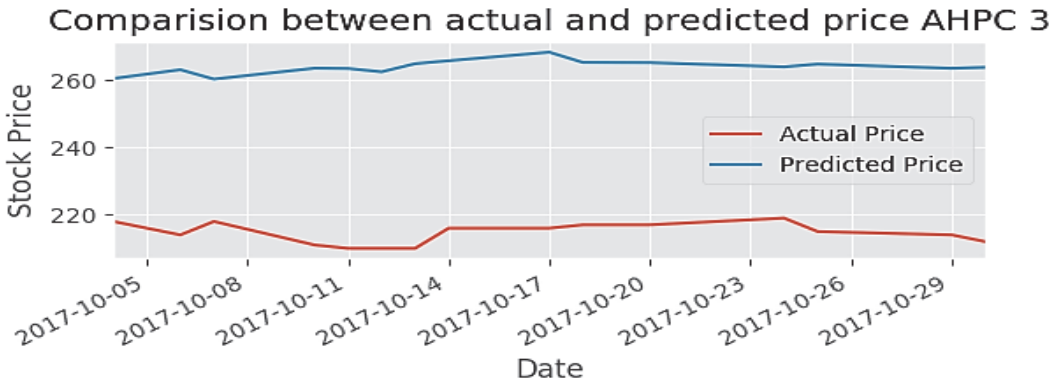

Figure 4.26: Actual and predicted stock price comparison

Average Prediction Error = ((|214.46- 264|)/ 214.46) * 100 %

=23.09 %

4.7.4. Comparison among four Companies for Investment Decision

Table 4.3: Setting Investment priority among four companies

Companies/Parameter	ADBL	ALICL	AHPC
1) Stock Sentiment Index	64.70 %	72.72%	78.94%
2)Moving Average	100MA<50MA	100MA<50MA	100MA=50MA
3) Accuracy	80%	77%	85%
4)InterestRate Correlation	-0.609354	-0.859071	0.62051
5)Maximum Average Prediction Error	2.08%	13.65%	26.17%

In this table ,there are five important parameters to decide whether to invest to any of these companies. The companies considered are ADBL,ALICL and AHPC . AHPC has maximum average prediction error, so investors need to consider other parameters to purchase and sell the stock of AHPC. ALICl has low accuracy with high average prediction error. so, it will be not better to invest in ALICL.

Similarly, the most appropriate company to invest is ADBL since its prediction error is low with predicted price high compare to other companies stock price . Regression accuracy is also high compare to other. In moving Average Analysis, the historical data also shows the best stock trends than before. Prediction error over the sample period is also very low compare to other.

5. Conclusion and Future Work

Finally, it has been concluded that predicting the stock price using RBF regression gives maximum accuracy and less prediction error than other regressions. The three important factors such as historical stock data, interest rate and stock sentiment are taken into consideration for stock market prediction. This regression technique gives similar accuracy when comparing the NEPSE stock market prediction with KSE-100 stock exchange.

The three companies listed in NEPSE are compared with number of parameters to set the investment priority for the investors. According to the result, ADBL has the highest investment priority compared to other companies. So, investor can invest their money in ADBL compared to other companies considered with less risk and high return.

As a future work, Optimized RBF technique can also be used that optimize spread and width parameter. Similarly, the new kernel function or technique can be designed which gives better performance with low error rate than these techniques.

6. References

[1] K. Raza, "Prediction of Stock Market using Machine Learning Techniques," in DOI:10.1109/7916583 .IEEE, 2017.
[2] M. Billah, S. Waheed and A. Mahifa, "Stock Market Prediction Using Improved Training Algorithm of Neural Network," in 987-1-5090-5785-6/16 ,2016/IEEE, Dec 2016.
[3] Z. Gaowei, X. Lingyu and Y. L. Xue, "Model and Forecast Stock Market Behaviour Integrating Investor Sentiment Analysis and Transaction Data," in 20:789, vol:20Cluster Computing. DOI: 10.1007/50586-017-0803-X, 2017.
[4] Y. Perwej and A. Perwej, "Prediction of Bombay Stock Exchange Market Returns using ANN and Genetic Algorithm," in JILSA, 2014.
[5] A. Lu and A. Mackinley, "Stock Market Prices do not follow Random Walks," in doi.org/10.1093/rfs/1.1.41 . Oxford Accademic, 03 April,2015.
[6] C. Arurangzeb, "Factors Affecting Performance of Stock Market Evidence From South Asian Countries," in IJARBSS, 02 Nov,2012.
[7] S. Fatima and G. Hussain, "Statistical Modeling of KSE-100 Index using Hybrid financial System," in NeuroComputing. vol:71. pp 2742-2746, 2008.
[8] I. Ali, K. Rehman, M. Khan and H. Afzal, "Casual Relationship between macro economic indicators and stock exchanges prices in Pakistan," in African Journal of Business Management vol 04, no:3 pp 312-314,2010, 2010.
[9] J. Bollen, H. Mao and X. Zeng, "Twitter Mood Predict the Stock Market," in Journal of Computational Science vol 02. n0:01. pp 1-8, 2011.
[10] S. Shen, H. Jiang and T. Zhang, "Stock Market Forecasting using Machine Learning Algorithm," in Standford University, 2012.
[11] S. Asur and B. Huberman, "Predicting the Future with Social Media," in ICWIIAT, 2010.
[12] N. Hung and Y. Zhawjun, "Profitability of Applying Simple Moving Average Trading Rules for Vietamese stock Market," in vol 02, no:03,pp 22-31Journal of Business Management, 2013.
[13] S. Sarkar, "An improved rough set data model for stock model prediction," in DOI: 10.1109/ICBIM.2014.6970963. IEEE, 09 Jan,2014.
[14] S. Bobde and V. Rajput, "Stock Market Prediction using Hybrid Approach," in SBN:987-1-5090-1666. 2016/IEEE, 2016.