Proceedings of IOE Graduate Conference, 2016
pp. 337–339

Analysis and Prediction of Stock Prices of Nepal using different Machine Learning Algorithms

Pushkar Khanal [1], Shree Raj Shakya [2]

[1,2] *Department of Mechanical Engineering, Pulchowk campus, Institute of Engineering, Tribhuvan University, Nepal*
Corresponding Email: [1] pushkarkhanal133@gmail.com

Abstract
Stock price data of from past four years were extracted and different machine learning algorithm were applied accordingly. The stock price data of most volatile to most stable prices were taken for analysis with hydropower stock being the most stable and Insurance prices the most volatile. The algorithm called Support Vector Machine that was combined with boosting produced the most accurate results with an accuracy of 65%. This learning algorithm predicted with more accuracy than the existing technical analysis methods.

Keywords
Stock Price – Machine Learninig Algoritms – Support Vector Machines – Regression

1. Introduction

Due to recent advancement in machine learning algorithms and artificial intelligence, these techniques are being applied to predict the pattern and analyze the trend in stock prices of various companies all around the world. The increase in accuracy, reliability and modifiability has also increased the reliance on intelligence trading system to help in predicting and analyzing the stock prices in different situation [1].

Stock prices are vulnerable to small and quick changes because of the underlying nature of the financial and economical structure that it is based upon. Stock price are very volatile as it is dependent upon many known parameters like previous day's closing price, P/E ratio, etc and also upon unknown factors like political upheavals, elections, business cycles, etc[2]. Due to such factors it is very hard to accurately predict the prices of stock market with sheer confidence. A trader in stock price market can use this algorithm to predict the prices which are going to rise and sell subsequently sell it before its price falls [3].

1.1 Stock Price Prediction Theory

The working of stock prices is generally dependent on two theories, random walk hypothesis and efficient-market hypothesis.

1.1.1 The Efficient Market Hypothesis (EMH)

The EMH hypothesizes that given the past trading history of the stock it is impossible to predict the stock price future prices as markets are efficient and current prices reflect all information. Thus predicting the stock price is just a mere game of chance rather than the careful use of skills and intellect[4].

1.1.2 The Random Walk Hypothesis

The random walk Hypothesis claims that past stock prices data cannot be exploited to predict the future prices because trends do not exist in stock prices as market is highly efficient. However, because of the advent of advanced machine learning algorithms and powerful computing resources, trading systems can exploit the underlying price patterns with an availability of enough amounts of data. With the easy availability of large amount of stock data, it is possible to challenge the efficient market hypothesis and predict stock prices with more accuracy and reduced time[5].

2. Methodology

2.1 Stock Prediction Methodologies

The prediction methods on the basis of type of data and the type of tool that each method is using to predict the market are categorized as:

1. Technical Analysis Methods
2. Fundamental Analysis Methods
3. Traditional Time Series Prediction Methods
4. Machine Learning Methods

2.2 Models for prediction and analysis of stock price

1. Moving Average
2. Single exponential smoothing
3. Linear Regression
4. Multiple Regression
5. Support Vector Machine
6. Decision Tree
7. Correlation

The research Methodology is summarized in a flowchart as shown in figure 1.

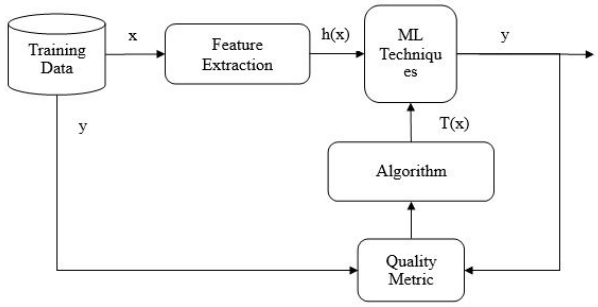

Figure 1: Methodology of Research

2.3 A flowchart of Machine learning Algorithm

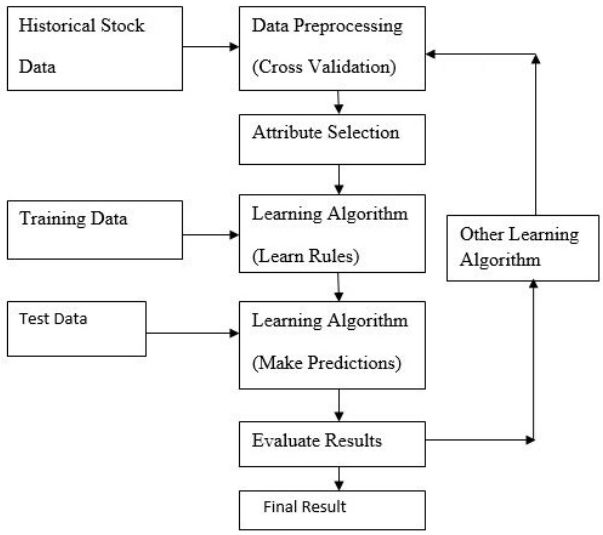

Figure 2: Flowchart of machine learning algorithm

2.4 Model Selection

Selection based on K-fold Cross validation

1. 1st model – Simple Linear Regression
   Output: Y
   Input: x – Different features
   $Y = mx + c$
2. 2nd model – Multiple Regression Model
   $Y = C + a_1 \times 1 + a_2 \times 2 + \ldots$
   Where $1, 2, 3, \ldots$ are features.
3. Quadratic Model – Extended form of above model using squared power of each features
4. Sinusoidal Model
5. SVM model
6. Decision Tree

2.5 Preprocessing of Historical Stock Data

The historical stock data was obtained from sharesansar.com. The stock prices of four different primary stocks of most volatile and stable stock prices were taken.

The available data of the stock had the following attributes
Date - LTP - %Change - High - Low - Qty - Turnover

The recent prices of stock are going to have a more impact on the price of today's stock and were subsequently given more weight compared to the recent ones. To proceed, we go along the time-line that is closer to today's price. We take date as x-axis with integer values attached to it, giving the most recent date the higher weight as compared to the later ones.

2.6 Method of validating different models

2.6.1 K-fold Cross Validation

Training data is divided into 10-folds. 9 out of 10 folds is used for training and 1 fold is used for testing. Training and test data is divided into the ratio of 4:1.

3. Result and Discussion

The result of Linear regression model is shown in table below.

Table 1: Linear Regression Results

Correlation Coefficient	0.785
Mean Absolute Error	11.23
Root Mean Squared Error	30.253
Relative Absolute Error	11.235
Root Relative Squared Error	25.236
Total Number of Instances	367

The result of Multiple Regression model is presented in table below.

Table 2: Multiple Regression Results

Number of Instances	232
Number of features	5
Number of Coefficients	6
Max Error	41%
RMSE	9%

The Result of Support vector machines is presented table below.

Table 3: Support Vector Machines

Root Mean Squared Error	0.456
Accuracy	59%

Table 4: Boosting

Root mean square error	0.576
Accuracy	65%

4. Conclusion

The algorithm called Support Vector Machine that was combined with boosting produced the most accurate results in comparison to most other machine learning algorithms and traditional technical methods. Linear regression also yielded lower mean squared errors while predicting the EMA pattern.

References

[1] V.H. Shah. Machine learning techniques for stock prediction. 2007.

[2] Forecasting the klse index using neural networks. 1995.

[3] S. Chaigusin. An investigation into the use of neural networks for prediction of stock exchange of thiland. 2011.

[4] Ali Ghezelbash. A new approach based on artificial intelligent models for forecasting of fininancial regression. 2015.

[5] M.D. Rechenthin. Machine learning classification techniques for the analysis and prediction of high frequency stock. 2014.