

Stock Price Trend Forecasting Using Supervised Learning Methods

Dr. K. SUJATHA

Professor, Department of Computer Science and Engineering,
Dadi Institute of Engineering and Technology (DIET), Anakapalle
Jawaharlal Nehru Technological University, Kakinada

V. SYAMALA

M.Tech Student, Department of Computer Science and Engineering,
Dadi Institute of Engineering and Technology (DIET), Anakapalle
Jawaharlal Nehru Technological University, Kakinada

Abstract: *The aim of the project is to examine a number of different forecasting techniques to predict future stock returns based on past returns and numerical news indicators to construct a portfolio of multiple stocks in order to diversify the risk. We do this by applying supervised learning methods for stock price forecasting by interpreting the seemingly chaotic market data. Proposed system use an efficient supervised learning method for future stock prediction. The advancement in technology provides an opportunity to gain steady fortune from stock market and also can help experts to find out the most informative indicators to make better prediction. Admin collect the past stock market record and stored in a database. The prediction of the market value is paramount importance to help in maximizing the profit of stock option purchase while keeping risk low.*

Proposed achieved expected predicted values by using efficient regressor methods. Regressor methods using the past stock market values for prediction. Taking average of past stock values and applied to the required regressor methods. User can easily predict the future values by using our application. We'll dive into the implementation part of this article soon, but first it's important to establish what we're aiming to solve. Broadly, stock market analysis is divided into two parts – Fundamental Analysis and Technical Analysis. Fundamental Analysis involves analyzing the company's future profitability on the basis of its current business environment and financial performance.

Keywords: *Stock Price, Machine Learning, Web Scrapping*

1. INTRODUCTION

Predicting the stock price trend by interpreting the seemly chaotic market data has always been an attractive topic to both investors and researchers. Among those popular methods that have been employed, Machine Learning techniques are very popular due to the capacity of identifying stock trend from massive amounts of data that capture the underlying stock price dynamics.

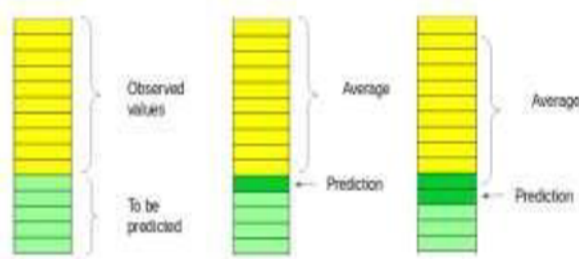
In this project, we applied supervised learning methods to stock price trend forecasting. According to market efficiency theory, US stock market is semi-strong efficient market, which means all public information is calculated into a stock's current share price, meaning that neither fundamental nor technical analysis can be used to achieve superior gains in a short-term (a day or a week). Indeed, our

initial next-day predication has very low accuracy around 50%. However, as we tried to predict long-term stock price trend, our models achieved a high accuracy (79%). Based on our prediction result, we built a trading strategy on the stock, which significantly outran the stock performance itself. The stock market is basically an aggregation of various buyers and sellers of stock. A stock (also known as shares more commonly) in general represents ownership claims on business by a particular individual or a group of people. The attempt [3] to determine the future value of the stock market is known as a stock market prediction. The prediction is expected to be robust, accurate and efficient. The system must work according to the real-life scenarios and should be well suited to real-world settings. The system is also expected to take into account all the variables that might affect the stock's value and performance.

There are various methods and ways of implementing the prediction system like Fundamental Analysis, Technical Analysis, Machine Learning, Market Mimicry, and Time series aspect structuring. With the advancement of the digital era, the prediction has moved up into the technological realm. The most prominent and [3] promising technique involves the use of Artificial Neural Networks, Recurrent Neural Networks that is basically the implementation of machine learning. Machine learning involves artificial intelligence which empowers the system to learn and improve from past experiences without being programmed time and again. Traditional methods of prediction in machine learning use algorithms like Backward Propagation, also known as Back propagation errors. Lately, many researchers are using more of ensemble learning techniques.

It would use low price and time [3] lags to predict future highs while another network would use lagged highs to predict future highs. These predictions were used to form stock prices. [1] Stock market price prediction for short time windows appears to be a random process. The stock price movement over a long period of time usually develops a linear curve. People tend to buy those stocks whose prices are

expected to rise in the near future. The uncertainty in the stock market refrain people from investing in stocks. Thus, there is a need to accurately predict the stock market which can be used in a real-life scenario. The methods used to predict the stock market includes a time series forecasting along with technical analysis, machine learning modeling and predicting the variable stock market. The datasets of the stock market prediction model include details like the closing price opening price, the data and various other variables that are needed to predict the



object variable which is the price in a given day. Predicting how the stock market will perform is one of the most difficult things to do. There are so many factors involved in the prediction – physical factors vs. physiological, rational and irrational behavior, etc. All these aspects combine to make share prices volatile and very difficult to predict with a high degree of accuracy.

```
#plot valid ['Prediction'] = 0
valid ['Predictions'] = preds
plt.plot(train['Close'])
plt.plot(valid[['Close',
'Predictions']])
```

2. RELATED WORK

Design and implementation of NN5 for Hong Kong stock price forecasting is proposed by M.T.Phipp, K. Poul, S.O. Choy, K. Reggie, S.c. Ng, I. Mark T.Jonathan, K. Kai, and W. Tak-Lam et al [1]. Various distributed methods have developed in the exchanging group for stock forecast undertakings. Among them is neural system (NN). In this project, the hypothetical foundation of NNs and the back propagation calculation is looked into. In this way, an endeavor to manufacture a stock purchasing/offering ready framework utilizing a back propagation NN, NN5, is displayed.

The framework is tried with information from one Hong Kong stock, The Hong Kong and Shanghai Banking Corporation (HSBC) Holdings. The framework is appeared to accomplish a general hit rate of more than 70%. Various exchanging methodologies are talked about. A best system for exchanging non-unpredictable stock like HSBC is suggested. Forecasting Stock Performance using

Intelligent Hybrid Systems is proposed by X. Wu, M. Fund and A. Flitman et al [2]. Foreseeing the future has dependably been one of humankind's wants. As of late, counterfeit savvy strategies, for example, Neural Networks, Fuzzy Logic, and Genetic Algorithms have picked up fame for this sort of uses. Much research exertion has been made to enhance the forecast precision and computational productivity. In this project, a hybridized neural systems and fluffy rationale framework, in particular the Feed Forward NeuroFuzzy (FFNF) display is proposed to handle money related estimating issue. It is discovered that, by separating a vast issue into reasonable "lumps", the proposed FFNF display yields better execution regarding computational proficiency, expectation precision and speculation capacity.

It additionally conquers the dark craftsmanship approach in customary NNs by joining "straightforwardness" into the framework. Stock Market Prediction System with Modular Neural Network is proposed by T. Kimoto, K. Asawaka, M. Yoda, and M. Takeoka et al [3]. a purchasing and offering timing expectation framework for stocks on the Tokyo Stock Exchange and examination of intenal portrayal. It depends on measured neural networks[1][2]. We built up various learning calculations and forecast strategies for the TOPIX (Tokyo Stock Exchange Prices Indexes) expectation framework. The forecast framework accomplished precise expectations and the recreation on stocks tradmg demonstrated a phenomenal benefit. The forecast framework was created by Fujitsu and Nikko Securities. Application of Neural Network to Technical Anaysis of Stock Market Prediction is proposed by H. Mizono, M. Kosaka, H. Yajma, and N. Komoda et al [4].

All through the tool compartment accentuations is set on neural system standards that development to or are themselves utilized as a part of designing, budgetary, and other down to earth applications. Predicting the Nigerian Stock Market Using Artificial Neural Network is proposed by S. Neenwi, P. O. Asagba, L. G. Kabari et al [7]. Anticipating a monetary time arrangement, for example, securities exchange patterns, would be an essential advance when creating venture portfolios. This progression is exceptionally testing because of unpredictability and nearness of a huge number of components that may influence the estimation of specific securities.

In this examination project, we have demonstrated by logical inconsistency that the Nigerian securities exchange isn't proficient however confused. Two years agent stock costs of a few banks stocks were broke down utilizing a bolster forward neural system with back-proliferation in Matlab 7.0. The recreation results and value estimates demonstrate that it is conceivable to reliably procure great degrees of profitability on the

Nigerian securities exchange utilizing private data from a simulated neural system marker.

3. Methodology

Stock market is a collection of buyers and sellers which show their interest with the trading of stocks which are released by the companies for elevating the capital and are bought by the investors in order to get a portion of the company. Stock Market is always aggressive, you don't know what will happen next as it is very difficult to predict the future stock price of the companies since it keeps fluctuating every day. As we all know a company's stock depends on many factors and taking into account the nonlinearities and discontinuities of the factors which are considered to impact stock markets. Some of the reasons are like company related news, political events natural disasters etc. stock price prediction is one of the most important issues to be investigated in academic and financial researches [1]. Multiple Regression Analysis, Principal Component Analysis, Hurst Exponent Analysis, and Grey Relation Analysis can be used during this initial step [Ince and Trafalis, 2007; Hurst, 1951; Kung and Wen, 2007]. This Model of Prediction will focus exclusively on predicting the daily trend (price movement) of individual stock .

This Model will make no attempt on deciding how much money to allocate to each prediction. In this present model Regression, Web Scrapping, Differential Evolution Optimization-based Fuzzy type-2 Clustering are considered where regression which can be linear and multiple is a predictive method. The remainder of the paper is organized as follows: Section 2 describes the web scrapping method to extract data from the Bombay Stock Exchange [BSE] website itself. Section 3 describes the application of Regression which consist of Multiple and linear regression analysis for the prediction. The structure of the Differential Evolution Optimization based Fuzzy type-2 Clustering system that is implemented in this paper is then presented in Section 4. The experiments and empirical results of proposed model are presented and described in Section 5. Finally, Section 6 provides a conclusion.

Web Scrapping: Web Scrapping or Web Harvesting is a method of extracting data from the websites. It may access the World Wide Web directly using the Hypertext transfer protocol, or through a web browser. In this proposed model web scrapping is used to extract the stock prices of a certain company you want from the Bombay stock exchange website at any time, as you know prices fluctuates every day and will save it in a csv file which will be our training data for the regression analysis and it there will be uncertainty in the prices along with dates Then, here the Differential Evolution

Optimization based Fuzzy type-2 Clustering system will start working.

Regression: Regression is a predicting method whose outcome is based on the given input. The simplest regression Technique is linear regression whereas advanced regression technique is multiple regressions. Future stock market prices. Nonetheless, using too many financial and economical factors can overload the prediction system [Thawornwong and Enke, 2003; Hadavandi et al., 2010; Chang and Liu, 2008; Esfahanipour and Aghamiri, 2010].

As a Result , in initials to make our prediction strong and more exact we have the challenge to determine only those inputs variable which have the strongest forecasting ability, basically the main factors on which a company profile depends. IN our proposed model multiple regression analysis is performed on 24 economic variables to both reduce the dimensionality of the variable set and identify which variables have a strong relationship with the market price of the BSE index for the subsequent testing period [2]. Example variables include past prices, volume, technical indicators, T-bill rates, certificate of deposit rates, credit ratings, producer/consumer price indexes, industrial production levels, and money supply levels.

Linear Regression Linear regression is a statistical Technique use to predict the relationship between the dependent and independent variable and is represented as $V=Y+WX$ where V is the dependent variable and X is the independent variable, Y is a constant and W is the slope of the regression line. The most basic machine learning algorithm that can be implemented on this data is linear regression. The linear regression model returns an equation that determines the relationship between the independent variables and the dependent variable.

4. WOK FLOW

Data Collection: Data collection is a very basic module and the initial step towards the project. It generally deals with the collection of the right dataset. The dataset that is to be used in the market prediction has to be used to be filtered based on various aspects. Data collection also complements to enhance the dataset by adding more data that are external. Our data mainly consists of the previous year stock prices. Initially, we will be analyzing the Kaggle dataset and according to the accuracy, we will be using the model with the data to analyze the predictions accurately.

Pre Processing Data: pre-processing is a part of data mining, which involves transforming raw data into a more coherent format. Raw data is usually, inconsistent or incomplete and usually contains many errors. The data pre- processing involves checking out for missing

values, looking for categorical values, splitting the data-set into training and tests set and finally do a feature scaling to limit the range of variables so that they can be compared on common environs.

Training the Machine: Training the machine is similar to feeding the data to the algorithm to touch up the test data. The training sets are used to tune and fit the models. The test sets are untouched, as a model should not be judged based on unseen data. The training of the model includes cross-validation where we get a well-grounded approximate performance of the model using the training data. Tuning models are meant to specifically tune the hyperparameters like the number of trees in a random forest. We perform the entire cross-validation loop on each set of hyperparameter values. Finally, we will calculate a cross-validated score, for individual sets of hyperparameters.

Then, we select the best hyperparameters. The idea behind the training of the model is that we some initial values with the dataset and then optimize the parameters which we want to in the model. This is kept on repetition until we get the optimal values. Thus, we take the predictions from the trained model on the inputs from the test dataset. Hence, it is divided in the ratio of 80:20 where 80% is for the training set and the rest 20% for a testing set of the data.

Data Scoring: The process of applying a predictive model to a set of data is referred to as scoring the data. The technique used to process the dataset is the Random Forest Algorithm. Random forest involves an ensemble method, which is usually used, for classification and as well as regression. Based on the learning models, we achieve interesting results. The last module thus describes how the result of the model can help to predict the probability of a stock to rise and sink based on certain parameters. It also shows the vulnerabilities of a particular stock or entity. The user authentication system control is implemented to make sure that only the authorized entities are accessing the results.

5. RESULTS

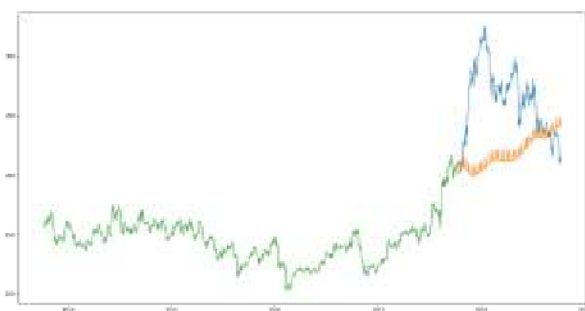


Fig2: Yearly chat of stock



Fig.3: No. of days stock variations

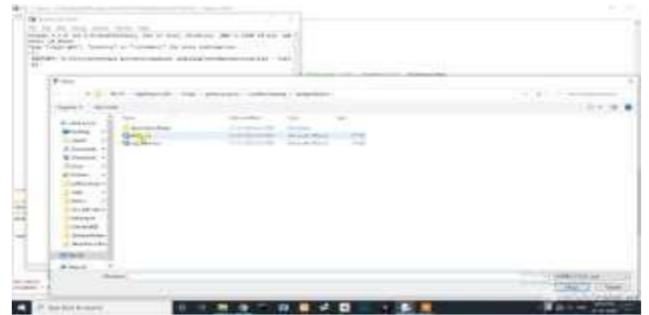


Fig.4: Selecting Stock Dataset



Fig.5: Expected prediction days



Fig.6: Stock predicting of some days



Fig.7: showing the predicting graph along with real graph .

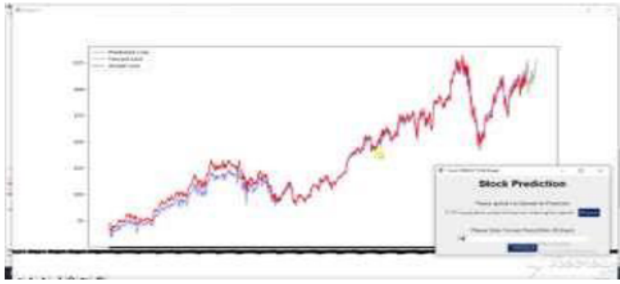


Fig.8: Showing the predicting graph along with real graph



Fig.9: Main Screen

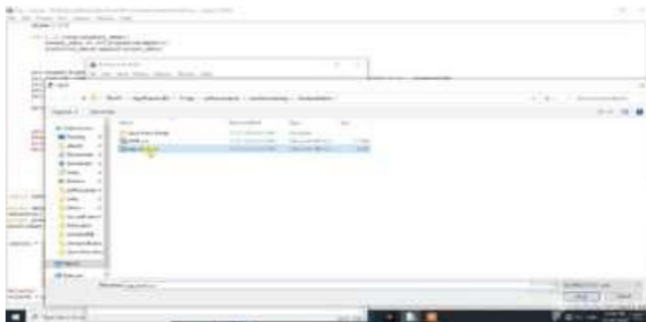


Fig.10: selecting new dataset



Fig.11: given new value for prediction



Fig.12: Showing new result for prediction

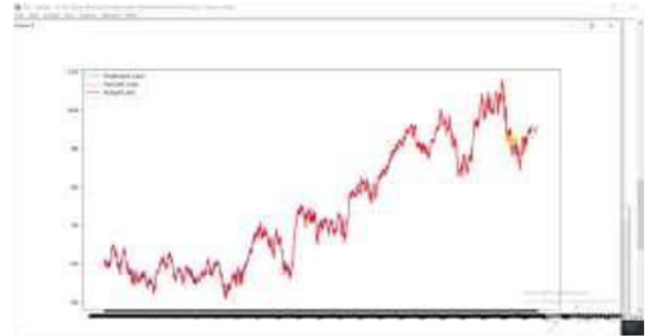


Fig.13: Showing graph for new dataset



Fig.14: Zooming to Prediction Line



Fig16: Zooming to the Result



Fig17: Zoomed to Result



Fig18: in the result prediction line matches the real time line and green line decide the future prediction.

CONCLUSION

The proposed model uses regression analysis as a ML technique with web scrapping which is followed by Fuzzy type-2 Clustering. A prediction system has been built that uses machine learning technique to produce periodically forecasts about stock market prices. The use of Prediction and regression helps us to find errors and improve accuracy of the system. It can be very useful for the investors to use this to gain maximum profit. There are many research directions which might be considered in the future work. Improving the accuracy of the predictive models is one of them. Accuracy can be improved by considering an entirely different aspect i.e., human sentiments.

As future scope of stock market is limitless, the demand for its data analysis will be ever increasing. By changing only the training data, the proposed system can be used for any stock markets of other countries. With few alterations, the system can be used for various purposes such as predicting prices of commodities like gold, predicting the fuel consumption of a vehicle as well as monitoring health of a patient.

REFERENCES

- [1]. I.A. Nayak, M. Pai and R. Pai, "Prediction Models for Indian Stock Market", *Procedia Computer Science*, vol. 89, pp. 441-449, 2016.
- [2]. W. Fenghua, X. Jihong, H. Zhifang and G. Xu, "Stock Price Prediction Based on SSA and SVM", *Procedia Computer Science*, vol. 31, pp. 625-631, 2014.
- [3]. Bini, T. Mathew, "Clustering And Regression Techniques For Stock Prediction", *Procedia Technology*, vol. 24, pp. 1248-1255, 2016.
- [4]. Elberzagher and K. Holl, "Towards Automated Capturing and Processing of User Feedback for Optimizing Mobile Apps", *Procedia Computer Science*, vol. 110C, pp. 215-221, 2017.
- [5]. A. Izzah and R. Widyastuti, "Prediksi Harga Saham Menggunakan Improved Multiple Linear Regression Untuk Pencegahan Data Outlier", *KINETIK*, vol. 2, no. 3, pp. 141-150, 2017.
- [6]. A. Khandeparkar, R. Gupta and B. Sindhya, "An Introduction to Hybrid Platform Mobile Application Development", *International Journal of Computer Applications*, vol. 118, no. 15, 2015.
- [7]. W.M. Lee, *Beginning Android Application Development* 1st Edition Kindle Edition, John Wiley and Sons Inc., 2012.
- [8]. A. Singh, S. Sharma and S. Singh, "Android Application Development using Android Studio and PHP Framework", *International Journal of Computer Applications*, vol. xx, no. xx.
- [9]. Robert K. Lai, Chin-Yuan Fan, Wei-Hsiu Huang and Pei-Chann Chang, "Evolving and clustering fuzzy decisiontree for financial Time series data forecasting", *An International Journal of Expert Systems with Applications*, Vol.36, No.2, pp. 3761-3773, March 2009.
- [10]. Shyi-Ming Chen and Yu-Chuan Chang, "Multi-Variable Fuzzy Forecasting Based On Fuzzy Clustering andFuzzy Rule Interpolation Techniques", *Information Sciences*, Vol.180, No.24, pp. 4772-4783, 2010.
- [11]. S AbdulsalamSulaimanOlaniyi, Adewole, Kayode S., Jimoh, R. G," Stock Trend Prediction Using RegressionAnalysis – A Data Mining Approach", *ARNP Journal of Systems and Software* Volume 1 No. 4, JULY 2011,Brisbane, Australia.
- [12]. D. Bhuriya, G. Kaushal, A. Sharma and U. Singh, "Stock Market Predication Using A Linear Regression," in *International Conference on Electronics, Communication and Aerospace Technology*, 2017.
- [13]. P. Guo, M. Waqar, H. Dawood , M. B. Shah Nawaz and M. A. Ghazanfar, "Prediction of Stock Market by Principle Component Analysis," in *13th International Conference on Computational Intelligence and Security*, 2017. International Conference on Recent Innovations in Electrical, Electronics & Communication Engineering - (ICRIEECE) 2477 Authorized licensed use limited to: UNIVERSITY OF BIRMINGHAM. Downloaded on May 10,2020 at 21:59:23 UTC from IEEE Xplore. Restrictions apply.