Machine Learning Models in Stock Market Prediction

Shrushti Doshi
MCA – I ,
Institute of Industrial and Computer Management and Research, (IICMR) Pune
shrushtidoshi22@gmail.com
9096465390

Shubhada Mathapati
MCA – I ,
Institute of Industrial and Computer Management and Research, (IICMR) Pune
byshubhada@gmail.com
7507571594

Shubham Varade
MCA-I,
Institute of Industrial and Computer Management and Research, (IICMR) Pune
shubhamvarade121@gmail.com
7218607991

Machine Learning Models in Stock Market Prediction

Abstract :- Stock market prediction has been a significant area of research in Machine Learning. Machine learning algorithms such as regression, classifier, and support vector machine (SVM) help predict the stock market. The currently overblown market economy has given rise to numerous variables which need to be considered before making a beneficial transaction in the stock market . Therefore, a Machine Learning approach is best suited for analysis of such a seemingly chaotic system. In this project we are using Machine learning, which give a prediction of various aspects of a particular stock or an index, such as future values of the opening price, closing price, index value etc. This will help investors and traders make better and faster decisions. This article presents a simple implementation of analysing and forecasting stock market prediction using machine learning.

Keywords: Stock market; Prediction; Machine learning; Artificial neural network

I. INTRODUCTION

A stock market is a public market where you can buy and sell shares for publicly listed companies. The stocks, also known as equities, represent ownership in the company. The stock exchange is the mediator that allows the buying and selling of shares. Stock market is the important part of economy of the country and plays a vital role in the growth of the industry and commerce of the country that eventually affects the economy of the country. Both investors and industry are involved in stock market and wants to know whether some stock will rise or fall over certain period of time. Stock Price Prediction using machine learning helps you discover the future value of company stock and other financial assets traded on an exchange. The entire idea of predicting stock prices is to gain significant profits. Application of Machine learning algorithms is generally focused on technical analysis but incorporation of the concepts of fundamental analysis into machine learning can be beneficial. This project tells about how various efforts have been taken in the application of Machine Learning to Stock forecasting and also suggests new potent ideas that can be worked upon. . Most stock traders nowadays depend on Intelligent Trading Systems which help them in predicting prices based on various situations and conditions, thereby helping them in making instantaneous investment decisions. . Most stock traders nowadays depend on Intelligent Trading Systems which help them in predicting prices based on various situations and conditions, thereby helping them in making instantaneous investment decisions.

II . LITERATURE SURVEY

In the last few decades forecasting of stock returns has become an important field of research. In most of the cases the researchers had attempted to establish a linear relationship between the input macroeconomic variables and the stock returns. After the discovery of nonlinearity in the stock market index returns, many literatures have come up in nonlinear statistical modelling of the stock returns, most of them required that the nonlinear model be specified before the estimation is done. But since stock market return is noisy, uncertain, chaotic and nonlinear in nature, ANN has evolved out to be better technique in capturing the structural relationship between a stock's performance and its determinant factors more accurately than many other statistical techniques. In literature, different sets of input variables are used to predict stock returns. In fact, different input variables are used to predict the same set of stock return data. Some researchers used input data from a single time series where others considered the inclusion of

heterogeneous market information and macro-economic variables. Some researchers even pre-processed these input data sets before feeding it to the ANN for forecasting.

III . METHODOLOGY

Creating a good stock price prediction model is particularly challenging because it is non-linear. As mentioned before, stock prices are influenced by people and not only socio-political-economical factors. Other aspects also affect the price viz. Fundamental research (qualitative analysis), based on the companies fundamentals, whose stock is the target, and Technical analysis, based on past behavior of the stock and the patterns associated with it. This makes the problem not only non-linear but highly dynamic. Stock market prediction using machine learning techniques is the right way forward.

In this section, We will be talking about the basics of how a technique works, its advantages, and its limitations.

A.) Linear Regression

One of the most straightforward techniques, Linear Regression, can be used to predict any continuous values, including the predictions of the stock price. Linear Regression, as the name suggests, is a linear technique, i.e., it finds the linear combination of the X variables that are used to predict the Y variable (the stock price in this case). The major advantage of this method is that it is high in interpretability as the user can know which factor influences the price of stock more and by how much. The disadvantage includes that it is highly limited in its scope. Many predictors cannot be used, which is required to solve the stock price prediction problem. Machine Learning-based packages such as sci-kit learn to allow the user to use Linear Regression in a Machine Learning framework. Some libraries in R also allow the same, but the disadvantage persists.

B) Support Vector Regression (SVR)

Considered once the major rival of Neural Networks, Support Vector Machine is a machine learning technique that can solve regression and classification problems. While initially designed to solve classification problems by maximizing the margin, it has presented a similar concept to solve regression-based problems only by adjusting the epsilon where the error of those data points inside the margin is not calculated. This provides the same advantage that a typical Support Vector Classifier offers: the reduction of overfitting, better generalization, and accurate results. Unlike Linear Regression, it has the capability of dealing with datasets that are in high dimensions. While we compare it with Linear Regression, one problem persists: the technique is linear, whereas the problem is often non-linear. This problem is solved by using kernels that can make this algorithm work in a non-linear manner. Among the traditional Machine Learning based techniques, this is the most advanced and accurate technique but again fails to solve highly complex and dynamic problems, which are something of a forte of deep learning-based methods.

C) K Nearest Neighbor

In theory, K Nearest Neighbor (KNN) seems to be the perfect technique. When we say that while each individual's actions are unique and possess free will when in a group, their behavior manifests patterns, thus making it predictable. This allows the data scientist the opportunity to look for similar patterns and predict the target. Therefore, KNN seems to be the ideal candidate. It is a distance-based technique that, for an observation, looks for the most similar records and then predicts the value based on the outcome of these records. Thus, in theory, for a situation in the market, KNN can look for the most similar historical situation, find how the market behaved, and predict the stock price. While the idea seems enticing and is highly simple to understand, it doesn't work as expected. The major problem with this method is that it will consider all the variables that we provide; thus, it gives equal importance to all the predictors all the time. In addition to this, it is a highly time-consuming technique and cannot be used for day trading. Lastly, the prediction is highly sensitive towards the value of K and the chosen distance metric, which can hugely alter the results, thus making the technique rigid and not as dynamic as needed.

D) Random Forest

Considered among the most potent tree-based techniques, Random Forest can predict the stock process as they can also solve regression-based problems. It uses bootstrapping and pasting techniques. It randomly picks different features and creates multiple

Decision, Tree models. This way can help predict the stock prices. Also, it has the advantage of being highly interpretable. However, even with these advantages and its advanced versions such as Gradient and XG Boost, the method is not sufficiently sophisticated to solve the highly complex problem of predicting stock prices. This leads us to the exploration of the often considered subset of Machine Learning – Deep Learning.

E) ANN

The most basic form of Deep Learning technique is the Artificial Neural Networks powered by backpropagation. ANN, for one, can solve non-linear and complex problems and adjust the weights of each of the inputs, thus making it an ideal technique to start with. As the input data for predicting the stock price can often be unstructured, for example, text-based news updates, blogs, articles, etc., ANN can also deal with unstructured data. ANN, however, is the most rudimentary deep learning technique, and other techniques generally outperform it and are discussed below.

F) Support Vector Machine (SVM)

Support Vector Machine (SVM) is used to create classifications and regression analysis by using learning algorithms. SVM is given a training set of examples marked as either of two categories. Based on these examples, SVM can further categorize new examples. These examples are represented by points and are mapped into a space to show the divide between categorizations.

G) C4.5 Decision Tree

C4.5 is an algorithm created by Ross Quinlan to construct a Decision Tree. C4.5 is built on Quinlan's earlier ID3 algorithm. C4.5 is a statistical classifier, meaning it is used for classification. Using the training data set, at each node of the tree, C4.5 decides what attribute best splits its set of examples into subsets of different classes. The attribute with the most standardized information gain is the decision-making attribute

Conclusion:

In this research, we sought to understand the set of information present in the financial market and identify the variables that drive stock prices, taking into account the activities performed and economy sectors, the various industry, macroeconomic and market indicators. The methodology presented may be adapted to other enterprises and their stocks. Theoretically, the discussion on the predictions of stock prices is still controversial and unproductive. However, from the empirical standpoint, this research proved to be very productive, with some methods for financial market prediction being developed and demonstrated.

While machine learning models can provide insights and assist in decision-making, they should not be relied upon as the sole basis for investment strategies. Incorporating other fundamental and qualitative analysis, staying informed about market news and trends, and consulting with financial professionals are crucial components for making informed investment decisions.

It is recommended to stay updated with the latest research in the field of machine learning for stock market prediction, as ongoing advancements may contribute to more accurate models and methodologies. Ultimately, a combination of diverse analytical approaches, human judgment, and prudent risk management is essential for navigating the complexities of the stock market.

REFERENCES :

[1]. K. Tsai and J. Wang, "External technology sourcing and innovation performance in LMT sectors", Research Policy, vol. 38, no. 3, pp. 518-526, 2009.

[2]. K. Han and J. Kim, "Genetic quantum algorithm and its application to combinatorial optimization problem", Evolutionary Computation, 2000, vol. 2, pp. 1354-1360, 2000.

[3]. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.

[4]. Dutta, A. Prediction of Stock Performance in the Indian Stock Market Using Logistic Regression. Intern. J. Bus. Inf. 2012, 7, 105–136.