Stock Prices Prediction Using Machine Learning and Deep Learning Techniques

Srivishnu Narayanan[1], Dr. Kayalvizhi Nagarajan[2]
[1]*M.E, Dept. of computer science and engineering, Easwari Engineering College (Affiliated to Anna University) Chennai, India*
[2]*Dept. of computer science and engineering, Easwari Engineering College (Affiliated to Anna University) Chennai, India*

Abstract **- 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. psychological, rational, and irrational behaviour, etc. All these aspects combine to make share prices volatile and very difficult to predict with a high degree of accuracy. Can we use machine learning as a game changer in this domain? Using features like the latest announcements about an organization, their quarterly revenue results, etc., machine learning techniques have the potential to unearth patterns and insights we did not see before, and these can be used to make unerringly accurate predictions.**

Index Terms **- Artificial intelligence, machine learning Data analyzing, deep learning, linear regression.**

I.INTRODUCTION

'Average' is easily one of the most common things we use in our day-to-day lives. For instance, calculating the average marks to determine overall performance, or finding the average temperature of the past few days to get an idea about today's temperature – these all are routine tasks we do on a regular basis. So, this is a good starting point to use on our dataset for making predictions.

The predicted closing price for each day will be the average of a set of previously observed values. Instead of using the simple average, we will be using the moving average technique which uses the latest set of values for each prediction. In other words, for each subsequent step, the predicted values are taken into consideration while removing the oldest observed value from the set. Here is a simple figure that will help you understand this with more clarity. We will implement this technique on our dataset. The first step is to create a data frame that contains only the Date and Close price columns, then split it into train and validation sets to verify our predictions.

A. Linear Regression

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.

The equation for linear regression can be written as:

$$Y = \theta_1 X_1 + \theta_2 X_2 + \dots \theta_n X_n$$

Here, x1, x2,….xn represent the independent variables while the coefficients θ1, θ2, …. θn represent the weights.

B. Artificial intelligence

Artificial intelligence (AI) refers to the simulation of human intelligence in machines that are programmed to think like humans and mimic their actions. Consider the behaviour of the digger wasp, Sphex ichneumoneus. When the female wasp returns to her burrow with food, she first deposits it on the threshold, checks for intruders inside her burrow, and only then, if the coast is clear, carries her food inside. The real nature of the wasp's instinctual behaviour is revealed if the food is moved a few inches away from the entrance to her burrow while she is inside: on emerging, she will repeat the whole procedure as often as the food is displaced.

Intelligence—conspicuously absent in the case of Sphex— must include the ability to adapt to new circumstances.

C. Machine learning

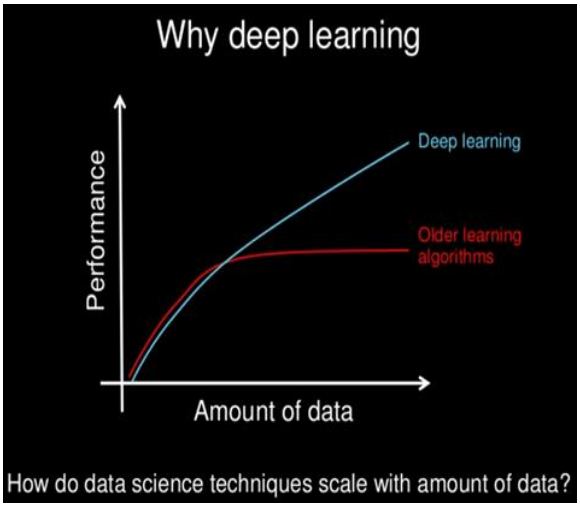
ML teaches a machine how to make inferences and decisions based on past experience. It identifies

patterns, analyses past data to infer the meaning of these data points to reach a possible conclusion without having to involve human experience. This automation to reach conclusions by evaluating data, saves a human time for businesses and helps them make a better decision.

D. Deep Learning

Deep Learning is an ML technique. It teaches a machine to process inputs through layers in order to classify, infer and predict the outcome. Deep learning is a subset of machine learning in artificial intelligence that has networks capable of learning unsupervised from data that is unstructured or unlabeled. Also known as deep neural learning or deep neural network.

- Deep learning is an AI function that mimics the workings of the human brain in processing data for use in detecting objects, recognizing speech, translating languages, and making decisions.
- Deep learning AI is able to learn without human supervision, drawing from data that is both unstructured and unlabeled.
- Deep learning, a form of machine learning, can be used to help detect fraud or money laundering, among other functions.



The core of deep learning according to Andrew is that we now have fast enough computers and enough data to actually train large neural networks. When discussing why now is the time that deep learning is taking off at Extract Conf 2015 in a talk titled "What data scientists should know about deep learning "

He also commented on the important point that it is all about scale. That as we construct larger neural networks and train them with more and more data, their performance continues to increase.

This is generally different to other machine learning techniques that reach a plateau in performance.

II. PROBLEM STATEMENT

We will 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.

Technical Analysis, on the other hand, includes reading the charts and using statistical figures to identify the trends in the stock market.

	Date	Open	High	Low	Last	Close	Total Trade Quantity	Turnover (Lacs)
0	2018-10-08	208.00	222.25	206.85	216.00	215.15	4642146.0	10062.83
1	2018-10-05	217.00	218.60	205.90	210.25	209.20	3519515.0	7407.06
2	2018-10-04	223.50	227.80	216.15	217.25	218.20	1728786.0	3815.79
3	2018-10-03	230.00	237.50	225.75	226.45	227.60	1708590.0	3960.27
4	2018-10-01	234.55	234.60	221.05	230.30	230.90	1534749.0	3486.05

There are multiple variables in the dataset – date, open, high, low, last, close, total_trade_quantity, and turnover.

The columns Open and Close represent the starting and final price at which the stock is traded on a particular day.

High, Low and Last represent the maximum, minimum, and last price of the share for the day.

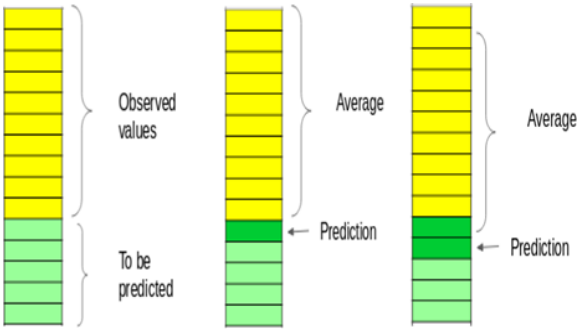
Total Trade Quantity is the number of shares bought or sold in the day and Turnover (Lacs) is the turnover of the particular company on a given date. Another important thing to note is that the market is closed on weekends and public holidays. Notice the above table again, some date values are missing – 2/10/2018, 6/10/2018, 7/10/2018. Of these dates, 2nd is a national holiday while 6th and 7th fall on a weekend.

The profit or loss calculation is usually determined by the closing price of a stock for the day; hence we will consider the closing price as the target variable. Let us plot the target variable to understand how it is shaping up in our data:

III. METHODOLOGY

Average' is easily one of the most common things we use in our day-to-day lives. For instance, calculating the average marks to determine overall performance, or finding the average temperature of the past few days to get an idea about today's temperature – these all are routine tasks we do on a regular basis. So, this is a good starting point to use on our dataset for making predictions.

The predicted closing price for each day will be the average of a set of previously observed values. Instead of using the simple average, we will be using the moving average technique which uses the latest set of values for each prediction. In other words, for each subsequent step, the predicted values are taken into consideration while removing the oldest observed value from the set. Here is a simple figure that will help you understand this with more clarity.



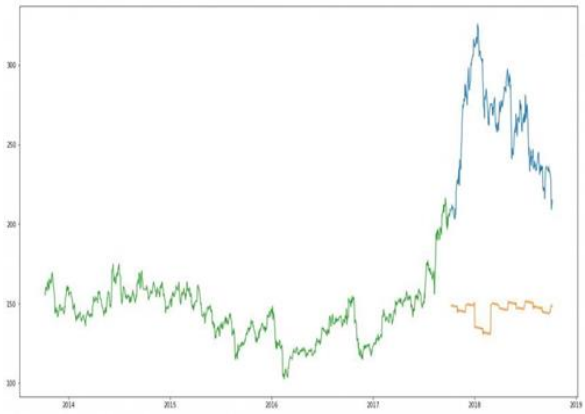
IV IMPLEMENTATION

We will first sort the dataset in ascending order and then create a separate dataset so that any new feature created does not affect the original data.

Note: I have used add_datepart from fastai library. If you do not have it installed, you can simply use the command pip install fastai. Otherwise, you can create these features using simple for loops in python. I have shown an example below.

Apart from this, we can add our own set of features that we believe would be relevant for the predictions.

For instance, my hypothesis is that the first and last days of the week could potentially affect the closing price of the stock far more than the other days. So, I have created a feature that identifies whether a given day is Monday/Friday or Tuesday/ Wednesday/ Thursday.



The RMSE value is higher than the previous technique, which clearly shows that linear regression has performed poorly.

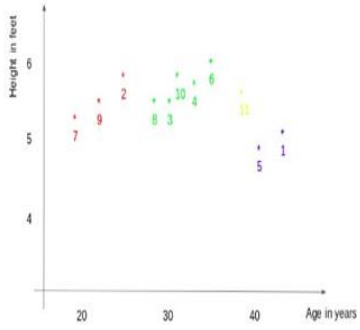
k-Nearest Neighbours

Another interesting ML algorithm that one can use here is kNN (k nearest neighbours). Based on the independent variables, kNN finds the similarity between new data points and old data points. Let me explain this with a simple example.

Consider the height and age for 11 people. On the basis of given features ('Age' and 'Height'), the table can be represented in a graphical format as shown below:

ID	Height	Age	Weight
1	5	45	77
5	4.8	40	72
6	5.8	36	60

ID	Age	Height	Weight
1	45	5	77
2	26	5.11	47
3	30	5.6	55
4	34	5.9	59
5	40	4.8	72
6	36	5.8	60
7	19	5.3	40
8	28	5.8	60
9	23	5.5	45
10	32	5.6	58
11	38	5.5	?



To determine the weight for ID #11, kNN considers the weight of the nearest neighbors of this ID. The weight of ID #11 is predicted to be the average of its neighbors. If we consider three neighbours (k=3) for now, the weight for ID#11 would be = (77+72+60)/3 =69.66 kg.

V.CONCLUSION

It has led to the conclusion that it is possible to predict stock market with more accuracy and efficiency using machine learning techniques. market indices by training on their previous values. considered are open, close, low, high and volume. Tangible value, thereby increasing the accuracy.

VI.ACKNOWLEDGMENT

I would like to thank Dr. Kayalvizhi Nagarajan ma'am for guiding me to write this paper. I would also like to show my gratitude towards Easwari Engineering College for giving an opportunity to write this paper and providing such a friendly environment for doing research and writing this paper.

REFERENCE

[1] Darmadi Komo, Chein-I Chang, Hanseok KO, "Neural Network Technology for Stock Market Index Prediction", International Symposium on Speech, Image Processing and Neural Networks, 13-16 April 1994

[2] D. Venugopal Setty, T.M.Rangaswamy and K.N.Subramanya, "A Review on Data Mining Applications to the Performance of Stock Market", International Journal of Computer Applications, (0975 – 8887) Volume 1 – No. 3, 2010

[3] Dase R.K. and Pawar D.D., "Application of Artificial Neural Network for stock market predictions: A review of literature", International Journal of Machine Intelligence, ISSN: 0975–2927, Volume 2, Issue 2, pp-14-17, 2010

[4] Akhter Mohiuddin Rather, "A prediction-based approach for stock returns using autoregressive neural networks", IEEE, 978-1-4673-0126-8, 2011.

[5] D. Ashok kumar and S. Murugan, "Performance Analysis of Indian Stock Market Index using Neural Network Time Series Model", International Conference on Pattern Recognition, Informatics and Mobile Engineering (PRIME), IEEE, 978-1-4673-5845-3, 2013.

[6] Aditya Nawani, Himanshu Gupta, Narina Thakur, "Prediction of Market Capital for Trading Firms through Data Mining Techniques", International Journal of Computer Applications (0975 – 8887) Volume 70– No.18, May 2013

[7] Chi Kin Chow, Tong Lee, "Construction of multi-layer feed forward binary neural network by a genetic algorithm Neural Networks", IJCNN Proceedings of the 2002 International Joint Conference, Honolulu, HI, USA, 2002

[8] Cao Q, Leggio KB, Schniederjans Mj., "A comparison between Fama French's model and artificial neural network s in predicting the Chinese stock market", Computers & Operations Research, Vol. 32, pp. 2499-2512, 2005.

[9] Satyajit Dhar, Tuhin Mukherjee, Arnab Kumar Ghoshal, "Performance Evaluation of Neural Network Approach in Financial Prediction: Evidence from Indian Market", Proceedings of the International Conference on Communication and Computational Intelligence, pp.597-602, 2010

[10] Textbook "Data Mining Techniques", by Michael J.A. Berry and Gordon S. Linoff, Second edition

[11] Gitansh Khirbat, Rahul Gupta, Sanjay Singh, "Optimal Neural Network Architecture for Stock Market Forecasting", International Conference on Communication Systems and Network Technologies, IEEE, 978-0-7695-4958-3, 2013

[12] Binoy B. Nair, M. Patturajan, V.P Mohandas, Sreenivasan R.R, "Predicting the BSE Sensex: Performance Comparison of Adaptive Linear Element, Feed forward and Time Delay Neural Networks", IEEE, 978-1-4673-0449-8, 2012.

[13] Kumar Abhishek, Anshul Khairwa, Tej Pratap, Surya Prakash, "A Stock Market Prediction Model using Artificial Neural Network", International Conference on Computing, Communications and Networking Technologies (ICCCNT), IEEE-20180, 26th -28th July 2012.

[14] Kunwar Singh Vaisla and Dr. Ashutosh Kumar Bhatt, "An Analysis of the Performance of Artificial Neural Network Technique for Stock Market Forecasting", (IJCSE) International Journal on Computer Science and Engineering, Vol. 02, No. 06, pp. 2104-2109, 2010

[15] J. G. Agrawal, Dr. V. S. Chourasia, Dr. A. K. Mittra, "State-of-the-Art in Stock Prediction Techniques", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol. 2, Issue 4, April 2013

[16] Pratap Kishore Padhiary and Ambika Prasad Mishra, "Development of Improved Artificial

Neural Network Model for Stock Market Prediction", International Journal of Engineering Science and Technology (IJEST), Vol. 3 No. 2, Feb 2011

[17] Yusuf Perwej and Asif Perwej, "Prediction of the Bombay Stock Exchange (BSE) Market Returns Using Artificial Neural Network and Genetic Algorithm", Journal of Intelligent Learning Systems and Applications, 4, 108-119, 2012.

[18] www.nseindia.com