Forex exchange using big data analytics

Jayakumar Sadhasivam\textsuperscript{1*}, M. Arun\textsuperscript{2}, R Deepa\textsuperscript{3}, V Muthukumaran\textsuperscript{4}, R Lokesh Kumar\textsuperscript{4} and R B Prasanna Kumar \textsuperscript{6}

\textsuperscript{1}School of Information Technology and Engineering (SITE), Vellore Institute of Technology, Vellore, India.
\textsuperscript{2}School of Computing, Kalasalingam Academy of Research and Education, Tamilnadu – 626128, India.
\textsuperscript{3}Chaitanya Bharathi Institute of Technology (A), Hyderabad, Telangana, India.
\textsuperscript{4}Department of Mathematics, School of Applied Science, REVA University, Bengaluru – 64.
\textsuperscript{5}Department of Analytics, School of Computer Science and Engineering, Vellore Institute of Technology, Vellore – 632 014.
\textsuperscript{6}Department of Computer Science, REVA University, Bengaluru – 64.
E-mail: \textsuperscript{*}jayakumarsvit@gmail.com

Abstract. Analysis and Prediction of forex has gained immense value in today’s economy. The stock price prediction is a difficult process owing to the irregularities in stock prices. Every trader wants to know if the pattern has been repeated in past to know what the possible output of the current situation will be. The primary objective is to propose a methodology that will use a historical dataset and provide a more accurate prediction on stock price. In this paper, we will be using machine learning pattern recognition algorithm on forex tick dataset. The learned model then will produce pattern from the given dataset and on the pattern of increasing or decreasing, the buyer will initiate a buy or sell the stock respectively. We will use python coding to execute the algorithm in jupyter notebook. Matplot library will help us to perform graphing in the process and Numpy will be helpful in doing statistical analysis of data.

1. Introduction
Every trader wants to find out the pattern of the forex before he makes any decision of making small or big investment in that forex. These changes in forex market reflect directly to the economy of the area. There is abundance of algorithms that could be found on the Internet that allows the user to predict the next change in forex market. Most of these are just a hoax and a way to manipulate people. We are going to find out patterns by plotting together the lines of those patterns on graph, which are very much similar to on another. Then we will perform back test on these results. A forward test can be performed on the upcoming data that has been produced after the prediction of the data but that data cannot be back tested to give the guarantee that it is indeed a suitable prediction.

We are using python here because it is a single threaded language, it uses single core of CPU, which means one script will use only one CPU (see ref. [8]-[14]).

We have many variables to be accounted for, not just the explicit ones but implicit ones too. Some of the explicit variables are percent change as pattern recognition, start point to current location percent change, fixed pattern length, fixed value/weight of pattern irrespective how old or new the data is.
For the implicit variables, we have two factors to be taken under consideration here. These are Opportunity and Accuracy. The way we find out the best case in machine learning is much like how companies decide the best suitable price for their products. The margin that we keep for the profit, that matters a lot but quantity sold is important too. A cheap product could sell many units but then we will have less profit overall.

Here we must also take into account the weigh opportunity, which means the amount of unit sold, comparing this to Accuracy, which is like the profit margin. So, we might have 60% accuracy in big volume opportunity, compared to 80% in less volume.

2. Related work
In [1], they have tested some trends that have given a contribution to central bank intervention. Firstly, using data that has high frequency and secondly, the study of event technique. This technique implies that it is better if we study clusters, sporadic intervention steps to carry out goal. And data that has more frequency can be seen to ameliorate the proportional daily routine tasks.

In [2], they used the basic principles of artificial neural networks to accurately predict the future rate of a particular forex. This is done in different steps; first, they collect the past data of a stock. After this, they create a neural network & train the network on the basis of the past data. Generally, 80-90% of the data is used to train the model. Then they use a different model to try to predict the value of the stock & compare the value with the previously obtained data set.

In [3], they have studied a wide range of method using sentimental analysis for forex market prediction. As the need of user rose, opinion mining became important to research areas. So, when the size of this data started increasing, the analysis and demand also seem to increase. To satisfy this need, more and different advanced techniques were introduced.

In [4], the results from this study imply that different kinds of techniques, work differently and all are capable enough to predict the forex market. They used Karachi Stock market to develop a technique for their study. The Multi-Layer Perceptron algorithm predicted 77 percent accuracy in market exchanges.

In [5], this paper summarizes important key features and techniques in machine learning, which are related to stock prediction. The paper recommends the use of linear regression and logistic regression for stock prediction and analysis. And this study suggests SVM to obtain accurate results. A constraint to this result is the necessity of the data used in prediction to be classification friendly.

In [6], data analysis shows that people trust their investment counselors and depend on their guidance to invest in market. This result shows that this trust is an essential factor that is much likely to affect the client’s transaction intention in the Foreign exchange. The connection between purchase intention and trust is strong, implying on this analysis, which was done on the users responses.

In [7], forex market counselors should be involved in guidance only after through study of the whole process. Based on the data that has been collected over past 3 months has led us to conclude that the Indian market is very susceptible to change and also very sensitive when it is compared to other countries.

3. Proposed method
In this paper we are using pattern recognition algorithm on forex tick dataset. This algorithm is used when a group of price in a time frame is taken and then converting them to percent change in an effort to normalize the data. This pattern is then map into the memory, and then move forward one price point and remap the pattern, for each pattern mapping into the memory a log is maintained to where the price is that point. The percent similarity of the current pattern is compared with the entire
previous patterns, and if the result has exceeded the certain threshold level, then the percent similarity gets considered. An estimated average outcome is generated from the aggregation of all their previous outcomes and if the result of the estimated average outcome is favourable then a customer might initiate a buy or else sells the shares.

In the function graphRawFX(), the dataset which we are currently using is plotted in the graph format. In the function percentchange(), a machine learned batch is created which has million of patterns and their results. A range of at is taken into consideration and a pattern is created in which the percent change will be performed. In this function percentchange(), at the starting forward percent change is performed. It means as long as the pattern goes, there are chances that actual end may be similar.

In the function currentpattern(), we are finding the current pattern from the tick dataset. The short term outcomes are also collected from this function, and later in the algorithm with the help of current pattern and stored pattern, a comparison will be done in order to find the similarity in the pattern.

In the function storingpattern(), the percent change values of patterns are being collected from the tick data. To compare the patterns, a percent calculation to calculate the similarity between each percent change is done. The pattern which has exceeded the certain threshold level is considered to be similar are getting stored, and average is calculated of the similar patterns and those values are also getting stored.

In the function patternrecognition(), each pattern in pattern array that is there we have, we will compare the similarity that with current pattern. The way this is done is to find the percent change, we will find the percent change for each pattern that was found instorepattern() and will compare it with the current pattern that was found in currentpattern().The percent change in order to find the similarity. After finding the similarity of all the patterns, an average similarity is calculated. The average similarity is then compared with certain threshold level, so as the average whose value is greater; a graph is plotted for those patterns.

A patterns are then plotted along with their possible outcomes in the same graph, then a visually prediction data can be generated which with we can see where the predicted pattern s going. The outcome patterns are color-coded with which it is easy to identify in the graph. The green color is used in the pattern outcome when it is above the threshold level value and red color signifies that the generated pattern is below the level. The pattern in cyan colored line is the current pattern and all the other pattern line, which is color-coded, are the similar patterns from the past.

4. Results
The dataset which we have used for prediction is represented in graphical format.
The average similarity is then compared with certain threshold level, so as the average whose value is greater; a graph is plotted for those patterns. The outcome patterns are color-coded with which it is easy to identify in the graph. The green color is used in the pattern outcome when it is above the threshold level value and red color signifies that the generated pattern is below the level. The pattern in cyan colored line is the current pattern.
The length of dataset on which the prediction is done is 62012. After the pattern is generated, back testing is performed in order to check the accuracy of the pattern. The given screenshot is taken after one actionable trade, the time the pattern took is 2.16 s, results after back testing is printed, entire time taken for the processing is 6.73 s, and the accuracy of the back testing is 100%.

5. Conclusion
In this analysis we studied how the patterns are recognized and analysis is done on them to make the future predictions about the forex. When function pattern Recognition() is completely executed we can see the visual representation of our predictions and the outcomes to those predictions. Thin lines are the comparable patterns that we found. Cyan line is the current pattern. Other lines are similar patterns from the past. Predicted outcome are the color-coded to reflect a positive or negative prediction. The Cyan dot marks where the pattern went. Only data in the past is used to generate patterns and predictions. Red and green dots represent each patterns outcome. Rise is the green dot, and if it’s a fall then a red dot. It’s not the degree at which it would rise or fall. Then the average rise or fall is calculated and plotted against the graph with dark blue dot. Then the Cyan plot shows the actual outcome of the plot. So, if the prediction is falling then consider selling, or else if it continues to rise then consider holding on to it.

References
[1]. Christopher J Neely 2005 Federal Reserve Bank of St. Louis Review 87(6) pp 685-717.
[2]. Vivek Rajput and Sarika Bobde 2016 International Journal of Computer Science and Mobile Computing pp 500-506.
[3]. Nirbhey Singh Pahwa, Neeha Khalfay, Vidhi Soni and Deepali Vora 2017 International Journal of Computer Applications.
[4]. Muthukumaran V and Ezhilmaran D 2020 International Journal of Information Technology and Web Engineering. 15(4) pp 18-36.
[5]. Ganesh Gopal Deverajan, Muthukumaran V, Ching-Hsien Hsu, Marimuthu Karuppiah, Yeh-Ching Chung and Ying-Huei Chen 2021 Transactions on Emerging Telecommunications Technologies.
[6]. Manikandan G, Perumal R and Muthukumaran V 2020 AIP Conference Proceedings 2277.
[7]. Ojaswa Swarnkar J S, Shubham Agarwal, Sanyam Jain and Nallakaruppan M K 2014 Int. J. Appl. Eng. Res. pp 2019–25.
[8]. Asha J M, Sadhasivam J and Rohit R 2017 Int. J. Mech. Eng. Technol. 8(11) pp 964–76.
[9]. Sadhasivam J, Jayavel S, Jeyakumar B and Merchant S 2017 Int. J. Civ. Eng. Technol. 8(11) pp 35–41.
[10]. Jayakumar S, Jayavel S and Senthilkumar M 2011 *Int. Conf. Netw. Commun. Comput.* pp 419–22.

[11]. Jayakumar S, Jayavel S and N S 2012 *Int. J. Comput. Sci.* pp 468–72.

[12]. Sadhasivam J, Asha Jerlin M and N. S 2016 *Int. J. Trend Res. Dev.* 3(6) 112.

[13]. Sadhasivam J, Charanya R, Harish Kumar S and Srinivasan A 2017 *IOP Conf. Ser. Mater. Sci. Eng.* 042033.

[14]. Sadhasivam J, Alamelu M, Radhika R, Ramya S, Dharani K and Jayavel S 2017 *IOP Conf. Ser. Mater. Sci. Eng.* 042032.