LITERATURE REVIEW:
THE ROLE OF MACHINE LEARNING TECHNIQUES IN FINANCIAL MARKETS

1 INTRODUCTION

The financial market is a complicated, volatile, dependent on numerous factors which would have not even be foreseen. The problem of stock market forecasting is always considered a Sisyphean task and with good reason. With information about financial markets, political events, economic events through the country and the world, influence the stock market prices. Predicting behaviour of these distinct areas which correlate to the financial market is a humongous and needs a competent enough machine learning algorithm that will be able to include all these parts of data.

To be able to create an approach for solving this, we should first understand some terminology of the data we have:

Stock Market Index:

A stock market index represents the movement average of many individual stocks whereas an index reflects mainly market movement rather than movement of a stock [1].

2 DESCRIPTION ABOUT REFERENCES

Atsalkis and Valavanis [1] in their research discuss various methods for evaluating stock market behavior and suggest that soft computing methods prove efficient for this purpose. Huang, Nakamori and Wang [2] look at Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Elman Backpropagation Neural Networks (BPNN) and Support Vector Machines (SVM) to understand and forecast the NIKKIEI 255 index in 2005. In 2006, Gavrischchaka and Banerjee also look at SVM as proposed method for volatility forecasting by presenting results for SP500 index [3]. Xie, Yu, Xu and Shang also in 2006 look at crude oil forecasting using SVM [4]. Another approach proposed by Pai and Lin is a hybrid ARIMA and SVM approach for stock price forecasting [5].

3 ANALYSIS

3.1 SVM

Huang, Nakamori and Wang strengthen with detailed reasoning why SVM is one of the good methods to be considered while discussing algorithms for forecasting stock markets versus methods like LDA, QDA and even algorithms that use artificial neural networks for deep learning of the data. The authors highlight that SVM is resistant to overfitting. They also bring up the use of radial kernel function in the SVM model by providing in-depth working of the algorithm and provide visible results by elaborately explaining that heteroscedastic models are more appropriate here. A unique approach proposed here is combining classification techniques with SVM by giving different weights based on individual performances. And the performance based on hit ratio does improve relative to SVM alone. [2]

A similar category of time series forecasting is crude oil price prediction which is taken up by Xie, Yu, Xu and Wang. The authors provide a similar outlook on the working of SVM that moves from classification to regression for time series prediction and deduce with its generalization ability. The paper also discusses the depth of the model functioning; and uses a good flow chart to show working of the model. It is also highlighted that a radial kernel is used which gives excellent performance when no additional knowledge is available. Along with detailed working of the algorithm, its performance is evaluated thoroughly since SVM provides a global solution. The authors give us great insights to conclude that in such a problem of time series forecasting a change in trend is more important than precision level. [4]

Another paper supporting the use of SVM is published by Banerjee and Gavrishchaka in 2006 where the aim is to be able to predict the stock market volatility on the SP500 index. It is depicted with a combined study of other papers on how volatility relates returns. The authors provide a comprehensive detail on how existing volatility models based on autoregressive conditional heteroskedastic (ARCH) type models are limited since they are unable to capture heterogeneity of traders acting at different tine horizons. Additionally, it is accentuated that SVM combines effectiveness of linear machines with regression power of nonlinear algorithms and how SVM overcomes the curse of dimensionality unlike traditional neural networks.

3.2 Hybrid Model – ARIMA + SVM

In 2004, Pai and Lin propose a novel technique which combines the auto regressive and classification strength from ARIMA and SVM. The authors point out an eligible approach to solving the forecasting problem by using ARIMA's ability to capture linear relations and SVM using the neural network model to capture non-linear details in the data. Thus, the hybrid model proposed by the authors has a potential however it needs to be explored further since the results shown in the paper do not display that it sufficiently performs better. Therefore, more research needs to be done on this area.

4 CONCLUSION

After exploring multiple state-of-the-art techniques for being able to understand how the stock market changes with the prices and to be able to predict direction of the trend, it can be comprehended that though there is no defined formula to help us with the solution. However, as we explore we can get as close to an optimum method as possible .

5 REFERENCES

[1] George S. Atsalakis and Kimon P. Valavanis, "Surveying stock market forecasting techniques – Part 2: Soft computing methods", Expert Systems with Applications 36, pp. 5932-5941, 2009

[2] Wei Huang, Yoshiteru Nakamori and Shou-Yang Wang, "Forecasting stock market movement direction with support vector machine", Computers & Operations Research 32, pp. 2513-2522, 2005

[3] Valeriy V. Gavrischchaka and Supriya Banerjee," Support vector machine as an efficient framework for stock market volatility forecasting", CMS 3, pp. 147-160, 2006

[4] Wen Xie, Lean Yu, Shanying Xu and Shouyang Wang, "A new method for crude oil price forecasting based on support vector machines", ICCS 2006 Part IV, pp. 444-451, 2006

[5] Ping-Feng Pai and Chih-Sheng Lin, "A hybrid ARIMA and support vector machines model in stock market forecasting", Omega 33, pp. 497-505, 2005