The authors thank two anonymous reviewers for very helpful suggestions. The paper has certainly benefited from appropriate inclusion and amendments following these comments.

**General Response**

Before we begin addressing the specific points and nuances raised by the reviewers, we think that it is necessary that we reiterate and reemphasize the aim of our work and what the outcome or the takeaway is.

The aim of the current work is to predict the direction that the price of a stock of a certain company will follow. This is effectively a gain/loss scenario where the objective is to identify the direction of change and not the magnitude. This is a deviation from the popular idea of predicting the monetary returns of a stock after a certain time period. We have tried to address this issue by using methods in machine learning, because the available data on this matter is overwhelming. Admittedly, we have used the word 'returns' in an ambiguous sense in the manuscript and have clarified it in the revised submission. However, with this point clarified, we would first like to address the following comment:

**Specific Comments and Responses**

**Comment**: "The proportion of positive and negative data is in range of 45:55”. Here you mean trading days with positive returns? Averaged across all your stocks (in all markets)? This is potentially misleading. You should include the median returns over various n-day periods in your (testing) dataset. If stock prices are stochastic but have positive drift, wouldn’t sign predictability get easier with increasing forecast horizon?"

**Response**: We have rephrased the sentence, in case of any confusion. However, the purpose here is to generalizeand develop a method to predict with reasonable efficiency whether the price of a stock will increase or decrease on day n+t as compared to day n. It seems that the mean accuracy is a more representative measure of the efficacy of the system as compared to the median returns. Certainly, in case of a positive drift in the prices of a stock it may be possible to have a better sign predictability, but that does not seem to have a conflict with our method of prediction. We have included this statement in Footnote ……..

In addition, it appears from the data that most stock prices do not exhibit a strictly negative or a strictly positive drift. Therefore, the purpose of selecting stocks of companies like Nike, Toyota, Facebook, Amazon, Apple etc. is to demonstrate that regardless of the background or domain of a company, regardless of specific fluctuations in the prices over time, the efficacy of our methods hold and these do not succumb to diminishing accuracies.

- **Reviewer 1**

**Summary**

This paper investigates the ability of two machine learning (ML) techniques (Random Forests and Gradient Boosted Trees) to forecast future stock prices. Specifically, forecasting is reformulated as a classification problem; essentially is the stock price expected to go up or down over a given time period. Model inputs are based on signals drawn from technical analysis, to which smoothing is applied by the authors. The ML techniques are presented in some detail with examples. The data set comprises 14 (mostly US) stocks with forecasts made over 3-90 days. Results are presented in terms of statistics such as accuracy, recall, etc. The ML models are held to outperform algorithms used in existing literature.

**Main Comments**

Exploring how the wide range of machine learning techniques can be applied in a financial/economic domain, and how they can be used to complement traditional models, is an interesting area.

The abstract should be rewritten to highlight novelty and the contributions of the paper.

**Response:** *We have now re-written the abstract highlighting the novel points.*

While the discussion of the EM hypothesis is well placed, anomalies could also be discussed. See suggested references below.

**Response:** *We have included the anomalies as obtained from the suggested references as part of the introduction. Thank you again*.

Malkiel (2003) offers a masterly discussion of the critique of the EMH, and suggests that, way back in 1973, he advised to buy a broad-based index fund that bought and held all the stocks in the market and that charged very low expenses. He admits that by the start of the twenty-first century, the intellectual dominance of the efficient market hypothesis had become far less universal. Many financial economists and statisticians began to believe that stock prices are at least partially predictable. Furthermore, a new breed of economists emphasized psychological and behavioral elements of stock-price determination, and came to believe that future stock prices are somewhat predictable on the basis of past stock price patterns as well as certain “fundamental” valuation metrics. We have discussed this issue further while explaining how variables are chosen in atypical models of stock market prediction. Indeed, many economists made controversial claim that these predictable patterns enable investors to earn excess risk-adjusted rates of return. From Shiller's (2000) behavioral 'bandwagon effect' to 'head and shoulders' and 'double bottoms' formations in stock prices, modest predictive power seems to exist (see, Lo, Mamaysky and Wang, 2000). Malkiel (2000) states that while the stock market may not be a mathematically perfect random walk, it is important to distinguish statistical significance from economic significance. The statistical dependencies giving rise to momentum are extremely small and are not likely to permit investors to realize excess returns. The anomalies with stock returns were often model and context specific and do not qualify as generalization.

Ref.

Lo, Andrew W., Harry Mamaysky and Jiang Wang (2000), “Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation,” Journal of Finance, 55, 1705-1765.

Malkiel, Burton G. (2003), The Efficient Market Hypothesis and Its Critics, CEPS Working Paper No. 91, April, Princeton University.

Shiller, Robert (2000), *Irrational Exuberance*, Princeton: Princeton University Press.

For me, too much of the paper is devoted to outlining the ML techniques. These are well established and fully documented elsewhere. Readers unfamiliar with the techniques should be directed to primary sources, or authoritative texts. Provide a brief description, giving an intuitive sense of how the ML techniques work. Focus on the benefits such models promise in relation to traditional approaches and to the problem at hand.

**Response:** *We have followed this suggestion carefully. The main sources where the techniques are explored and emphasized are cited in the main text along with a short introduction to the analytical and intuitive aspects of the models developed.*

While perhaps common in ‘Information System’ journals, my recommendation is to remove the outline algorithms and examples (e.g. Tables 2, 3, 4, etc). While one figure may be useful to aid understanding of a tree, the extended example and corresponding figures should be removed (or at least relegated to a separate appendix).

To appear in the finance/economics journal, the paper should be targeted to such an audience. Figure 2 is too basic to merit inclusion. Similarly figure 1 adds little value and could be easily summarised in a few lines.

Derivation of established formula/results (e.g. Chebyshev’s Inequality) and basic definitions (e.g. an indicator variable) need not be included and certainly not proved.

Similarly, it is unnecessary to provide formula or examples for well-established technical indicators [incidentally, why include day 1 in the RSI formula; 15 prices are required for 14 daily returns]. Rather, focus on why these indicators are considered to have predictive ability. See also comments below.

**Response:** *We have shortened the overall presentation of the theorems to aid better readability for economics/finance journals. The appropriate references are cited and we have deleted the discussion and proof of theorems, as suggested.*

The paper does not adequately discuss feature selection or reference supporting literature to justify the choice of input attributes believed to have predictive ability. What do the models under investigation tell us about variable importance? Do certain technical indicators prove more useful for prediction? If these models really are outperforming existing benchmarks, from a finance perspective we want to understand why.

**Response:** In recent years, the stock market analysis and prediction have been studied with the aid of methods such as machine learning and text mining. Data mining studies use daily stock data. For example, prediction studies based on support vector machines (SVMs) (Cao and Tay, 2001; Ince and Trafalis, 2007) have been conducted to determine pattern categories. In addition, artificial neural networks (ANNs) (Kimoto et al., 1990; Kohara et al., 1997) have been employed to achieve good predictions even in the case of complex relationships of variables. Typically, autoregressive integrated moving average (ARIMA) model (Pai and Lin, 2005; Wang and Leu, 1996) are used for identifying and predicting time series variation. Notwithstanding, since behavior and individualized responses play a significant role in dictating the stock turnovers and prices,, a few studuies have engaged with word analysis of news articles (Mittermayer, 2004; Nikfarjam et al., 2010; Kim et al., 2014) and its predictive ability. However, most of these studies have some limitations for short-term prediction. First, without filtering for outliers, the predictions based on all historial data leads to potential errors. Second, although the total completion price is determined by a variety of factors such as the foreign purchase closing price and domestic selling completion amount, this set needs to be expanded in order to reduce omitted variables bias. Variables of importance may include, categories of financial ratios, macro, labour market and housing variables and measures of sentiment and leverage (Black et al. 2014; Cochrane, 2008, etc). With respect to the current paper, it is important to note that the main purpose is to implement two distinct methods on stock data and highlight their advantage over other non-ensemble techniques within the machine learning approaches for analyzing and predicting stock prices. This does not warrant conducting a regression analysis. Therefore, our engagement with feature extraction and assigning of importance to respective variables will follow available wisdom, except that the outcomes will be more efficient due to the choice of models.

(Snehanshu, ADD THIS TO **RELATED LITERATURE** SECTION)

Ref:

Black, AJ, Klinkowska, O, McMillan, DG and McMillan, FJ. (2014), ‘Predicting stock returns: Do commodities prices help?’, Journal of Forecasting, 33, 627-639.

Cao, L. and Tay, F. E. (2001). Financial forecasting using support vector machines. Neural Computing & Applications, 10(2):184–192.

Cochrane, J. (2008), ‘The dog that did not bark: A defense of return predictability. Review of Financial Studies, 21, 1533-1575.

Ince, H. and Trafalis, T. B. (2007). Kernel principal component analysis and support vector machines for stock price prediction. IIE Transactions, 39(6):629–637.

Kim, Y., Jeong, S. R., and Ghani, I. (2014). Text opinion mining to analyze news for stock market prediction. Int. J. Advance. Soft Comput. Appl, 6(1).

Kimoto, T., Asakawa, K., Yoda, M., and Takeoka, M. (1990). Stock market prediction system with modular neural networks. In Neural Networks, 1990., 1990 IJCNN International Joint Conference on, pages 1–6. IEEE.

Kohara, K., Ishikawa, T., Fukuhara, Y., and Nakamura, Y. (1997). Stock price prediction using prior knowledge and neural networks. Intelligent systems in accounting, finance and management, 6(1):11–22.

Pai, P.-F. and Lin, C.-S. (2005). A hybrid arima and support vector machines model in stock price forecasting. Omega, 33(6):497–505.

Wang, J.-H. and Leu, J.-Y. (1996). Stock market trend prediction using arima-based neural networks. In Neural Networks, 1996., IEEE International Conference on, volume 4, pages 2160–2165. IEEE.

Mittermayer, M.-A. (2004). Forecasting intraday stock price trends with text mining techniques. In System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on, pages 10–pp. IEEE.

Nikfarjam, A., Emadzadeh, E., and Muthaiyah, S. (2010). Text mining approaches for stock market prediction. In Computer and Automation Engineering (ICCAE), 2010 The 2nd International Conference on, volume 4, pages 256–260. IEEE.

A complete description of the data should be presented before the results. You indicate that the start date is when the company went public. Does this mean different companies are examined for different periods? The reasons for selecting the 14 companies, or why they span different counties are not clear. Was this in part to aid comparison with the results of existing papers?

Similarly, you need to clarify:

·         The time period(s) over which training and test sets were chosen (and why).

·         If a range of different training/testing configurations were trialled.

·     The number of attributes used (preferably with summary statistics). For example, was PROC calculated over different periods? PCA is referenced earlier, was it used? Were all features smoothed in the same manner? Using what alpha value?

·         The ML parameters (e.g. number of trees, number of variables used for each split)

**Response:** Explained clearly, thank you again.

The sources on randomness in a random forest (along with their importance) should be clarified. In random forests, a random subset of features is selected for each branching decision.

**Response:** Explained.

Results should be presented against a suitable baseline drawn from existing literature. A model that always predicts an upward trend may prove instructive. Can you provide a statistical justification that these models are superior? Are the results consistent if repeated over different time periods?

**Response:** The paper discusses the merit of using Random Forests and XG Boost systems as compared to other techniques used under non-ensemble procedures. Indeed, it has been shown in several studies that trained SVM applied on Korean stock market (Kim, 2003), or DT plus ANN applied on Taiwan market (Tsai and Wong, 2009) reach 56% and 67% accuracy overall. Compared to this the present techniques attain 78% accuracy. In addition, the reason for choosing Random Forests over Decision Trees is because Random Forests uses a significant amount of voting-based conclusions as compared to that of Decision Trees. It runs a bagging based routine by using a large number of de-correlated Decision Trees (consider growing forests in random fashion) to classify a predicted class. This course of operation is highly suitable for the stock data and associated classification, as it meticulously examines the feature space to make better judgments over which mass class to finalize as the expected outcome.

Could other performance statistics be included? For example AUC or Brier scores (you mention ROC curves in the appendix but do not use them).

**Response**: Included in the current version.

Section 6.3 felt out of place and left me confused. Why consider pharmaceutical stocks? How is your study of sign predictability related to “why certain stocks did not succumb to the aggravated economic crisis”?

**Response**: Snehanshu, I don't know this.

“Comparison of the accuracy achieved in this work to the accuracies achieved in available literature.” On what basis has figure 15 been compiled? The caption is confusing. Were the 3 additional models run using identical data and features? If these were to be used as comparator models this should be made clear in the earlier methodology.

**Response**: Snehanshu, I don't know this.

Over what dates is figure 16 plotted? Does the large stock price drop represent a stock split (and if so why haven’t you accounted for this)? What constitutes a ‘buy’ signal or a ‘sell’ signal? Are there particular RF voting thresholds that must be breached to trigger a new signal? If so, why are there consecutive buys (and consecutive sells)?

Perhaps if this was clarified you could attempt to assess the economic impact of a trading strategy based on the ML models.

**Response**: Snehanshu, I don't know this.

**Minor Comments**

·         Rephrase “Market risk, strongly correlated with forecasting errors, needs to be minimized to ensure minimal risk in investment”. I assume the point you want to make here is that minimising forecast error would minimise risk.

·         “measuring the change in price of a stock compared to its price t days back” seems like a clumsy way of saying the “t-day return”.

·         Wouldn’t investors be more interested in establishing which assets might provide positive excess returns?

·         “Wisdom of Crown” should be “Wisdom of Crowds”?

·         “some stocks tend to develop linear trends in the long-run”. Do you mean that they exhibit momentum or directional trends?

·         “Tree based learning methods are non-metric”. Perhaps clarify in terms of discrete/continuous versus nominal/ordinal.

·         “Boosting a classifier means combining the results of many weak predictors to make a strong prediction”. Isn’t it more that this? A RF does that.

·         Why not present all results as percentages?

·         Figure 14 has many series (mostly blue) but only 4 legend items.

·         “The proportion of positive and negative data are in range of 45:55”. Here you mean trading days with positive returns? Averaged across all your stocks (in all markets)? This is potentially misleading. You should include the median returns over various n-day periods in your (testing) dataset. If stock prices are stochastic but have positive drift, wouldn’t sign predictability get easier with increasing forecast horizon?

·         Fix Lane (1984) bibliography entry.

·         Review paper to ensure abbreviations are consistently introduced and used.

·         Take care to clarify/standardise use of terms such as ‘accuracy rate’, ‘hit rate’ and ‘success rate’.

**Response:** Modified according to specific suggestions.

**Suggested references**

Atsalakis, G. S., & Valavanis, K. P. (2009). Surveying stock market forecasting techniques–Part II: Soft computing methods. *Expert Systems with Applications*, *36*(3), 5932-5941.

Christoffersen, P. F., & Diebold, F. X. (2006). Financial asset returns, direction-of-change forecasting, and volatility dynamics. *Management Science*, *52*(8), 1273-1287.

Moskowitz, T. J., Ooi, Y. H., & Pedersen, L. H. (2012). Time series momentum. *Journal of financial economics*, *104*(2), 228-250.

Nyberg, H. (2011). Forecasting the direction of the US stock market with dynamic binary probit models. *International Journal of Forecasting*, *27*(2), 561-578.

- **Reviewer 2**

**Title :    From forecasting to classification : Predicting the Direction of Stock Market Price Using   Tree-Based Classifier.**

In this paper the authors studied the prediction of stock market direction using two ensemble learning  algorithms:  random forest and XGboost. After preprocessing the data six technical indicators are calculated and the two algorithms are implemented using scikit learn and other software packages taking the technical indicators as input features. Discussion of results are made  in the case of data taken from 10  specific companies  and it is concluded that these two methods outperform  logistic regression, neural network ,and SVM.

*The following are some of my observations and suggestions:*

1.The paper is too long. It can be shortened by omitting sections 4.1-4.4,,5.1-5.3 ( which contain detail description of decision trees, random forest, XGboost, along with a worked out example, some derivations,etc.) as  these are standard material available in data mining /machine learning  text books. Section 4.5 and 4.6 about chebychev inequality and its proof can also be omitted. However , a  very brief description of random forest and XGboost with the algorithms may be given in the appendix. Again  the appendix A with key definitions also are to be  omitted. Also the presentation in sections 1,2,and 3 may be shortened.

Response: We have shortened the paper and deleted known proofs.

2.The authors claim that  random forest and XGboost outperform  SVM, ANN, logistic regression based on the values of performance metrics for RF and XGboost. But the corresponding values for SVM and other methods have not been computed for the data set used by the authors. To justify this claim a table containing the values of performance metrics( such as accuracy ,etc)for the three methods SVM, RF, XGboost for the **same data set** be prepared and the values be compared. This  may  further be illustrated using ROC curves for the three methods for the same data set.

Response: This is carried out in the main text. Please see section.....

3. Authors  mention  that return on the  stock can be predicted using the model. But  the model predicts only upward and downward movement. How can one determine the return on stock ? Similarly with out a quantitative predicted value of stock how can one make   efficient portfolio management?

Response: In the introduction and related analysis, we have claimed that the two models used can suitably (with considerable accuracy compared to trained SVM and other models) predict the direction the stock prices can take. In terms of investment plans, the direction can be useful for making plans on purchase or sell.

4.Stock price prediction is a more general problem then  stock direction prediction  as the former gives more information than later. So in the title   from forecasting to classification does not seem reasonable. This part may be omitted.

Response: We are considering it.

5. The size of the data set should  be mentioned in sec.7. Deep learning method  may be explored only if  available data size is big.

Response: We have mentioned this in detail. Thank you for the suggestions!