

Review of “A Random Forest Approach to Predicting Clean Energy Stock Prices”

Don Krapohl

Background and Motivation

Forecasting stock prices is of intense interest to investors and companies. Random Forest (RF) algorithms have shown great promise in predicting direction and price of stocks and are easier to train and interpret than more complex and more widely used models such as Support Vector Machines (SVM) or Artificial Neural Networks (ANNs). The authors undertake to provide directional stock forecasts using several RF model algorithms and well-known technical indicators to measure and compare the accuracy and predictive power of each model algorithm. The intent is to apply these analytical tools to predict directional movement of clean energy companies, which the authors assert is a clear gap in the literature (Sadorski, 2021). The authors further state that, while stock prices have generally been considered to be unpredictable (Malkiel, 2003) there are multiple patterns based on momentum and psychology that provide possible sources of prediction targets (Gray and Vogel, 2016; Lo et al. 2000; Basak et al. 2019). Using this potential and industry-accepted technical indicators the authors demonstrate the desirability of Random Forest models as compared to more complex and opaque models that would be less approachable for most investors.

Random forests do well on this forecasting task because they make no assumptions about linearity and can capture non-linear relationships other approaches miss. RF also provides approaches for preventing overfitting and by giving information about the contribution of each indicator to the prediction adds a great deal to interpretability and explainability.

Methods Used

The study uses Random Forest model algorithms to predict the future direction of five clean energy Exchange Traded Funds (ETFs) listed on the US stock exchanges using daily data from 2008 to 2020. The data were obtained from Yahoo finance. The features used are common technical indicators including relative strength index, stochastic oscillators, moving average convergence/divergence, price rate of change, on-balance volume, advance vs decline, and 200-day moving average (Murphy, 1999). The prediction is a binary indicator predicting whether the future price of the fund moves up or down over a 1- to 20-day period. The features were selected to provide potential patterns based on momentum, trend, and volume that could support the predictor.

The models are composed of ensembles of multiple decision trees. The number of trees selected was 500. Each tree was trained on a random subsample with replacement of 2/3 of the dataset with the predictions averaged to reduce the variance of the model (Bootstrap Aggregation or Bagging). The remaining 1/3 not seen by a tree was held as an Out Of Bag (OOB) sample for validation. The forecasts on each tree used 3 randomly chosen predictors, the number of predictors chosen as the floor of the square root of the number of predictor variables in the study, which was 10 (Breiman, 2001). Ten-fold cross validation with 10 repeats was used to control overfitting with OOB error estimates giving unbiased accuracy estimates.

Significance of the Work

The results support the authors' assertion of the advantages of tree-based ensemble methods over logistic regression and logit models. The accuracy rates of the Random Forest vs logit were 85-90% vs 55-60% with this range seen in both upward and downward movement forecasts. Because the explainability of the models the authors determined that on-balance volume and the 200-day moving average are the most significant predictors of future price direction with the highest accuracy and Gini index measures. The prediction accuracy is stable after 10 days indicating the models have stable utility in medium-range prediction.

This paper adds weight to the utility of Random Forest model techniques for pricing direction. More specifically it fills a hole within the research for clean energy stock predictions. The paper also cites the existing areas of research in clean energy stock prediction and its tie to oil prices, which lend additional features that would be used in future work. The results support the existing body of work that shows that Random Forest has high accuracy in stock price prediction (Ballings et al., 2015)

Connection to Other Work

This paper relies heavily on work by Billings and Basak to establish the basic viability of Random Forest in price prediction (Sadorsky, 2021; Ballings et al., 2015; Basak et al., 2019) . While studies have demonstrated the accuracy of the modeling techniques, they are largely unused in commercial technology with neural networks and support vector machines predominating. Traditional econometric models tend to use modeling approaches that rely on or assume an underlying linearity or that data can be made linear approaches (Henriques & Sadorsky, 2008; Reboredo et al., 2017). This work demonstrates the applicability of Random Forest models for price predictions as well as the accuracy of those approaches but provides application of the techniques within the lesser-studied clean energy fund equities. The results of the models show superior accuracy on non-linear data without the complexity of neural networks with additional insight provided through the model architecture and feature metadata.

Relevance to Capstone Project

This paper is directly relevant to my capstone project on stock market index prediction using random forests. From this study I have gained additional insight into key indicators I may want to use in my own work. The multi-step forecast horizons will help me to gain additional accuracy over prediction timeframes and help me evaluate how accuracy changes over time. The number of trees used as well as the random feature selection at each split and the cross-validation approach will help me in my approach to implementing Random Forest over my financial data.

References

- Ballings, M., Van den Poel, D., Hespeels, N., & Gryp, R. (2015). Evaluating multiple classifiers for stock price direction prediction. *Expert Systems with Applications*, 42(20), 7046–7056.
- Basak, S., Kar, S., Saha, S., & Khaidem, L. (2019). Predicting the direction of stock market prices using tree-based classifiers. *North American Journal of Economics and Finance*, 47, 552–567.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- Gray, W. R., & Vogel, J. (2016). Momentum and other anomalies. *Journal of Financial Economics*, 122(1), 1–21.
- Henriques, I., & Sadorsky, P. (2008). Oil prices and the stock prices of alternative energy companies. *Energy Economics*, 30(3), 998–1010.
- International Energy Agency. (2020). World energy investment 2020. IEA.
- Lo, A. W., Mamaysky, H., & Wang, J. (2000). Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation. *Journal of Finance*, 55(4), 1705–1765.
- Malkiel, B. G. (2003). The efficient market hypothesis and its critics. *Journal of Economic Perspectives*, 17(1), 59–82.
- Murphy, J. J. (1999). Technical analysis of the financial markets. New York Institute of Finance.
- Reboredo, J. C., Rivera-Castro, M. A., & Ugolini, A. (2017). Downside and upside risk spillovers between oil and renewable energy stocks. *Energy Economics*, 61, 69–81.
- Sadorsky, P. (2021). A random forests approach to predicting clean energy stock prices. *Energy Economics*, 100, 105339.