Investigating the relationship between trading volumes and stock market returns using Random Forest (RF) algorithm.

Leonard Mushunje.

Midlands State University, Department of Applied Mathematics and Statistics.

leonsmushunje@gmail.com

Abstract.

Trading returns are surely the key focus for all investors within financial markets. On the other hand, an automatic analysis of these returns against other factors of interest is considerably becoming a common way to investors and other key stakeholders in the stock markets. Despite the prior knowns and unknowns about stock returns and trading volumes, this paper aims to provide an exploration of these two using machine learning based approach. Purposefully, we intended to explore the statistical relationship between stock returns and trading volumes using Johannesburg stock market data for a spanning period of 2000 to 2017. To meet our goals, we used the random forest (RF) algorithm based on R programming. We performed the random forest regressions to see the cause and effect relations between these two and as well the Spearman correlation test and the Variance–Covariance matrix analysis to fully evaluate the state of the relationship between stock returns and trading volumes. Our results suggested that there is a positive correlation between trading volumes and stock returns as doubly supported by the correlation tests and regressions done. Above all we noted that, random forest as part of machine learning (ML) best estimates our variable regressions and correlations with a better precision and turnaround time which is of noble significance. Therefore an emphasized recommendation on the use of such tools is strongly and highly encouraged by this paper.

Keywords: stock returns, random forest, trading volumes, correlation, machine learning.