ABSTRACT

Based on previous literature, which develops an automated risk forecasting system for a large panel data of S&P 100 stocks using 118 features and five machine learning algorithms, we aim to improve the performance of these five machine learning algorithms before combining them into the automated system. We define a more precise and comprehensive method for retrieving and cleaning the stock data. Then, we adjust those 118 features and fit the same five ML algorithms by keeping similar parameters as the original paper does but adding validation-step into the model selection for stock volatility prediction. Given data with different time intervals, day, week, month, and quarter, we record the out-of-sample $R^2$ respectively and compare them to that of a benchmark VAR model. The result shows we improve the machine learning models performance given the out-of-sample $R^2$ explains more proportion of the volatility.