Improved Value Investing with Machine Learning and an Innovative K-Nearest Neighbor Algorithm Variant: Development, Evaluation, and Deployment

He, Jun Yi (School: College Park High School)

The majority of approaches to forecast stock market behavior are plagued with scalability, profitability, and applicability issues. The present work comprehensively developed and evaluated 5 machine learning algorithms to overcome them using a fundamental analysis approach. There are two main contributions: conducting state-of-the-art experimentation on value investing with machine learning and developing an innovative, dimensionality-weighted k-Nearest Neighbors (DWKNN) Algorithm. Random Forest was found to obtain the highest return on investment, while keeping a strong ratio between return and risk over 12 test years. An innovative process involving picking the top 20 stocks to invest in based on the algorithms' confidence enhanced returns by approximately 30%. Metrics such as comparing precision as well as z-scores of test year with training years suggested that the out-of-sample data may be reasonably used for forecasting. Decision tree visualization allows for potentially novel insights into how to perform fundamental analysis. Finally, the returns of the algorithms were found to outperform most mutual and hedge funds in the real world. The DWKNN and a weight optimization algorithm were developed. The algorithm outperformed both regular KNN and WKNN on the same value investing dataset. It was further validated on other datasets to demonstrate its range of applications. Overall, the machine learning algorithms have been suggested to scalably outperform benchmarks, and the DWKNN has been suggested to obtain superior results. The value investing algorithms are being validated by industry experts and deployed.

Awards Won: