Abstract

The thesis investigates the application of machine learning in portfolio construction. The analysis was conducted on a dataset consisting of 442 American stocks. Initially, we cluster stocks using Principal Component Analysis and K-means algorithms. Then we select stock from each cluster based on return/risk metrics. Where risk was estimated by Value at Risk, and return was predicted using Random Forest and GARCH models. This leaves us with 11 stocks for every monthly period during 2020. The results indicate that the portfolios constructed from the selected stocks were able to outperform the market benchmark. However, the return predictions were not accurate enough. Thus, the portfolio from selected stock using the 1/N approach achieved better results than the portfolio optimized by the Mean-Variance model.