

ABSTRACT

IMPLEMENTATION OF DATA SCIENCE TECHNIQUES IN STOCKS PORTFOLIO MANAGEMENT

Portfolio management can be a formidable task that comprises of portfolio construction, selection and assessment, nonetheless there are tools and techniques in data science that can help with these tasks and in creating optimal portfolios. This quantitative study aims to create optimal portfolios using a set of techniques and methods derived from data science. The study relied on intraday data of KOMPAS100 stocks and market prices along with bond yields for risk-free rate from January 2010 to December 2021.

The study tested two clustering analysis techniques, *k-means* and *agglomerative hierarchical clustering*; to select and group stocks into portfolios. Two portfolio weights optimisation methods were also evaluated, the classical *mean-variance optimisation* (Markowitz, 1952) and a more recent, machine learning based *hierarchical risk parity* (Lopez de Prado, 2016). The portfolios created by these techniques were tested against the market and a set of randomly selected stocks portfolios.

The results showed that optimal portfolios can be created from the clusters produced the clustering analysis methods combined with portfolio weights optimised by *hierarchical risk parity*. In the validation stage, the optimal clusters from *k-means clustering* and *agglomerative hierarchical clustering* provided holding returns of 63.19% and 58.38%, with risk-adjusted Sharpe of 0.676 and 0.678 respectively. This study is intended to be a reference for portfolio managers, investment managers, individual investors and other researchers in the field of financial management with interest in financial data science.

Keywords: portfolio management, optimal portfolio, financial data science, mean-variance analysis, hierarchical risk parity, clustering analysis

INTISARI

IMPLEMENTASI TEKNIK DATA SCIENCE DALAM MANAJEMEN PORTOFOLIO SAHAM

Manajemen portofolio bisa menjadi tugas berat yang terdiri dari konstruksi portofolio, seleksi dan penilaian, namun ada teknik dalam ilmu data yang dapat membantu proses ini dan menciptakan portofolio yang optimal. Studi kuantitatif ini bertujuan untuk menciptakan portofolio yang optimal dengan menggunakan seperangkat teknik dan metode yang diturunkan dari ilmu data. Kajian ini mengandalkan data harga harian saham KOMPAS100 dan harga pasar serta imbal hasil obligasi untuk tingkat bebas risiko dari Januari 2010 hingga Desember 2021.

Penelitian ini menguji dua teknik analisis kluster yaitu, *k-means* dan *agglomerative hierarchical clustering*; untuk memilih dan mengelompokkan saham menjadi portofolio. Disamping itu, pengoptimalan bobot portofolio dilakukan menggunakan metode klasikal *mean-variance optimisation* (Markowitz, 1952) dan metode lebih terkini yaitu, *hierarchical risk parity* yang berbasis pembelajaran mesin (Lopez de Prado, 2016). Portofolio yang dihasilkan oleh rangkaian metode ini adalah diuji terhadap pasar dan portofolio dibentuk dari saham yang dipilih secara acak.

Hasil penelitian menunjukkan bahwa portofolio optimal dapat dibuat dengan menggunakan kluster yang dihasilkan oleh metode analisis kluster yang dikombinasikan dengan bobot portofolio yang dioptimalkan menggunakan *hierarchical risk parity*. Pada tahap validasi, kluster optimal dari *k-means* dan *agglomerative hierarchical clustering* memberikan *holding return* sebesar 63,19% dan 58,38%, dengan Sharpe yang disesuaikan dengan risiko masing-masing sebesar 0,676 dan 0,678. Penelitian ini ditujukan untuk menjadi referensi bagi manajer portofolio, manajer investasi, investor individu dan peneliti lain di bidang manajemen keuangan yang tertarik pada ilmu data keuangan.

Kata Kunci: manajemen portofolio, portofolio optimal, ilmu data keuangan, analisis kluster, analisis mean-variance, hierarchical risk parity