

Klasterisasi Saham pada Sembilan Sektor Saham dengan Pendekatan Data Mining = Stock Clustering in Nine Sectors using Data Mining Approach

Dhea Indriyanti, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20512710&lokasi=lokal>

Abstrak

Dalam beberapa tahun terakhir, jumlah investor saham di Indonesia meningkat pesat, sehingga perlu dilakukan analisis tentang saham yang dapat membantu investor dalam rencana investasinya. Pengelompokan saham dapat membantu investor memilih saham yang sesuai untuk investor. Sayangnya, harga saham terus berubah dari waktu ke waktu. Akibatnya, kegiatan memilih saham untuk investasi bukanlah hal yang mudah. Selain itu, data time series saham dipengaruhi oleh banyak faktor yang mempertimbangkan time frame, kemudian menjadikan data pada setiap sektor memiliki jumlah faktor yang banyak yang disebut data dimensi tinggi. Karena itu, penting untuk menggunakan teknik yang cocok untuk mengelompokkan data dimensi tinggi. Penelitian ini menyajikan dua pendekatan yang dapat digunakan untuk data dimensi tinggi, yaitu subspace clustering dan dimension reduction. Pendekatan subspace clustering menggunakan metode High Dimensional Data Clustering (HDDC), sebuah teknik klasterisasi berbasis model berdasarkan Gaussian Mixture Model, dengan menggunakan algoritma Expectation-Maximization (EM). Pada pendekatan dimension reduction menggabungkan teknik reduksi fitur dan teknik klasterisasi yang sudah sering digunakan yaitu K-Means. Penelitian ini menggunakan dua pendekatan reduksi fitur, yaitu ekstraksi fitur menggunakan Principal Component Analysis (PCA) dan pemilihan fitur menggunakan Correlation Attribute Evaluation. Luaran dari penelitian ini adalah 2 klaster terbentuk di sektor agrikultur, 3 klaster di sektor pertambangan, 4 klaster di sektor industri dasar dan kimia, 2 klaster di sektor aneka industri, 2 klaster di sektor industri konsumsi, 2 klaster di sektor properti dan real estate, 2 klaster di sektor infrastruktur, 2 klaster di sektor keuangan, dan 4 klaster di sektor perdagangan. Dari perhitungan indeks validasi klasterisasi, teknik seleksi fitur memberikan performa yang lebih baik.

.....In recent years, the stock investor in Indonesia has been increasing rapidly, hence it is required to conduct analysis about the stock that helps the investor in their investment plan. Clustering is beneficial to select the appropriate stock for investors. Unfortunately, stock prices keep varying from time to time. Consequently, it is not an easy work to select the stock for investment. In addition, stock price time series data influenced by many factors, so the factors in this study consider the time frame that makes the data in each sector has a larger number of features that called high dimensional data. In this study, high dimensional data are obtained by the time frame of each factor. Therefore, it is important to use a suitable technique to cluster high dimensional data. This study presents two approaches that can be used for high dimensional data, namely subspace clustering and dimension reduction. The subspace clustering approach uses High Dimensional Data Clustering (HDDC), a model-based clustering based on Gaussian Mixture Model, with using the Expectation-Maximization (EM) algorithm. The dimension reduction approach combines feature reduction techniques and common clustering technique, that is K-Means. This study uses two feature reduction approaches, namely feature extraction using Principal Component Analysis (PCA) and feature selection using Correlation Attribute Evaluation. The output of this study are 2 clusters formed in agricultural sector, 3 clusters formed in mining sector, 4 clusters formed in basic and chemical industry

sector, 2 clusters formed in various industrial sector, 2 clusters formed in consumption industry sector, 2 clusters formed in property and real estate sector, 2 clusters formed in infrastructure sector, 2 clusters formed in financial sector, and 4 clusters formed in trade sector. Based on the clustering validation index, feature selection techniques provide better performance.