

Prediksi Return Saham Berbasis *Clustering* Menggunakan *K-Means Clustering* dan *Convolutional Neural Network*

Rizaldi Maulid Fathurachman¹, Deni Saepudin²

^{1,2}Fakultas Informatika, Universitas Telkom, Bandung

¹rizaldimauidf@students.telkomuniversity.ac.id, ²denisaepudin@telkomuniversity.ac.id

Abstrak

Prediksi harga saham di masa depan merupakan sebuah masalah yang menantang. Dalam penelitian ini telah diterapkan pendekatan *clustering* untuk mengelompokkan saham-saham yang menunjukkan kesamaan dalam prediksi harga. Proses ini membantu dalam mempermudah prediksi harga, karena saham yang ada dalam 1 cluster yang sama diharapkan mempunyai pola yang lebih mirip. Melalui *clustering*, saham-saham yang memiliki prediksi harga yang sama dikelompokkan berdasarkan cluster yang telah terbentuk. Kemudian, dilakukan training model menggunakan data rata-rata dari harga saham dalam setiap cluster untuk menghasilkan prediksi harga yang lebih tepat. Dalam penelitian ini metode *K-Means* digunakan untuk mengelompokkan saham berdasarkan rata-rata return dan standar deviasi, sedangkan *CNN* digunakan untuk memprediksi harga saham harian. Selanjutnya prediksi return dapat dihitung berdasarkan prediksi harga yang telah diperoleh. Dataset harga saham LQ45 dibagi menjadi dua bagian, yaitu training data dan test data. Data training digunakan untuk melatih model *CNN* untuk menganalisis data harga saham historis yang dapat membantu memprediksi return saham di masa depan, sedangkan data test digunakan untuk mengevaluasi kinerja model yang telah dilatih. Hasil penelitian ini menunjukkan bahwa, dengan melihat kriteria elbow maka *clustering* dengan metode *K-Means* akan diambil dengan jumlah clusternya adalah sebanyak 4 cluster. Dan setelah itu di lakukan prediksi melalui *CNN* dan hasilnya mampu menghasilkan prediksi harga saham harian dengan baik, dibuktikan dengan nilai *RMSE* prediksi untuk 1 hari ke depan pada *cluster* 1 adalah 0,09, Untuk *cluster* 2, *RMSE* yang diperoleh adalah 0,05, Sementara itu, *cluster* 3 memiliki *RMSE* sebesar 0,07, dan untuk *cluster* 4 memiliki *RMSE* yang diperoleh yaitu 0,57. Setelah itu, dihasilkan dari prediksi saham-saham di setiap *cluster* yang menghasilkan prediksi *return* tertinggi di setiap harinya. Kemudian dibuat prediksi kinerja portofolio menggunakan *equal weight* dengan hasil rata-rata *return* sebesar 0,0009 dan standar deviasi sebesar 0,0209 yang lebih baik dibandingkan dengan indeks LQ45.

Kata kunci: Prediksi *return* saham, *clustering*, *K-means*, *Convolutional Neural Network (CNN)*, portofolio investasi.

Abstract

Predicting future stock prices is a challenging problem. In this research, a clustering approach will be applied to group stocks that show similarities in price movements. This process helps in making price prediction easier, as stocks in the same cluster are expected to have more similar patterns. Through clustering, stocks that have similar price movements are grouped based on the clusters that have been formed. Then, a training model is performed using average data from stock prices in each cluster to produce more precise price predictions. In this research, the K-Means method is used to group stocks based on average return and standard deviation, while CNN is used to predict daily stock prices. Furthermore, return predictions can be calculated based on the price predictions that have been obtained. The LQ45 stock price dataset is divided into two parts, namely training data and test data. Training data is used to train CNN models to analyze historical stock price data that can help predict future stock returns, while test data is used to evaluate the performance of the model that has been trained. The results of this study indicate that, by looking at the elbow criteria, clustering with the K-Means method will be taken with the number of clusters is 4 clusters. And after that, predictions are made through CNN and the results are able to produce daily stock price predictions well, as evidenced by the RMSE value of predictions for the next 1 day in cluster 1 is 0.09, For cluster 2, the RMSE obtained is 0.05, Meanwhile, cluster 3 has an RMSE of 0.07, and for cluster 4 has an RMSE obtained which is 0.57. After that, the predictions of stocks in each cluster are generated which produce the highest return predictions on each day. Then a portfolio performance prediction is made using equal weight with an average return of 0.0009 and a standard deviation of 0.0209 which is better than the LQ45 index.

Keywords: Stock return prediction, clustering, K-means, Convolutional Neural Network, investment portfolio.

1. Pendahuluan

1.1 Latar Belakang

Dalam penelitian ini dibahas mengenai tantangan dalam memprediksi harga saham di masa depan sebagai masalah utama dalam prediksi return saham. Masalah ini menarik karena dapat membantu investor