

ABSTRACT

This research was conducted to build models that can be used to classify short term stock's price movement pattern of 16 quotes listed in Indonesia Stock Exchange (IDX). The models classify Technical Indicator patterns into two classes (Up and Down). There are three main issues in using technical indicators for modeling. First, technical indicators data are nonlinear, nonstationary and have long memory dependency. Second, the use of multiple technical indicators can lead to redundancy caused by multicollinearity. Third, interpretation of technical indicator is open and subjective.

To handle those issues, the models were built in three phases: Technical Indicators transformation, dimensional reduction, and classification. The purpose of transformation is to denoise the original data. The dimensional reduction was implemented using Principal Component Analysis (PCA). PCA does the dimensional reduction by extracting and selecting the subset of extracted features. In selecting the subset, three selection methods were implemented: Broken Stick, Bootstrapped Kaiser Guttman and Accumulated Variance. The reduced dataset was used as input to train three Long Short Term Memory (LSTM) based classification models, which were: Shallow LSTM, Stacked LSTM dan Stacked LSTM with Dropout on Feed-Forward Layer.

The performance of classification models was analyzed using statistic method (Friedman test). The result shows that the Broken Stick method, which selects only 1-2 PC, consistently be the best performer in the three classification models. The suggestion of 1-2 PC reflects 27%-41% of the variance of the dataset. The result also shows that Stacked LSTM with the implementation of dropout on the feed-forward layer is the best classification model with an average accuracy of 64.09%. The comparison of models with PCA and without PCA shows that the implementation of PCA succesfully increase the average of accuracy by 8.21%.

Keywords : Stock, Price's Movement, Classification, Time Series, PCA, LSTM.

INTISARI

Penelitian ini dilakukan untuk membangun model klasifikasi arah pergerakan harga 16 saham Bursa Efek Indonesia (BEI). Tiga isu yang dihadapi dalam pembangunan model klasifikasi. Pertama data runtun data waktu finansial yang bersifat *nonlinear*, *nonstationary* dan memiliki *long memory dependency*. Kedua, pemanfaatan beberapa Technical Indicators berpotensi menyebabkan redundancy yang disebabkan multicollinearity. Ketiga, interpretasi pola yang dihasilkan indicator technical bersifat terbuka dan subjektiv.

Pembangunan model klasifikasi terdiri atas tiga tahap yaitu transformasi TI, reduksi dimensi dan klasifikasi. Transformasi TI bertujuan untuk mentransformasi data histori harga saham menjadi data dengan informasi tertentu. Tahap reduksi dimensi dilakukan menggunakan metode *Principal Component Analysis* (PCA). Pada penelitian ini dilakukan eksperimen terhadap 3 metode pemilihan PC yaitu *Broken Stick*, *Bootstrapped Kaiser Guttman* dan metode berbasis akumulasi *variance*. Tahap klasifikasi dilakukan menggunakan metode *Long Short Term Memory* (LSTM). Pada penelitian ini dilakukan eksperimen klasifikasi menggunakan tiga arsitektur LSTM yaitu *Shallow LSTM*, *Stacked LSTM* dan *Stacked LSTM + implementasi dropout* pada *layer feed forward*.

Hasil uji *Friedman* memperlihatkan metode pemilihan PC yang secara konsisten menghasilkan unjuk kerja klasifikasi terbaik adalah metode *Broken Stick*. Metode *Broken Stick* mengajukan penggunaan 1-2 PC dengan akumulasi *variance* terjelaskan sebesar 27% sampai 41%. Analisis hasil klasifikasi juga menunjukkan bahwa arsitektur *Stacked LSTM + implementasi dropout* pada *Layer Feed Forward* merupakan arsitektur yang menghasilkan rata-rata akurasi tertinggi sebesar 64.09%. Akurasi tertinggi diperoleh PT. Tempo Scan Pacific, Tbk (TSPC) sebesar 76.79% dan akurasi terendah diperoleh PT. Bank Mandiri, Tbk (BMRI) sebesar 55%. Hasil perbandingan dengan model tanpa implementasi PCA memperlihatkan bahwa PCA berhasil menaikkan rata-rata akurasi klasifikasi sebesar 8.21%.

Kata kunci -- Saham, Klasifikasi, *time series*, *Machine Learning*, PCA, LSTM