

ABSTRACT

COMPARATIVE STUDY OF SUPPORT VECTOR MACHINE AND LINEAR DISCRIMINANT ANALYSIS FOR STOCK PRICE PREDICTION BASED ON SENTIMENT ANALYSIS WITH TECHNICAL INDICATORS

Ahmad Helmi
20/454524/PA/19555

This undergraduate thesis will explore the application of sentiment analysis with the technical indicators to the dataset onto the machine learning model to see whether there is improvement in the models. The sentiment analysis that will be used in this thesis is the news headlines of the financial industry ranging from 2008-2016.

This research aims to compare the SVM and LDA performance on a dataset with the added technical indicators and also incorporate the hyperparameter tuning on the SVM model. Three scenarios will be implemented in this research. The first scenario is to experiment with just the sentiment analysis dataset with no added technical indicators into the dataset. The second scenario is to add the technical indicators to the dataset. The third scenario involves doing the hyperparameter tuning on the SVM model. These scenarios will use the same preprocessing method, the same 80-20 data splitting, and a random state of 0.

LDA consistently outperforms SVM in scenarios where no additional features are included, demonstrating significantly better performance. Even with the introduction of technical indicators, LDA remains robust, while SVM shows only modest improvements. However, when SVM undergoes extensive hyperparameter tuning, it can surpass LDA in performance, yet LDA proves to be more reliable and consistent, especially in cases where optimization is not feasible.

Keywords: Stock price prediction, technical indicators, SVM, LDA, sentiment analysis

INTISARI

COMPARATIVE STUDY OF SUPPORT VECTOR MACHINE AND LINEAR DISCRIMINANT ANALYSIS FOR STOCK PRICE PREDICTION BASED ON SENTIMENT ANALYSIS WITH TECHNICAL INDICATORS

Ahmad Helmi
20/454524/PA/19555

Skripsi ini akan mengeksplorasi penerapan analisis sentimen dengan indikator teknis pada dataset ke dalam model pembelajaran mesin untuk melihat apakah terdapat peningkatan pada model tersebut. Analisis sentimen yang akan digunakan dalam penelitian ini adalah tajuk berita dari industri keuangan selama periode 2008-2016.

Penelitian ini bertujuan untuk membandingkan kinerja SVM dan LDA pada dataset dengan penambahan indikator teknis, serta menggabungkan penyesuaian hyperparameter pada model SVM. Tiga skenario akan diimplementasikan dalam penelitian ini. Skenario pertama adalah bereksperimen dengan hanya menggunakan dataset analisis sentimen tanpa penambahan indikator teknis. Skenario kedua adalah menambahkan indikator teknis ke dalam dataset. Skenario ketiga melibatkan penyesuaian hyperparameter pada model SVM. Ketiga skenario tersebut akan menggunakan metode pra-pemrosesan yang sama, pembagian data 80-20 yang sama, serta menggunakan random state 0.

LDA selalu lebih unggul dari SVM saat tidak ada fitur tambahan, dengan performa yang jauh lebih baik. Bahkan setelah indikator teknis ditambahkan, LDA tetap stabil, sedangkan SVM hanya menunjukkan sedikit peningkatan. Namun, dengan tuning hyperparameter, SVM bisa melampaui LDA, meski LDA tetap lebih andal dan konsisten terutama saat optimasi tidak memungkinkan.

Kata kunci: Prediksi harga saham, indikator teknis, SVM, LDA, analisis sentimen