

## FORECASTING STOCK INDEX MOVEMENT DIRECTION WITH *CPL* LINEAR CLASSIFIER

Jerzy Krawczuk

Faculty of Computer Science, Białystok University of Technology, Białystok, Poland

**Abstract:** Stocks, indexes, commodities, and precious metals price prediction is a difficult task where many approaches are used: traditional technical analysis, econometric time series or modern data mining techniques. One particular data mining technique - linear classifier is described in this article. Prediction based on linear classifier is done using current market state, which can be described by various data sets (attributes, features). The simplest form of this model could use data from yesterday's price movement. Advanced models are using more historical price movements. Very advanced models include various historical price movements for indexes from other countries and other instruments like currencies, commodities, etc. Using more features requires extended time to estimate model parameters. We build the linear classifier models by the minimisation of a convex and piecewise-linear function which is very efficient comparing to other functions. Computational costs for building the model are similar to linear programming. We also use feature selection method called RLS. Those techniques allow us to explore data with many features. Four scenarios are considered, in each scenario a different amount of market data is used to create a model. In the simplest scenario only one day's change in price is taken, in the most complicated one 421 historical prices of 43 different instruments are taken. Best results were achieved by using middle range of 52 attributes. In this scenario, the model was right 53.19% times. Meaning the directions of daily change in S&P500 index (up or down) were predicted correctly. This doesn't seem a lot, but if those predictions would have been used for investing, they could produce a total profit of 77% in the tested time period from November 2008 to March 2011 (2 years 4 months), or an average of 28% per year.

**Keywords:** market forecast, market prediction, linear classifier, convex and piecewise-linear function

## **PROGNOZOWANIE KIERUNKU ZMIANY INDEKSÓW GIEŁDOWYCH ZA POMOCĄ KLASYFIKATORA LINIOWEGO TYPU CPL**

**Streszczenie** Prognozowanie cen akcji i wartości indeksów giełdowych jest zadaniem trudnym, gdzie używanych jest wiele technik takich jak: analiza techniczna, ekonometryczne szeregi czasowe, techniki eksploracji danych. Artykuł ten przedstawia jedną z metod eksploracji danych - klasyfikator liniowy. Klasyfikator ten w przeprowadzonym eksperymencie został użyty do prognozowania wartości indeksu giełdy amerykańskiej. Prognozowanie takie oparte jest o dane opisujące obecny stan giełdy. Stan giełdy można opisać różną ilością danych (atrybutów, cech). W najprostszym przypadku może to być tylko jednodniowa zmiana ceny prognozowanego indeksu. W bardziej rozbudowanym modelu można użyć wielu cen historycznych. W modelu jeszcze bardziej rozbudowanym można użyć danych z innych giełd, kursów walut, cen towarów jak np. ropa. Użycie dużej ilości danych wymaga dłuższego czasu obliczeń parametrów modelu. W prezentowanym podejściu klasyfikator liniowy budowany jest w oparciu o minimalizację wypukłej i odcinkowo-liniowej funkcji kryterialnej. Metoda ta jest bardzo wydajna o koszcie zbliżonym do programowania liniowego. Dodatkowo użyta została metoda selekcji cech RLS. Techniki te pozwoliły na efektywną eksplorację danych o wielu wymiarach. W artykule przedstawiono cztery scenariusze o różnej ilości danych opisujących giełdę. W najprostszym użytku tylko jednej danej, w najbardziej rozbudowanym 421 danych o 43 instrumentach finansowych. Najlepsze wyniki uzyskano dla pośredniego modelu o 52 cechach, w którym model przewidział prawidłowo 53.19% kierunków dziennych zmian indeksu S&P500. Otrzymany wynik nie wydaje się być wysoki, jednak gdyby inwestowano w indeks zgodnie z modelem zysk z takich inwestycji wyniósłby 77% w okresie od października 2008 do marca 2011, dając średnio 28% zysku rocznie.

**Słowa kluczowe:** klasyfikator liniowy, prognozowanie giełdy, funkcje wypukłe i odcinkowo-liniowe