GPW Analizer – badania nad modelem wnioskującym

# Wstęp

GPW Analizer jest systemem, który ma wspomagać decyzję inwestorów giełdowych. Po sparametryzowaniu przez użytkownika zapytania ma dać odpowiedź, czy dana inwestycja się zwróci. Przeznaczony jest dla inwestorów długoterminowych, tj. takich, którzy nastawiają się na mniejsze ryzyko, gdyż wpłacają pieniądze na dłuższe okresy czasu. Obsługiwanym rynkiem jest rynek regulowany Giełdy Papierów Wartościowych w Warszawie. Program wyliczający prognozy jest osobnym podprojektem o nazwie Stock Analyzer, został napisany w języku JAVA z użyciem pakietu WEKA. Niniejszy dokument jest sprawozdaniem z prac podjętych nad znalezieniem modelu, który jak najlepiej przewidzi przyszłe kursy spółek.

## Okres przewidywania

GPW Analizer przewiduje wartości akcji spółek na następujące okresy:

* 30 dni
* 60 dni
* 90 dni
* 180 dni
* 360 dni

## Dane statystyczne

Stock Analyzer jest programem, który dzięki danym statystycznym prognozuje przyszłe notowania. Do analiz zgromadziliśmy następujące dane:

* Giełdy zagraniczne – wartości. Dane pochodzą ze strony http://finance.yahoo.com
  + Dow Jones Industrial Average
  + NASDAQ Composite
  + Eurostoxx50
* Kursy walut – wartości. Dane ze strony http://nbp.pl/.
  + Dolar
  + Euro
* Indeksy GPW – kurs zamknięcia oraz wartość obrotu. Dane ze strony http://www.gpw.pl.
* Akcje Spółek – kurs zamknięcia, wolumen obrotu, liczba transakcji, wartość obrotu. Dane ze strony http://www.gpw.pl.

W bazie danych posiadamy powyższe dane od 01.01.2007r.

## Zasada dokonywania analiz

Akcje spółek są trudne do przewidzenia, gdyż zależą od wielu czynników. Pierwszym problemem była odpowiedź na pytanie, jak wnioskować wartości akcji. Można scalić ze sobą otrzymane dane i na ich podstawie wyliczyć ostateczną wartość akcji. W Stock Analyzer zdecydowaliśmy się przyjąć inny model wnioskujący, bowiem wnioskowanie odbywa się na następującej zasadzie:

* Przewidywanie wartości giełd zagranicznych.
* Na podstawie ww. przewidywań, oraz danych historycznych – waluty.
* Na podstawie ww. przewidywań, oraz danych historycznych – indeksy GPW.
* Na podstawie ww. przewidywań, oraz danych historycznych – wartości akcji.

Przy czym do przewidzenia pierwszych trzech punktów używamy wszystkich wcześniejszych danych, natomiast do wartości akcji bierzemy pod uwagę jedynie dane historyczne i indeksy, do których dana spółka należy.

## Algorytmy użyte do predykcji

Do predykcji wartości używamy następujących algorytmów zaimplementowanych w pakiecie WEKA:

* GaussianPricesses
* MultiLayerPerceptron
* RBFNetwork
* SMOreg
* IBk
* KStar
* LWL
* AddictiveRegression
* Bagging
* Stacking
* Vote
* ConjunctiveRule
* M5Rules
* ZeroR
* DecisionStump
* M5P
* REPTree

Docelowo ostateczny wynik ma być średnią wyników poszczególnych algorytmów.

## Od czego zależny jest wynik? Co tak naprawdę chcemy ustalić?

Ostateczny wynik jest zależny przede wszystkim od wyników poszczególnych algorytmów. Jednak aby model był dokładniejszy postanowiliśmy każdemu z algorytmów nadać wagę i to właśnie ustalenie odpowiedniej wagi jest jednym z dwóch celów badań. Drugim jest ustalenie liczby dni, którą należy wziąć do poszczególnej analizy (30, 60, 90, 180, 360). Manipulując tymi parametrami spróbujemy ustalić model, który w przeszłości uzyskałby najlepsze wyniki.

## Błąd modelu

Aby móc określić, który model jest lepszy, musimy ustalić, co przyjmiemy za jego błąd. Badanie dla określonych parametrów da nam wynik w postaci tablicy z składającej się z n elementów, w którym każdy oznacza pomyłkę algorytmu w stosunku do realnej wartości wyrażony w procentach. Następnie każdy element tej tablicy mnożymy przez współczynnik, który przewartościuje nam błędy. Współczynnik ten wyraża się wzorem:

gdzie:

* z – nasz współczynnik
* w – wartość wyliczona przez model
* maxW – największa wartość spośród wszystkich wyliczonych

Na sam koniec wyliczamy statystykę, tj. funkcję elementów tablicy zdefiniowaną jako średnią arytmetyczną:

Tym samym otrzymujemy błąd naszego pomiaru.