

ABSTRAKT

mgr Piotr Łukasz Bednarski

„Wykorzystanie algorytmów NLP do automatycznej analizy treści informacji finansowych w procesach prognozowania notowań giełdowych”

W dzisiejszym świecie rynków finansowych szybkość reakcji na informacje jest kluczowa dla skutecznego podejmowania decyzji inwestycyjnych. W niniejszej pracy doktorskiej zaproponowane zostało innowacyjne podejście do prognozowania notowań giełdowych wykorzystujące zaawansowane techniki przetwarzania języka naturalnego (NLP) oraz sieci neuronowe o długiej pamięci krótkotrwałej (LSTM).

Głównym celem tej pracy było zaproponowanie nowatorskiej architektury modeli uczenia maszynowego wykorzystujących algorytmy NLP, w tym sieci neuronowe LSTM oraz modele językowe typu BERT, do automatycznego przetwarzania i analizy treści wiadomości finansowych, a następnie zbadanie ich skuteczności. W tym kontekście tradycyjne modele ekonometryczne, takie jak ARMA-GARCH, stanowiły punkt odniesienia.

Proponowany model opiera się na integracji dwóch narzędzi: modeli językowych typu BERT, które pozwalają na zrozumienie kontekstu i znaczenia wiadomości finansowych, oraz sieci LSTM, umożliwiających uczenie się długoterminowych zależności w danych sekwencyjnych – w tym przypadku szeregach czasowych danych finansowych. W tym celu stworzony został autorski moduł FENE (Enkoder Wiadomości Ekonomicznych i Finansowych), którego celem jest automatyczna analiza tekstu wiadomości i wyekstrahowanie z niego informacji, które mogą poprawić jakość prognoz.

Dodatkowo zbadane zostało hybrydowe podejście, które łączy tradycyjne modele ekonometryczne z autorskim modulem FENE. Proponowana metoda wykorzystuje zalety obu podejść, zwiększając trafność prognoz poprzez lepsze wykorzystanie dostępnych danych oraz bardziej precyzyjne zrozumienie kontekstu informacji finansowych.

W celu oceny skuteczności zaproponowanych modeli zastosowano różnorodne mierniki trafności prognoz oparte na błędach prognoz, a także testach statystycznych.

Przeprowadzone badania sugerują, że autorskie podejście oparte na sieciach neuronowych LSTM i modelach językowych typu BERT może przewyższać tradycyjne metody ekonometryczne, takie jak modele typu ARMA-GARCH, w przewidywaniu notowań giełdowych na podstawie wiadomości finansowych. Uśredniony błąd prognoz RMSE autorskiego modelu jest niższy o 4.56% w porównaniu z modelami typu ARMA-GARCH i nawet o 25.4% niższy dla najbardziej medialnych spółek. Ponadto modele hybrydowe wykazują potencjał poprawy trafności prognoz poprzez wykorzystanie synergii pomiędzy różnymi technikami analizy danych. Błąd RMSE modeli hybrydowych jest do 1.92% niższy niż błąd klasycznych modeli ekonometrycznych. Niniejsza praca

stanowi istotny krok w kierunku wykorzystania metod uczenia maszynowego w celu lepszego zrozumienia i przewidywania zachowania notowań giełdowych.

Wnioski płynące z tej pracy wskazują na znaczący potencjał wykorzystania danych tekstowych zawartych w wiadomościach finansowych do usprawnienia procesu przewidywania notowań giełdowych. Integracja zaawansowanych technik NLP z modelami prognozowania rynkowego może przyczynić się do bardziej precyzyjnych i skutecznych decyzji inwestycyjnych.