

Prof. dr hab. Janusz Brzeszczyński
Katedra Rynku i Inwestycji Kapitałowych
Instytut Finansów
Wydział Ekonomiczno-Socjologiczny
Uniwersytet Łódzki

Recenzja pracy doktorskiej

mgr Piotra Bednarskiego

pod tytułem:

“Wykorzystanie algorytmów NLP

do automatycznej analizy treści informacji finansowych

w procesach prognozowania notowań giełdowych”

Wstęp

Recenzowana rozprawa doktorska dotyczy wykorzystania algorytmów przetwarzania języka naturalnego (*ang.* natural language processing – NLP) do analizy treści informacji i prognozowania notowań giełdowych na przykładzie 15 spółek giełdowych, 5 kryptowalut oraz 2 metali szlachetnych.

Oryginalny wkład omawianej pracy obejmuje propozycję autorskiej metody prognozowania szeregów czasowych z uwzględnieniem takich algorytmów wykorzystujących NLP jak mechanizm uwagi, technologia ‘word embeddings’ oraz głębokie sieci neuronowe.

Przedstawione w pracy badanie oparte jest na bazach wiadomości finansowych zarówno w postaci nagłówków jak i szerszego zakresu pełnych tekstów.

Cel, hipotezy badawcze i struktura recenzowanej rozprawy doktorskiej

Główny cel recenzowanej pracy określony został (na str. 11) jako zbadanie możliwości zastosowania najnowszych rozwiązań z dziedziny NLP do zautomatyzowanej analizy wiadomości oraz postów o tematyce finansowej, gospodarczej i politycznej wraz z uwzględnieniem ich w finansowych modelach predykcyjnych, a także stworzenie narzędzi (oprogramowania), które wspomagają wyliczanie takich prognoz.

Cel ten został zrealizowany przy pomocy 6 celów szczegółowych, które dodatkowo wsparte są 4 pytaniami badawczymi oraz 7 hipotezami badawczymi (przedstawionymi na str. 11 – 14).

Praca składa się ze wstępu, 6 głównych rozdziałów oraz rozdziału przedstawiającego konkluzje i zakończenie.

W rozdziale 1 zaprezentowane zostały ogólne dyskusje dotyczące rynku finansowego oraz roli informacji w prognozowaniu cen instrumentów finansowych. Rozdział 2 przedstawia opis modeli klasy ARMA-GARCH oraz modeli uczenia maszynowego. W rozdziale 3 szczegółowo opisane zostały techniki NLP oraz ich zastosowanie w analizach informacji finansowych. W Rozdziale 4 omówiona została metodologia badań wykorzystania technik NLP w analizie treści informacji finansowych, w tym architektura autorskiego systemu prognostycznego, opis treningu i walidacji modeli oraz metody oceny trafności prognoz. Rozdziały 5 i 6 prezentują wyniki badania empirycznego.

Literatura cytowana w pracy składa się z 204 pozycji w postaci artykułów opublikowanych w czasopiśmie, monografii oraz innych źródeł (w tym odwołań do stron internetowych).

Merytoryczna zawartość recenzowanej rozprawy doktorskiej

Podjęta przez Autora próba konstrukcji autorskiej metody prognozowania szeregów czasowych z uwzględnieniem algorytmów wykorzystujących NLP zasługuje na wysoką ocenę i w mojej opinii cel pracy został osiągnięty.

Pewne wątpliwości natomiast budzić może aspekt ekonomiczny przeprowadzonych badań, co w szczegółach omawiam w uwagach zamieszczonych w dalszej części recenzji.

Ocena formalnej strony recenzowanej rozprawy doktorskiej

Praca napisana jest dobrym językiem i została starannie przygotowana pod względem redakcyjnym. Znajdują się w niej tylko drobne nieścisłości stylistyczno-językowe, w tym na przykład:

- W kontekście zagadnień dyskutowanych w tej rozprawie doktorskiej, sądzę że lepszym terminem niż "ekwilibrium" jest po prostu "równowaga".
- W przypadku terminu "zwrot", w literaturze z zakresu rynków finansowych lepszym (i bardziej ugruntowanym) terminem jest "stopa zwrotu" (jest to termin bardziej właściwy niż proste tłumaczenie z jęz. angielskiego słowa "return", które jest – nawiasem mówiąc – skrótem od terminu "rate of return").

- Zamiast “wyliczanie prognoz” sugerowałbym raczej “tworzenie prognoz”, “formułowanie prognoz” lub “generowanie prognoz”.

- Podobnie w przypadku “wyciągnięcie konkluzji” sugerowałbym “sformułowanie konkluzji”.

Uwagi do recenzowanej rozprawy doktorskiej

Moje uwagi do omawianej pracy są następujące:

1) Recenzowana rozprawa doktorska od strony koncepcyjnej, i z perspektywy teorii rynku finansowego, umiejscowiona jest w kontekście hipotezy rynku efektywnego (*ang.* efficient market hypothesis), co stanowi przedmiot zwięzłej dyskusji w Rozdziale 1 w sekcji 1.1.7. Prowadzi to do określonych konsekwencji interpretacyjnych z punktu widzenia przedstawionych wyników badawczych, tzn. brak efektywności rynku (w tym, w szczególności, brak efektywności na poziomie słabym implikujący na gruncie teoretycznym, że istnieje informacja z przeszłości, w postaci np. historycznych cen instrumentów finansowych, która może stanowić podstawę decyzji inwestycyjnych prowadzących do ponadprzeciętnych wyników względem całego rynku) oznacza, że rezultaty zastosowania zaproponowanych metod w recenzowanej pracy mogą zostać wykorzystane nie tylko do wnioskowania na temat samej efektywności danego rynku, lecz również do formułowania oczekiwań względem skuteczności strategii inwestycyjnych na nich opartych (i do dalszych interpretacji uzyskanych wyników z takiej właśnie perspektywy).

Z tego też względu konstrukcja pracy pozostawia pewien niedosyt, gdyż w oparciu o uzyskane wyniki empiryczne dokonana została analiza z wykorzystaniem miar dokładności prognoz (tj. ME, MAE, MSE, RMSE, MPE oraz MAPE), jednak nie jest jasne jakie rezultaty przyniosłoby faktyczne zastosowanie zaproponowanych metod w praktyce w postaci osiąganych stóp zwrotu, a także innych miar korygujących o ryzyko tj. wskaźnik Sharpe’a / zmodyfikowany wskaźnik Sharpe’a, indeks Treynora czy miara CEQ (*ang.* Certainty Equivalent Returns) itp., względem rynku (co więcej, taka analiza powinna również uwzględniać odpowiednie koszty transakcyjne, które w praktyce znacząco różnią się między sobą w zależności od rynku oraz konkretnego instrumentu itp.).

Miary dokładności prognoz są oczywiście bardzo ważne i dostarczają wielu użytecznych informacji na temat jakości danej metody statystycznej, lub w szczególności modelu ekonometrycznego, a także bezpośrednio służyć mogą do prowadzenia interesujących analiz porównawczych, jednak niewiele one mówią na temat rzeczywistej skuteczności inwestycyjnej oraz, w kontekście hipotezy rynku efektywnego, nie są one w stanie zbyt dużo pomóc w kwestii wnioskowania na temat tego jak zaprezentowane wyniki empiryczne wyglądają w kontekście przywołanej teorii (w tym przypadku: hipotezy rynku efektywnego, do której bezpośrednio odniesienie znajduje się w Rozdziale 1 na początku omawianej pracy).

2) Dyskusja na temat hipotezy rynku efektywnego powinna zostać także znacznie rozwinięta o takie alternatywne koncepcje teoretyczne jak np. hipoteza rynku adaptacyjnego (*ang.* *adaptive market hypothesis*) czy finanse behawioralne, co nie tylko wzbogaciłoby pracę od tej strony ale również ułatwiłoby interpretację uzyskanych wyników empirycznych (nawiasem mówiąc, w roku 2013 została nawet przyznana Nagroda Nobla - którą otrzymali Eugene Fama, Lars Peter Hansen oraz Robert J. Shiller - za wkład w naukową dyskusję na temat funkcjonowania rynku finansowego oraz teorii rynków efektywnych, co obejmowało także powstałe później alternatywne koncepcje próbujące wyjaśniać to czego nie wyjaśniała wcześniej hipoteza rynku efektywnego zaproponowana przez Eugene Fama ponad 50 lat temu).

W recenzowanej pracy w sekcji 1.1.7 na str. 32 są jedynie wspomniane takie kwestie jak założenie o racjonalnym zachowaniu inwestorów oraz o przesłankach pozwalających przypuszczać, że hipoteza efektywności rynkowej “nie jest w mocy przynajmniej w krótkich okresach”, jednak wątki te powinny zostać znacznie bardziej rozwinięte (np. o dyskusję wskazującą, że hipoteza efektywności rynkowej może być w praktyce prawdziwa tylko okresowo, co stanowi podstawę m.in. hipotezy rynku adaptacyjnego), gdyż jest to w mojej opinii bardzo ważne zagadnienie w kontekście tematu omawianej rozprawy oraz przeprowadzonych w niej badań.

3) W części empirycznej, podstawę analiz stanowią zbiory wiadomości z wielu różnych źródeł, które wskazane są w tabelach 5.17 oraz 5.18 na str. 166-167. Nie jest jednak jasne co było kryterium ich wyboru, tym bardziej że grupa ta obejmuje zarówno bardzo duże serwisy informacyjne (np. *bbc.com*) jak i znacznie mniejsze serwisy lub strony internetowe.

4) W badaniu empirycznym wykorzystane zostały modele klasy GARCH i przedstawione wyniki (zaprezentowane w tabelach na str. 263 – 313) wskazują, że najczęściej miały one opóźnienia o długości 1 lub 2 okresów, jednak w pracy tej brak jest dyskusji na temat testów diagnostycznych, tj. testy na autokorelację oraz heteroskedastyczność składnika losowego. W przypadku tego typu modeli oraz danych, jest niestety prawdopodobne, że przy takich opóźnieniach cały czas występuje jeszcze zarówno heteroskedastyczność jak i autokorelacja, lecz nie zostało to zbadane (a przynajmniej wyniki odpowiednich testów nie są zaprezentowane w pracy). Sprawa ta jest o tyle istotna, że modele te stanowią ważne narzędzie w omawianej rozprawie (są one wykorzystywane jako modele referencyjne), więc wszelkie ewentualne błędy w estymacji ich parametrów (gdyby w określonych przypadkach występowała autokorelacja lub/oraz heteroskedastyczność) prowadziłyby do nieprawidłowych (lub przynajmniej niedokładnych) ostatecznych wniosków z przeprowadzonych badań.

5) Przedmiotem analiz zaprezentowanych w pracy są ceny akcji 15 spółek giełdowych, 5 kryptowalut oraz 2 metali szlachetnych. Nie wiadomo jednak co stanowiło kryteria selekcji tych właśnie instrumentów.

W przypadku akcji nie jest jasne dlaczego zostało wybranych akurat tych 15 spółek giełdowych? Jakie zastosowano kryterium (poza tym, że są to składowe indeksu S&P 500)?

Czy było to np. kryterium wielkości (mierzone kapitalizacją giełdową?) czy kryterium płynności? Na str. 127 jest wskazane jedynie, że są to “duże i znane” spółki oraz że charakteryzują się one “dużą płynnością”, jednak nie jest to poparte żadnymi danymi (tzn. czy jest to np. 15 najbardziej płynnych akcji na giełdzie w USA i według jakiej miary płynności zostały one tak sklasyfikowane?).

Podobne wątpliwości można sformułować w stosunku do pozostałych instrumentów, tzn. dlaczego akurat wybrano złoto i srebro jako metale? Jeśli kryterium także była płynność, lub wolumen obrotu na rynku międzynarodowym, to takie metale jak miedź czy aluminium mają historycznie większe obroty, więc ich rynki są uznawane za większe niż np. rynek srebra.

6) Zaproponowana autorska metodologia badawcza oparta na module FENE została bardzo dokładnie i klarownie opisana w Rozdziale 4 oraz dalej w Rozdziale 5 ilustrującym jej zastosowanie w badaniu empirycznym. Za szczególnie wartościowe uważam wykorzystanie w jej ramach danych na temat liczby oraz długości informacji na temat analizowanych instrumentów. Sądzę, że rozwinięcie tego właśnie kierunku badań (w połączeniu dodatkowo z analizą samej treści napływających na rynek wiadomości) ma duży potencjał publikacyjny w renomowanych czasopismach międzynarodowych.

7) Jak zaznaczyłem powyżej, w przeprowadzonych analizach z zastosowaniem zaproponowanych metod powinny zostać uwzględnione również wyniki strategii inwestycyjnych wskazujących jaki rezultat inwestycji przyniosłyby różne narzędzia. Co więcej, badanie takie powinno obejmować także analizę odporności na wybór okresu, tzn. prezentować spektrum wyników dla różnych okresów w których symulowane są tego rodzaju inwestycje. Rezultaty takiego badania byłyby bardziej informatywne niż różnice w miarach dokładności prognoz (tj. ME, MAE, MSE, RMSE, MPE czy MAPE), które z natury rzeczy są marginalne i trudno jest na ich podstawie właściwie ocenić jaka jest faktyczna przewaga danej metody and innymi rozważanymi metodami.

8) W kontekście powyższej uwagi, wspomniane wcześniej miary dokładności prognoz powinny zostać uzupełnione o miary zgodności kierunku (*ang.* direction quality measures), które szczególnie z punktu widzenia tematu recenzowanej rozprawy doktorskiej stanowią ważne narzędzia oceny, gdyż np. dla inwestorów giełdowych w praktyce ważniejszy niż precyzja prognoz punktowych jest sam kierunek zmiany ceny danego instrumentu finansowego, tzn. znak prognozowanej względem rzeczywistej stopy zwrotu, niż różnice (oceniane przy pomocy miar ME, MAE, MSE, RMSE, MPE, MAPE itp.).

Konkluzja

Recenzowana rozprawa doktorska dotyczy problematyki o znaczeniu zarówno metodologicznym jak i aplikacyjnym. Praca ta moim zdaniem stanowi interesujący wkład do badań nad algorytmami NLP i prognozowaniem rynków finansowych. Wyrażam jednocześnie

nadzieję, że moje uwagi sformułowane powyżej mogą posłużyć również jako zachęta do podjęcia dalszych badań lub lepszego umiejscowienia uzyskanych przez Autora dotychczasowych wyników w kontekście ekonomicznym (w szczególności w odniesieniu do teorii rynku finansowego), a kontynuacja pracy Autora prowadzić będzie do publikacji w renomowanych czasopismach międzynarodowych.

W podsumowaniu, uważam, że rozprawa doktorska mgr Piotra Bednarskiego pt. *“Wykorzystanie algorytmów NLP do automatycznej analizy treści informacji finansowych w procesach prognozowania notowań giełdowych”* spełnia wymogi merytoryczne i formalne określone w art. 187 ust. 1 oraz ust. 2 ustawy z dnia 20 lipca 2018 roku Prawo o szkolnictwie wyższym i nauce (Dz. U. z 2023 poz. 742 z późn. zmianami) i wnioskuję o dopuszczenie jej do publicznej obrony.



Prof. dr hab. Janusz Brzeszczyński

Łódź, 23.12.2024