

Bolesław Borkowski

Warszawa, 06.09.2015 r.

Katedra Ekonometrii i Statystyki

Wydział Zastosowań Informatyki i Matematyki

SGGW w Warszawie

## **R E C E N Z J A**

**rozprawy doktorskiej mgr inż. Salama Hassana Mhesn Al-augby**

**na temat**

**„*Text Mining Method in Evaluation of Media's Impact on Market Value Ratios*”**

**napisanej pod kierunkiem dr hab. inż. Kesry Nermenda, prof. US**

**przy współudziale promotora pomocniczego dr hab. Agnieszki Majewskiej**

### **1. Uzasadnienie podjęcia tematu badawczego**

Praca o charakterze metodologiczno – empirycznym jest z dziedziny nauk ekonomicznych, z zakresu zarządzania. W dysertacji doktorant wykorzystał metody text miningu do oceny wpływu informacji medialnych ze źródeł internetowych na kształtowanie się wskaźników wartości rynkowej. Jednym z najbardziej aktualnych problemów badawczych w badaniach ekonomicznych jest prognozowanie przyszłych cen akcji za pomocą wszelkich możliwych zestawów zmiennych. Metodologia prognozowania cen akcji ciągle ewoluuje. Liczba i jakość informacji, jakie w codziennym funkcjonowaniu otaczają inwestorów giełdowych sprawia, że rynek staje się coraz bardziej skomplikowany i trudny do przewidzenia. W związku z tym coraz powszechniej i w szerszym stopniu naukowcy piszą o związkach zmiennych nieekonomicznych (nazywanych czynnikami behawioralnymi) ze stopami zwrotu, które mogą w znaczny sposób wpływać na decyzje inwestorów giełdowych poprzez oddziaływanie na ich emocje. Od niedawna do standardowego zestawu cech objaśniających przyszłe ceny akcji wykorzystywane są informacje nieustrukturyzowane, czyli analizy tekstu zapisanego w języku naturalnym. Wielkość i różnorodność oraz zmienność tych informacji wymagają specjalistycznego narzędzia i metod analizy. Do analizy materiału empirycznego Autor słusznie wykorzystał metody text mining. Metody eksploracji danych tekstowych Text Mining łączą w sobie techniki Data Mining z analizowaniem różnorodnych dokumentów tekstowych. Dzięki tym metodom z nieustrukturyzowanych danych o charakterze tekstowym

można dostrzec nieznaną wcześniej informację, zależności i związki pomiędzy nimi a także zwiększyć bazę wiedzy. Głównym obszarem, na którym skupiły się badania w dysertacji, było zabarwienie emocjonalne informacji wpływających do inwestorów giełdowych. Dokonano ich podziału na „dobre”, „złe” i „neutralne”. Dzięki takiemu prostemu zabiegowi uzyskano opis nastrojów rynku, które przekładają się bezpośrednio na emocjonalne reakcje inwestorów giełdowych. Tego typu badania prowadzone są przez przedstawicieli nurtu promującego uwzględnianie psychologii w ekonomii. Nurt psychologiczny w analizach giełdowych opiera się na przekonaniu, że ruchy cen na giełdach są uzależnione w znacznym stopniu od nastawienia mentalnego inwestorów zarówno indywidualnych, jak i instytucjonalnych. Realizacja badań jest więc istotna nie tylko z punktu widzenia pojedynczego inwestora ale także dla inwestorów instytucjonalnych.

Wybór tematu rozprawy przez Doktoranta jest więc bardzo ważnym i aktualnym problemem badawczym. Praca dotyczy wykorzystania metody *text miningu* do oceny wpływu informacji medialnych ze źródeł internetowych na kształtowanie się wskaźników wartości rynkowej spółek notowanych na giełdzie papierów wartościowych.

Metody te zostały zweryfikowane na danych empirycznych pochodzących z giełd krajów Zatoki Perskiej. Takie podejście pozwoliło na bezpośrednie porównanie przydatności różnych statystycznych metod klasyfikacji. Uzyskane w ten sposób wyniki należy uznać za interesujące. Dlatego też uważam, że podjęty przez Doktoranta temat badawczy jest problemem aktualnym i w pełni uzasadnionym.

## **2. Struktura pracy, materiał empiryczny, metody badawcze**

Głównym celem pracy było „Opracowanie podstaw metodycznych do analizy wpływu mediów na kształtowanie się wskaźników wartości rynkowej na przykładzie rynków kapitałowych krajów Zatoki Perskiej z wykorzystaniem text miningu”.

Realizacja tego zadania została oparta na dwóch celach dodatkowych:

- ✓ Opracowanie algorytmu text miningu z serwisów internetowych na potrzeby pomiaru nastrojów rynku.
- ✓ Opracowanie autorskiego oprogramowania (text miningu) do analizy wiadomości z serwisów internetowych.

Na tej podstawie sformułowano następującą hipotezę naukową:

„Pojawianie się informacji medialnych wpływa na kształtowanie się wskaźników wartości rynkowej na giełdzie. Metoda text miningu umożliwia szybsze przekazywanie informacji o nastrojach rynku na rynku kapitałowym”.

Cele pracy, jak i hipoteza badawcza zostały sformułowane prawidłowo w kontekście analizy wpływu mediów na kształtowanie się wskaźników wartości rynkowej akcji na przykładzie rynków kapitałowych krajów Zatoki Perskiej z wykorzystaniem metody text mining.

Rozprawa doktorska mgr inż. Salama Al-augby odnosi się więc do tematyki aktualnej i ważnej, tak dla teorii jak i praktyki gospodarczej w zakresie szeroko rozumianych nauk o zarządzaniu.

Układ pracy jest podporządkowany potrzebom weryfikacji sformułowanej tezy. Przeprowadzona weryfikacja na rzeczywistych danych empirycznych algorytmu zbudowanego przez Autora i wykorzystywanego do klasyfikacji wiadomości według ich wpływu wykazała dużą jego przydatność. Doktorant wykazał się dużą wiedzą na temat metod analizy wielowymiarowej, potrafi umiejętnie wykorzystać wiedzę teoretyczną do analizy dużych baz danych, poprawnie stosując różne techniki eksploracji danych.

Rozprawa, napisana w języku angielskim, składa się ze wstępu, pięciu rozdziałów, podsumowania, bibliografii oraz załączników. Zajmuje łącznie 208 stron, przy czym 8 stron to streszczenie pracy w języku polskim. Struktura pracy jest spójna i logiczna; sprzyja ona realizacji celu pracy, a jednocześnie w pełni odpowiada tematyce rozprawy. Bibliografia pracy jest bardzo obszerna, obejmuje 418 pozycji opublikowanych głównie w języku angielskim. Zakres cytowanej literatury jest właściwy i odpowiada zagadnieniom prezentowanym w pracy.

### **3. Wartość merytoryczna pracy**

Praca ma charakter teoretyczno – empiryczny. Do części teoretycznej, oprócz dwóch pierwszych rozdziałów, w których Doktorant omówił podstawowe zagadnienia związane z podjętym tematem pracy, zaliczam opracowanie autorskiego algorytmu do rozpoznawania tekstu oraz opracowanie koncepcji systemu informatycznego do wspomaganie decyzji inwestycyjnych na rynku finansowym poprzez analizy wiadomości z mediów masowych w internecie. W rozdziale pierwszym Autor poprawnie omawia w sposób jednak zbyt szczegółowy zagadnienia związane z konstrukcją i funkcjonalnością na rynkach finansowych

i kapitałowych. Dla czytelnika z Polski ciekawym fragmentem jest opis funkcjonowania rynku w krajach Zatoki Perskiej. Autor opisuje struktury i znaczenie rynków kapitałowych oraz zmiany, jakie zaszły na rynkach kapitałowych, w ostatnim okresie w Zatoce Perskiej. Autor ponadto, na podstawie przeglądu literatury, zwraca uwagę na ważność analiz behawioralnych i wskazuje bardzo istotną rolę mediów na rynkach kapitałowych.

W drugim rozdziale Autor skupia się na przeglądzie metod analizy danych i tekstu, wykorzystywanych na potrzeby rynku kapitałowego. Autor opisuje metody eksploracji danych pod kątem analizy informacji pochodzących z mediów masowych oraz wskazuje na potrzebę wykorzystania metod klasyfikacji danych. W tym rozdziale Autor również przedstawia przydatność metod taksonomicznych w budowie miar agregatowych do badania zjawisk wielowymiarowych.

Do najważniejszej części pracy zaliczam rozdziały III, IV i V, które mają zasadnicze znaczenie dla realizacji celu głównego oraz celów szczegółowych sformułowanych w rozprawie.

Rozdział trzeci stanowi próbę weryfikacji hipotezy badawczej poprzez wykorzystanie metody eksploracji danych do analizy tytułów artykułów prasowych (nagłówków wiadomości) zamieszczonych w serwisach internetowych w horyzoncie czasowym od 1 października do 31 grudnia 2012 roku. Do obliczenia i pomiaru ekspansywności mediów Doktorant wykorzystał dwa rodzaje danych, pierwsze z nich to nagłówki wiadomości z internetowych serwisów informacyjnych oraz informacje ekonomiczne pozyskane z serwisu *alarabia.net* oraz z serwisu *reuters.com* w języku angielskim. Wstępna klasyfikacja zebranych nagłówków przeprowadzona została przez eksperta rynku kapitałowego (jak Go wybrano?). Jaki procent tych wiadomości tekstowych został odrzucony na etapie wstępnym?. Rzeczywiste dane rynkowe (drugi rodzaj danych) pobierano manualnie z różnych giełd i stron internetowych (jakich i jaki był ich dobór?) o dziennym interwale. Wyjaśnienia wymaga także dobór przez Doktoranta wskaźników finansowych ograniczonych do: stopy zwrotu z akcji –  $R_t$ , współczynnika ryzyka  $\beta$ ,  $Mar\ Cap$ ?, wskaźnika ceny do zysku –  $P/E$ , wskaźnika ceny do wartości księgowej oraz wskaźnika wypłaty dywidendy –  $DY$ . W rozdziale tym doktorant wykorzystuje metody budowy miary agregatowej *Vector Measure Construction Method (VMCM)* oraz metody kalkulacji mierników ekspansywności mediów, co jest podstawą do wykorzystania wielowymiarowej analizy porównawczej do wyznaczenia homogenicznych grup obiektów i okien czasowych niezbędnych do prawidłowego przeprowadzenia badania. Opis zaproponowanej metody *text miningu* do eksploracji danych służy do wzbogacenia zestawu wskaźników ilościowych, jakie wykorzystał doktorant w rozprawie, o wskaźniki

ekspansywności mediów w relacjonowaniu sytuacji gospodarczej (GNI, BNI, i NNI). Zaproponowana metoda rozpoznawania tekstu (podrozdział 3.3) stanowi jeden z głównych i cennych elementów rozprawy. Dzięki takiemu podejściu Autor przedstawia sposób pomiaru wpływu masowych mediów na wskaźniki ekonomiczne istniejące na rynkach kapitałowych. W celu kategoryzacji tekstu (wpływ pozytywny, negatywny czy neutralny) słusznie zastosowano: *Harvard IV Psychosocial Dictionary*, oprogramowanie *Wordsmith 4* oraz zbiór wyrażen frazeologicznych (idiomów) z zakresu nauk ekonomicznych w języku angielskim. Rozdział czwarty – to weryfikacja zaproponowanej metody badania wpływu środków masowego przekazu na kształtowanie się wskaźników rynkowych krajów Zatoki Perskiej. W profesjonalny sposób Autor przedstawił i przeanalizował wpływ środków masowego przekazu, przy wykorzystaniu wskaźników ekspansywności mediów, na kształtowanie się cen akcji spółek sektora bankowego i energetycznego. Doktorant wykazał dużą wiedzę z zakresu inżynierii finansowej oraz metod analizy danych wielowymiarowych. Data mining oraz różne techniki eksploracji danych są jednymi z najbardziej dynamicznie rozwijających się dziedzin w informatyce. Najistotniejszą częścią pracy doktorskiej jest opracowanie algorytmu text miningu z serwisów internetowych na potrzeby pomiaru nastrojów rynku (szczegółowy opis Autor przedstawił w drugim załączniku). Opracowany przez Doktoranta program analizowania wiadomości (NAP) jest zaimplementowany w środowisku programowania Python. Pomiaru wpływu środków masowego przekazu na kształtowanie się wskaźników rynkowych Doktorant dokonał w kilku etapach. W pierwszej kolejności wyznaczył współczynniki korelacji Pearsona, następnie oszacował najlepszy horyzont czasowy oraz dokonał podziału na jednorodne grupy. Do wydzielenia homogenicznych grup autor wykorzystał metodę k – średnich a obliczenia wykonał przy użyciu pakietu Statistica 10. Natomiast optymalny okres dla badania zmian wskaźników rynkowych giełdy został obliczony za pomocą miary odległości euklidesowej. Analiza materiału empirycznego została przeprowadzona w sposób poprawny przy wykorzystaniu najnowszych metod analizy wielowymiarowej. Chciałbym także podkreślić dobrą znajomość u Doktoranta istniejących pakietów programowych oraz umiejętność ich wykorzystania do badań.

W rozdziale piątym Doktorant prezentuje autorską koncepcję systemu informatycznego (DSS) wykorzystującego algorytm text miningu oraz opracowane miary agregatowe do badania dynamiki zmian na rynkach kapitałowych krajów Zatoki Perskiej. Na rynku dostępnych jest już wiele różnych pakietów, programów służących do analizy nieustrukturyzowanych danych tekstowych. Są to programy ukierunkowane do rozwiązywania konkretnych zagadnień. Propozycja budowy software'u ukierunkowanego na

wspomaganie podejmowania decyzji przez inwestorów w krajach Zatoki Perskiej jest – moim zdaniem – w pełni uzasadniona. Głównym założeniem tej koncepcji jest analiza wpływu wiadomości pochodzących z mediów masowych w internecie na podejmowanie decyzji inwestycyjnych, na rynkach finansowych krajów Zatoki Perskiej. System umożliwia dowolny wybór tekstu z podanych przez użytkownika nagłówków gazet, na podstawie którego program wyszukuje wiadomości lub słowa kluczowe służące do segregacji dobrych, złych i neutralnych wiadomości. Autor przedstawia ponadto formalny opis wymagań funkcjonalnych systemu z uwzględnieniem projektu interfejsu użytkownika. W pracy koncepcja budowy DSS przedstawiona jest w jasny i zrozumiały sposób dla czytelnika.

W podsumowaniu rozprawy zawarte są odniesienia do celu pracy i hipotezy badawczej oraz potencjalne kierunki dalszych badań. Przeprowadzone badania jednoznacznie potwierdziły postawioną przez Autora rozprawy doktorskiej hipotezę. Cele, zarówno główny jak i dodatkowe, zostały osiągnięte.

#### **4. Konkluzja**

W pracy zauważalne jest dobre przygotowanie merytoryczne Autora oraz wysokie umiejętności posługiwania się przez niego informacją i prawidłowym jej przetwarzaniem. Doktorant umiejętnie selekcjonuje informacje, formułuje prawidłowe wnioski oraz zmierza w kierunku uzyskania finalnego rezultatu. Wyniki analizy przedstawione w pracy potwierdzają przyjęte założenia badawcze. Podejście doktoranta do badanej problematyki w świetle aktualnej wiedzy jest właściwe. Doktorant przedstawił problem badawczy, sformułował hipotezę badawczą, dokonał przeglądu dotychczasowego dorobku naukowego w tym zakresie, zaproponował własne rozwiązania metodyczne i dokonał weryfikacji metod na danych empirycznych. Zaletą pracy jest dobra znajomość omawianych zagadnień, poparta połączeniem teorii z wynikami empirycznymi. Praca doktorska o charakterze naukowym ma istotne walory poznawcze, analityczne oraz aplikacyjne.

**Biorąc pod uwagę ocenę formalną, metodyczną i merytoryczną pracy należy stwierdzić, że odpowiada ona wymaganiom stawianym pracom doktorskim. Reasumując, uważam, że recenzowana praca spełnia stawiane wymogi i wnoszę o jej przyjęcie oraz dopuszczenie Doktoranta do publicznej obrony.**

*Bolesław Borkowski*

