

Prof. dr hab. Magdalena Osińska
Uniwersytet Mikołaja Kopernika
Wydział Nauk Ekonomicznych i Zarządzania
Katedra Ekonometrii i Statystyki

Toruń, dn. 24 sierpnia 2015 r.

RECENZJA

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 współdziałaniu promotora pomocniczego dr hab. Agnieszki Majewskiej

Uwagi ogólne o przedmiocie, celu i treści pracy

Techniki Data Mining zyskują współcześnie na znaczeniu z uwagi na ogromny zakres informacji, jaki codziennie dostępny jest dla każdego człowieka dzięki rozwojowi internetu, a także technik i technologii mobilnych. Informacje te można podzielić na najnowsze – pojawiające się w ostatniej chwili oraz starsze, które tworzą źródła archiwalne. Zbiory informacji powiększają się z każdą chwilą, co stawia poważne zadanie przed metodami i narzędziami ich eksploracji. W pierwszej dekadzie XXI w. powstał termin Big Data, który jest już nie tylko terminem marketingowym, ale przede wszystkim naukowym i praktycznym. Jak pisze S. Lohr (2012) w artykule pt. *The Age of Big Data* – Big Data to również zaawansowane trendy w technologii, przede wszystkim komputerowej. Z kolei P. Zikopoulos i Ch. Eaton (*Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data*, 2011) wskazują, że Big Data oznacza nową erę w eksploracji i wykorzystaniu danych. Big Data charakteryzuje się trzema cechami, o których można powiedzieć, że są duże: wielkość, różnorodność i zmienność. Takimi też cechami charakteryzują się informacje napływające na rynki finansowe, którymi zajmuje się w rozprawie doktorskiej Salam Al-augby. Inwestorów na rynkach finansowych interesują przede wszystkim informacje najnowsze, dostępne w danym momencie, takie, które bezpośrednio można wykorzystać w podejmowaniu decyzji finansowych i zarządzaniu finansami. Informacje te mogą mieć zarówno postać liczbową jak i tekstową. Z uwagi na krótki czas między napływem informacji

a podjęciem decyzji, informacje tekstowe muszą być pośrednio kwantyfikowane, ale w bardzo prostej postaci, tak aby inwestor był w stanie szybko tę informację przetworzyć. Najczęściej stosuje się prostą klasyfikację pozwalającą podzielić informacje na trzy grupy: informacje dobre (pozytywne), informacje złe (negatywne) oraz informacje neutralne. Jednak zakwalifikowanie nowej informacji tekstowej do jednej z tych trzech grup nie zawsze jest proste i wymaga wspomagania narzędziami statystycznymi oraz informatycznymi.

Oceniana rozprawa doktorska przygotowana przez p. mgr inż. S. Al-augby wpisuje się w nurt badań związanych z technikami analizy dużych zbiorów informacji, dotyczy bowiem wykorzystania metody *text miningu* do oceny wpływu informacji medialnych ze źródeł internetowych na kształtowanie się wskaźników wartości rynkowej spółek notowanych na giełdzie papierów wartościowych. Ilość i jakość informacji, jakie w codziennym funkcjonowaniu otaczają inwestorów giełdowych sprawia, że rynek staje się coraz bardziej skomplikowany. W związku z tym inwestorzy coraz częściej polegają na odczuciach i emocjach, jakie na pierwszy rzut oka wywołują napływające informacje. Te odniesienia behawioralne wpływają na zachowania inwestorów, a to z kolei skutkuje zmianami stóp zwrotu, które nie zawsze sprzyjają właściwej wycenie instrumentów finansowych. Głównym obszarem, na którym skupiły się badania w dysertacji, było zabarwienie emocjonalne informacji wpływających do inwestorów giełdowych. Dzięki podziałowi informacji na trzy wspomniane wyżej klasy uzyskano opis nastrojów rynku, które przekładają się bezpośrednio na emocjonalne reakcje inwestorów giełdowych. Dlatego też recenzowana praca wpisuje się także w nurt finansów behawioralnych, uwzględniających elementy psychologiczne zarówno w modelach jak i w analizach empirycznych. Z powyższych powodów temat rozprawy, jakim jest *Text Mining Method in Evaluation of Media's Impact on Market Value Ratios*, uznaję za aktualny, wyodrębniony problem badawczy, odpowiedni do rozwiązania na poziomie rozprawy doktorskiej.

Autor postawił w pracy następujący cel główny: *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 wsparta poprzez sformułowanie dwóch celów dodatkowych, a mianowicie:

- Opracowanie algorytmu *text miningu* z serwisów internetowych na potrzeby pomiaru nastrojów rynku.

- Opracowanie autorskiego oprogramowania (*text miningu*) do analizy wiadomości pochodzących z serwisów internetowych.

Rozważania zawarte w pracy zmierzają do wykazania prawdziwości hipotezy, że *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.*

Zarówno cele pracy, jak i hipoteza badawcza zostały sformułowane trafnie i stanowią spójną podstawę opracowania metodyki służącej do analizy wpływu mediów na kształtowanie się wskaźników wartości rynkowej akcji na przykładzie rynków kapitałowych krajów Zatoki Perskiej z wykorzystaniem *text miningu*, co nastąpiło w kolejnych częściach pracy. Reasumując, rozprawa doktorska mgr inż. Salama Al-augby odnosi się do tematyki nie tylko aktualnej i ważnej, ale także bardzo potrzebnej zarówno dla teorii jak i praktyki gospodarczej w zakresie szeroko rozumianych nauk o zarządzaniu. Jednocześnie stwierdzam, że zarówno cele pracy jak i postawiona hipoteza badawcza mają charakter naukowy i stanowią spójną całość.

Na podstawie sformułowanej powyżej oceny ogólnej przedmiotu, celu i treści pracy stwierdzam, że temat rozprawy doktorskiej został wybrany i sformułowany prawidłowo, przedmiot badań uważam za istotny dla nauki i praktyki, a ujęcie problemu stanowi wyodrębnione zadanie badawcze, odpowiadające wymogom rozprawy doktorskiej.

Ocena zawartości rozprawy

Rozprawa, która została napisana w języku angielskim, składa się ze wstępu, pięciu rozdziałów, podsumowania, bibliografii oraz załączników. Zajmuje łącznie 208 stron, przy czym 17 stron przeznaczono na streszczenie pracy napisane w języku polskim. W pracy zamieszczono także spis skrótów stosowanych przez Autora, co znacznie ułatwiło jej lekturę. Struktura pracy jest spójna i logiczna; sprzyja ona realizacji celu pracy, a jednocześnie w pełni odpowiada tematyce rozprawy. Bibliografia pracy jest bardzo obszerna obejmuje bowiem aż 418 pozycji opublikowanych głównie w języku angielskim, pochodzących w większości z ostatnich 20. lat. Zakres cytowanej literatury jest właściwy i odpowiada zagadnieniom prezentowanym w pracy. Bardzo detaliczna analiza spisu literatury wskazuje na to, że zabrakło tam pozycji wydanych przez M. Walesiaka czy E. Gatnara, dotyczących analizy taksonomicznej czy metod klasyfikacji, jednak zagadnienia te są reprezentowane w innych artykułach czy książkach, które zostały zacytowane.

We wstępie Autor dokonał przedstawienia motywacji, celu i hipotezy badawczej oraz skrótu zawartości poszczególnych rozdziałów. Wstęp jest przejrzysty i stanowi dobre wprowadzenie do merytorycznej części rozprawy.

W rozdziale pierwszym Doktorant omawia zagadnienia związane z konstrukcją i funkcjonalnością rynków finansowych i kapitałowych, w szczególności przybliżając te zagadnienia, które dotyczą funkcjonowania rynków w sześciu arabskich krajach Zatoki Perskiej, tj. w Bahrajnie, Kuwejcie, Omanie, Katarze, Arabii Saudyjskiej oraz Zjednoczonych Emiratach Arabskich. Autor dokonuje także przeglądu klasycznych metod analizy rynku kapitałowego, metod analizy zdarzeń czy metod behawioralnych oraz wskazuje na istotną rolę mediów w kształtowaniu procesów zachodzących na rynkach kapitałowych na podstawie przeglądu literatury. Przegląd literatury dokonany w pracy doktorskiej przez mgr inż. S. Al-augby należy ocenić wysoko.

W drugim rozdziale uwaga 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.

Rozdziały I i II mają charakter przeglądowy. Opis wszystkich uwzględnionych tam elementów nie był konieczny dla realizacji celu pracy (np. opis analizy technicznej, fundamentalnej czy portfelowej), wystarczyło powołać się na stosowną literaturę. Można jednak zrozumieć, że Doktorant będący obcokrajowcem poznał w ten sposób dorobek pracowników Wydziału Nauk Ekonomicznych i Zarządzania Uniwersytetu Szczecińskiego, w którym prowadzony jest przewód doktorski. Rozdziały III, IV i V 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 (ściśle: nagłówków wiadomości) zamieszczonych w serwisach internetowych w okresie od 1 października do 31 grudnia 2012 roku. Ponadto, w rozdziale tym Doktorant wykorzystuje metody budowy miary agregatywnej Vector Measure Construction Method (VMCM) oraz metody kalkulacji mierników ekspansywności mediów, co jest podstawą do zastosowania wielowymiarowej

analizy porównawczej do wyznaczenia homogenicznych grup obiektów i okien czasowych niezbędnych do prawidłowego przeprowadzenia badania. Za obiekty badawcze posłużyły spółki giełdowe z sektorów bankowego i energetycznego, notowane na giełdach papierów wartościowych w sześciu wspomnianych krajach Zatoki Perskiej. Opis zaproponowanej metody *text miningu* do eksploracji danych służy do wzbogacenia zestawu wskaźników ilościowych, które wykorzystał Doktorant w rozprawie o wskaźniki ekspansywności mediów w relacjonowaniu sytuacji gospodarczej (GNI, BNI, i NNI). Zaproponowana metoda rozpoznawania tekstu, stanowi jeden z głównych elementów rozprawy. Dzięki takiemu podejściu Autor przedstawia sposób pomiaru wpływu masowych mediów na wartości wskaźników na rynkach kapitałowych, obserwowane głównie w krótkim okresie. W celu wyboru słów słusznie zastosowane zostały: *Harvard IV Psychosocial Dictionary*, oprogramowanie *Wordsmith 4* oraz zbiór wyrażen frazeologicznych (idiomów) z zakresu nauk ekonomicznych w języku angielskim.

W rozdziale czwartym opisana została konstrukcja i implementacja zaproponowanego algorytmu. Doktorant weryfikuje go na podstawie analizy danych pochodzących z rynku kapitałowych sześciu krajów Zatoki Perskiej. Dzięki zastosowanym metodom eksploracji danych Autor przedstawia wpływ mediów internetowych na zmiany cen akcji na badanych rynkach.

W rozdziale piątym przedstawiona została koncepcja systemu informatycznego zaprojektowanego i służącego do wspomagania podejmowania decyzji inwestycyjnych na rynku finansowym, poprzez analizę wiadomości pochodzących z mediów masowych w internecie za pomocą algorytmu *text miningu* oraz opracowanych miar agregatowych do badania dynamiki zmian na rynku kapitałowym. System umożliwia wybór dowolnego tekstu z podanych przez użytkownika nagłówek artykułów, 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.

Istotnym elementem pracy są załączniki. W pierwszym załączniku zawarto słownik wyrazów nie mających wielkiego wpływu na emocje inwestorów podczas podejmowania decyzji na rynku kapitałowym. Wyrazy te są usuwane automatycznie z tekstu wiadomości w celu minimalizacji ilości informacji, które będą poddawane dalszej analizie dla inwestorów. Drugi załącznik zawiera schemat blokowy autorskiego algorytmu *text miningu* na potrzeby

badania wpływu masowych mediów na funkcjonowanie rynku kapitałowego. W pozostałych załącznikach znajdują się dane źródłowe, wyniki badań, program *text miningu* oraz wygenerowane rankingi.

W podsumowaniu rozprawy zawarte są odniesienia do celu pracy i hipotezy badawczej oraz potencjalne kierunki dalszych badań. Kierunki dalszych badań należałoby rozszerzyć o porównanie zaproponowanej i zastosowanej w pracy metody na rynkach rozwiniętych (*developed markets*), które, ze względu na swoją płynność, cechują się jeszcze szybszym przepływem informacji aniżeli rynki wschodzące (*emerging markets*), do których należy zaliczyć giełdy w krajach zatoki Perskiej. Informacje na temat płynności papierów wartościowych wzbogaciłyby i uwiarygodniły także wyniki dotyczące państw omawianego w pracy regionu.

Zasadniczy wkład Autora w rozwiązanie postawionego problemu badawczego widoczny jest w rozdziałach: trzecim, czwartym i piątym. Badanie empiryczne przeprowadzone w pracy zostało starannie zaprojektowane i wykonane, a ponadto właściwie oprogramowane, przez co wpisuje się w nurt zarządzania finansami na rynkach kapitałowych. W wyniku tego zrealizowany został nie tylko cel naukowy pracy, ale także cele utylitarne wyrażone w formie dwóch celów pomocniczych.

Zrealizowane badanie za pomocą narzędzi *text miningu* przyczynić się może do wzbogacenia wiedzy w zakresie krótkookresowej analizy rynków kapitałowych, dokonywanej i modelowanej w sposób teoretyczny za pomocą nurtu badawczego związanego z mikrostrukturą rynków kapitałowych, o której pisali m.in. M. O'Hara, M. Doman czy K. Bień-Barkowska. Wiedza o reakcjach inwestorów indywidualnych oraz instytucjonalnych na informacje pojawiające się w mediach jest także istotnym elementem wspierającym rozwój finansów behawioralnych. Z drugiej strony zaproponowany algorytm oraz oprogramowanie stanowią narzędzia wspomagające zarządzanie inwestycjami finansowymi.

Zastosowane metody analizy oraz przedstawioną interpretację uzyskanych wyników należy uznać za wystarczające z punktu widzenia rozwiązania postawionych w pracy problemów. Postawiony i zrealizowany w pracy problem opracowania metody *text miningu* w celu oceny wpływu informacji medialnych ze źródeł internetowych na kształtowanie się wskaźników wartości rynkowej spółek notowanych na giełdach papierów wartościowych w krajach Zatoki Perskiej został rozwiązany w sposób oryginalny. Przeprowadzone badania jednoznacznie potwierdziły postawioną w pracy hipotezę. Cele, zarówno główny jak i dodatkowe, zostały osiągnięte. Autor wykazał się rzetelnym warsztatem badawczym,

a ponadto dobrą znajomością literatury przedmiotu oraz umiejętnością stawiania i rozwiązywania problemów naukowych. Recenzowana praca doktorska została przygotowana starannie zarówno w sensie językowym, jak i formalnym. Język pracy jest konkretny i komunikatywny, a wersja angielskojęzyczna w pełni zrozumiała.

Podsumowując, do podstawowych zalet recenzowanej pracy należy zaliczyć:

- Automatyzację pozyskiwania danych do obliczania wartości wskaźników ekspansywności mediów w relacjonowaniu wydarzeń gospodarczych.
- Propozycję autorskiego algorytmu do rozpoznawania tekstu, który został zaimplementowany w programie komputerowym. Algorytm służy do dostarczania inwestorom informacji pochodzących z publikacji medialnych dotyczących rynków kapitałowych w postaci wskaźników (GNI, BNI, NNI), które wspomagają podejmowanie podejmowania decyzji na rynku kapitałowym.
- Zastosowanie metod eksploracji danych do badania najkrótszego czasu, niezbędnego do reakcji inwestorzy na wiadomości dotyczące giełdy oraz spółek.
- Adaptację metody Vector Measure Construction Method (VMCM) do budowy miary agregatywnej w celu badania dynamiki zmian na rynku kapitałowym w kontekście informacji pojawiających się w nagłówkach publikowanych w portalach internetowych.
- Opracowanie koncepcji systemu informatycznego do wspomagania podejmowania decyzji inwestycyjnych na rynku finansowym, poprzez analizy wiadomości publikowanych w portalach internetowych.

Omawiana praca doktorska ma istotne walory poznawcze, analityczne oraz aplikacyjne. Przeprowadzone badania są interesujące i mają istotne znaczenie dla zarządzania inwestycjami prowadzonymi na rynkach kapitałowych.

Ocena końcowa

Rozprawa doktorska Pana mgr. inż. Salama Al-augby jest wartościowym studium potwierdzającym wiedzę i warsztat badawczy Doktoranta, pozwalającym na rozwiązywanie problemów badawczych w dziedzinie nauk ekonomicznych w dyscyplinie nauki o zarządzaniu. Praca w całej rozciągłości spełnia wymagania stawiane rozprawom doktorskim w ustawie z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki. W związku z tym wnoszę do Rady Wydziału Nauk Ekonomicznych i Zarządzania Uniwersytetu Szczecińskiego o przyjęcie pracy doktorskiej

i dopuszczenie mgr. inż. Salama Al-augby do jej publicznej obrony. Proponuję także rozważenie publikacji rozprawy w formie książkowej.

A handwritten signature in black ink, appearing to read 'M. D. ...' with a stylized, cursive script.