

DZIEKAN  
Kolegium Analiz Ekonomicznych

dr hab. Joanna Plebaniak, prof. nadzw. SGH

Warszawa 18.02.2016

25.02.16

Prof. zw. dr hab. Jerzy Kisielnicki  
Uniwersytet Warszawski  
Wydział Zarządzania

**Opinia o rozprawie doktorskiej Pana mgr Przemysława Pospieszego pt. Zastosowanie technik eksploracji danych do estymacji pracochłonności i czasu trwania projektów informatycznych, napisanej pod kierunkiem naukowym Dr hab. Andrzeja Kobylińskiego, prof. SGH w Instytucie Informatyki i Gospodarki Cyfrowej, Kolegium Analiz Ekonomicznych, Szkoły Głównej Handlowej w Warszawie**

Przedstawiana recenzja została napisana zgodnie z uchwałą Rady Kolegium Analiz Ekonomicznych Szkoły Głównej Handlowej w Warszawie z dnia 8 grudnia 2015 roku.

Decyzje jak i materiały potrzebne do wykonania recenzji otrzymałem 10 stycznia 2016 roku.

Przedstawiona do oceny rozprawa doktorska Pana mgr Przemysława Pospieszego liczy około 13 arkuszy wydawniczych.

**Zakres pracy i jej struktura**

Recenzowana praca poświęcona jest analizie wybranych problemów związanych z procesami zarządzania projektami. Doktorant w swojej dysertacji zajmuje się problematyką szacowania pracochłonności i czasu trwania projektów informatycznych. Praca jest próbą szacowania tych parametrów już na ich początkowym etapie realizacji a mianowicie na etapie przystąpienia do realizacji projektu tworzenia systemu informatycznego. Jako cel swoich badań, Autor podjął bardzo ambitne zadanie badawcze jakim jest, opracowania modeli estymacji wspomnianych wielkości (pracochłonność i czas) charakteryzujących projekt. Uzyskane wyniki jak i opracowana procedura estymacji ma zarówno wartość naukową jak też może zostać zastosowane w praktyce gospodarczej.

Celem pracy, jak pisze Autor na str. 17, jest cyt.

„Celem głównym niniejszej pracy jest budowa agregacyjnych modeli predykcyjnych z użyciem ogólnych modeli liniowych, sieci neuronowych oraz drzew decyzyjnych CHAID do estymacji pracochłonności i czasu trwania projektów informatycznych.”

Autor na stronach 17 i 18 przedstawił jako uzupełnienie celu podstawowego następujące cele: poznawcze, metodyczne i aplikacyjne. Uzupełnieniem przedstawionych celów są sformułowane następujące hipotezy badawcze cyt.:

1. Predykcyjne techniki eksploracji danych (data mining) mogą znajdować zastosowanie w zarządzaniu projektami informatycznymi, wspomagając proces estymacji pracochłonności i czasu trwania inicjatyw na ich inicjalnym etapie oraz potencjalnie przyczyniać się do wzrostu prawdopodobieństwa zakończenia projektu sukcesem. Przez to stanowią one narzędzie konkurencyjne do metod tradycyjnych oraz metod wykorzystujących linie kodu źródłowego lub punkty funkcyjne.
2. Ogólne modele linowe, wielowarstwowe sieci neuronowe oraz drzewa decyzyjne CHAID charakteryzują się *dostatecznie dobrą* (co to znaczy? – pytanie recenzenta) zdolnością predykcyjną pracochłonności i czasu trwania projektów informatycznych oraz odpornością na braki i szumy w danych, umożliwiając potencjalne ich wdrożenie w praktyce.
3. Agregacyjne predykcyjne modele eksploracji danych zastosowane do estymacji projektów informatycznych na początkowym etapie umożliwią otrzymywanie dokładniejszych szacunków badanych zjawisk niż użyte indywidualnie algorytmy.

Uwagi, do tak sformułowanych hipotez, przedstawię w dalszej części opinii.

Struktura ocenianej pracy i zawarte w niej treści merytoryczne są następujące. Praca składa się, obok wprowadzenia i zakończenia, z czterech rozdziałów merytorycznych. Dysertacja zawiera spis literatury oraz prezentacje ośmiu załączników w których podane są wyniki z przeprowadzonych badań.

Rozprawa doktorska została podzielona na cztery rozdziały. Rozdziały pierwszy i drugi mają charakter teoretyczny i przedstawiają analizę problemu estymacji zawartą w literaturze przedmiotu. Na ich podstawie Doktorant buduje procedurę badawczą. Rozdziały trzeci i czwarty poświęcone są przeprowadzonym badaniom i stanowi tzw. praktyczną część rozprawy.

Rozdział pierwszy zatytułowany "Zarządzanie projektami informatycznymi i metody estymacji parametrów projektu" to ogólne wprowadzenie do tematyki. Omówiono w nim czynniki wpływające na projekty informatyczne oraz kryteria ich oceny. Jak zaznacza się w tym rozdziale, jest to punkt wyjściowy do przedstawienia problematyki szacowania parametrów przedsięwzięć informatycznych. Taka kolejność prezentacji jest logiczna i nie nasuwa zastrzeżeń. Wychodzi się tu z założenia, że poprawna estymacja parametrów analizy projektu wpływa na realizację zamierzonych celów, którymi są prawidłowe określenie pracochłonności i długości trwania projektu informatycznego. W kolejnych podrozdziałach

została omówiona problematyka estymacji projektów i miar stosowanych w procesie ich szacowania. Przedstawione zostały w tej części pracy zarówno klasyczne techniki estymacji, takie jak: ocena przez analogię, szacowanie eksperckie czy dekompozycja, jak i bardziej zaawansowane metody parametryczne oparte na liniach kodu źródłowego (COCOMO/COCOMO II, SLIM, SEER-SEM) oraz na wymiarowaniu oprogramowania z użyciem punktów funkcyjnych (FPUG, NESMA, COSMIC). W końcowej części rozdziału Doktorant omawia, na podstawie literatury przedmiotu i swojego doświadczenia, zalety i wady stosowanych obecnie technik szacowania badanych parametrów projektów informatycznych.

Rozdział drugi p.t „Wiedza i jej odkrywanie” poświęcony jest problematyce tzw. odkrywania wiedzy w zarządzaniu projektami oraz stosowanym technikom eksploracji danych. Rozdział zawiera przegląd literatury z zakresu zastosowania technik data mining do estymacji parametrów projektu. Przedstawione zostały w tej części pracy obszary wiedzy wyróżniane w procesie zarządzania projektami oraz opisano informacje zbierane na każdym z etapów przeprowadzonej analizy procesu estymacji. Opis dotyczył: zbiorów danych (zawierających takie charakterystyki przedsięwzięć, jak budżet, pracochłonność), czas trwania, zastosowany język programowania, metodyka realizacji pracy. Punktem ciężkości analizy jest proces odkrywania wiedzy w zbiorach danych w tym eksploracja danych (data mining). W poszczególnych podrozdziałach przedstawiono stosowane w praktyce rodzaje technik data mining, jako narzędzia pozyskiwania wiedzy. Analiza literatury, przeprowadzona w tym rozdziale, jak pisze Doktorant, była niezbędna z punktu widzenia potrzeby opracowania podejścia agregacyjnego. W podejściu tym zostały użyte takie techniki, które umożliwią wdrożenie modeli predykcyjnych data mining w praktyce.

Rozdział trzeci zatytułowany „Estymacja projektów informatycznych z użyciem technik data mining” rozpoczyna część empiryczną rozprawy. Składa się on z dwóch podrozdziałów. Pierwszy poświęcony jest przygotowaniu danych a drugi natomiast zagadnieniom budowy tzw. modeli indywidualnych i agregatowych. W rozdziale tym przedstawiono wielobranżową bazę projektów informatycznych ISBSG. Baza ta która jest podstawą przeprowadzonych obliczeń, zawiera dane o zrealizowanych ponad 6000 projektów.

W rozdziale przedstawiono, oddzielnie dla pracochłonności i czasu trwania inicjatyw, proces budowy oraz konfiguracji modeli agregacyjnych. Do analizy zostały zastosowane następujące techniki predykcyjne: ogólny model linowy, sieci neuronowe oraz drzewa

decyzyjne CHAID. W końcowej części rozdziału przedstawiono uzasadnienie wyboru tych technik.

Rozdział czwarty „Ewaluacja modeli i analiza wyników” stanowi ewaluację indywidualnych modeli oraz agregacyjnych (uśrednionych) Ewaluacja jest przeprowadzona pod względem dokładności estymacji oraz możliwości wykorzystania ich w praktyce. Ostatnia część rozdziału czwartego przedstawia propozycję metodyki wdrożenia agregacyjnego modelu w praktyce. W zasadzie jest to zarys koncepcji metodyki a nie metodyka jaka jest stosowana w nauce.

Przestawioną strukturę pracy oceniam pozytywnie. Budowa całości rozprawy jest logiczna i odpowiada jej tytułowi.

### **Ocena rozprawy**

Na wstępie pragnę przedstawić swoją opinie o samym temacie pracy i jej zakresie.

Uważam, że wybór tematu był bardzo dobrą decyzją. Prace z zakresu projektowania systemów informatycznych w aspekcie analizy ryzyka są z punktu widzenia praktyki gospodarczej bardzo istotne. Jak przedstawi się w literaturze poświęconej teorii zarządzania projektami (por. między innymi: M. Trocki, J. Kisielnicki) , dla oceny projektu, niezależnie od jego przeznaczenia, są najbardziej istotne następujące parametry: zakres, budżet, czas. Autor wybrał z tej triady jako elementy estymacji: pracochłonności i czasu trwania projektów informatycznych. Tak więc z wymienionych wcześniej parametrów wybrał jeden a mianowicie czas trwania tego projektu. Drugi wybrany parametr pracochłonność, powiązany jest w pośrednio z budżetem przeznaczonym na realizację projektu.

Doktorant do przeprowadzenia badań wybrał, wspomniana wcześniej, bazę danych ISBSG. Nie korzystałem w swoich badaniach z tej bazy. Jej znajomość wynika tylko z zapoznania się ze stronami internetowymi na których jest rekomendowana. Wydaje się ona być odpowiednia do przeprowadzenia planowanych badań nad zamierzoną estymacją parametrów projektów informatycznych. Baza ISBSG, jak pisze się w internetowych materiałach informacyjnych, zawiera historyczne dane o zakończonych projektach informatycznych. Realizowane były one w wielu instytucjach publicznych i prywatnych i pochodziły z różnych gałęzi przemysłu i administracji. Dotyczą one zarówno nowych inicjatyw jak i modyfikacji istniejących systemów. Tu nasuwa się następująca uwaga. Czy dobranie jednej bazy danych jest reprezentatywne dla prowadzonych badań? Szkoda, że Autor nie weryfikował otrzymanych wyników na jednostkowych projektach realizowanych w Polsce. Jest wiele projektów informatycznych, które między innymi finansowane lub obecnie

są realizowane przez Unię Europejską. Może zbiór danych nie jest tak imponujący jak bazy ISBSG ale na pewno jest to cenny materiał do przeprowadzenia porównań.

Pozytywnie oceniam zasadniczy dobór narzędzi do przeprowadzenia badań. Zgadzam się ze stwierdzeniem Autora, że cyt. „ Szacowanie projektu jest niezwykle skomplikowanym przedsięwzięciem ze względu na niepewność i niejasność związaną z produktem finalnym jaki ma podlegać wytworzeniu. Odbywa się ono na podstawie niepełnej wiedzy, bazując na szerokim zakresie założeń.”. Autor potwierdza tu znane badania Schindlera, który stwierdził, że w trakcie realizacji projektów informatycznych ryzyko to zmniejsza się dość wyraźnie.

Doktorant w rozprawie pisze między innymi cyt. „Główną miarą, na podstawie której określany jest budżet i harmonogram projektów jest pracochłonność. Definiuje się ją jako ilość pracy wymaganej do wykonania określonego zadania lub jednostki produkcji, wyrażonej przeważnie w osobodniach”. Jest to prawda, jednak kształtowanie się tych parametrów zależy bardzo mocno od kwalifikacji zespołu realizującego i przyjętej metodyki pracy. Zaskakujące dla mnie jest to, że w pracy stwierdza się o stosunkowo małej wadze tych czynników. Brak w pracy uzasadnienia tego stanowiska. We wszystkich moich badaniach jednoznacznie wynika, że pracochłonność jest wynikiem działania wielu czynników z których najsilniej na nią wpływają kwalifikacje zespołu realizującego. Jak wyjaśniał Autor, problematyka kwalifikacji zespołu realizującego projekt nie jest rozpatrywana a to ze względu na brak wiarygodnych danych w analizowanej bazie ISBSG.

Uważam, że Doktorant przecenia rolę w szacowaniu pracochłonności, rozmiar „oprogramowania w postaci fizycznych linii kodu źródłowego (SLOC)” jak też obliczonym punktom funkcyjnym, które odpowiadają przyszłym funkcjonalnościom systemu. Może dawniej były to istotne parametry, ale nie są już we współczesnym projektowaniu.

Uważam, iż dyskusyjny jest wybór zmiennych do przeprowadzonej analizy. I tak Doktorant na stronie 180 podaje, że na pracochłonność mają wpływ (wg jednej z metod szacunku - sieci neuronowej) :

1. Industry sector 0,18
2. Relative size 0,17
3. Application type 0,17
4. Resource level 0,11
5. Architecture 0,09
6. Development platform 0,07
7. Package customization 0,06
8. Used methodology 0,05

9. Language type 0,04

10. Agile 0,04

Ale dalej już pisze cyt. iż „Wpływ poszczególnych zmiennych na badane zjawisko jest odmienny niż w przypadku ogólnego modelu liniowego”. To stanowisko potwierdzają uzyskane wyniki (strona 169), które wskazują nieco inne parametry, wpływu dokładność estymacji pracochłonności na podstawie wyznaczonych współczynników ważności dla następujących zmiennych (kolejność malejąca):

1. Relative size 0,27

2. Industry sector 0,18

3. Application type: 0,10

4. Used methodology 0,08

5. Architecture 0,08

6. Resource level 0,06

7. Development platform 0,06

8. Language type 0,04

9. Package customization 0,04

10. Development type 0,04

11. Agile 0,03

Zastanawiające jest, że wcześniejsze stwierdzenie o odmienności uzyskanych wyników, zostaje w na dalszych stronach dysertacji złagodzone. I tak podsumowując przeprowadzone badania pisze się, że: zbudowany model ogólny liniowy, wielowarstwowa sieć neuronowa oraz drzewo decyzyjne CHAID dobrze odwzorowują badane zjawisko i generują dokładne predykcje czasu trwania projektów. Lepsze (ale w jakim znaczeniu?) rezultaty jak pisze Doktorant uzyskiwał algorytm sieci neuronowej MLP. „Błędy generowane przez wszystkie techniki były niewielkie i na podobnym poziomie”. Te stwierdzenia, w świetle uzyskanych rezultatów, są dyskusyjne.

Każdy z zastosowanych w pracy (wg Autora) trzech modeli był dobrze dopasowany do danych i może być zastosowany w praktyce do predykcji czasu trwania inicjatyw informatycznych. Analizując otrzymane wielkości nie jestem o tym wniosku przekonany. Uważam bowiem, że pracochłonność zależy w największym stopniu od wielkości projektu, który w badaniach Autora nie zawsze jest na pierwszym miejscu a przewaga nad następnymi czynnikami nie jest zbyt wielka. Przypuszczam, że jest to spowodowane tym, że w badanej zbiorowości, projekty duże stanowią zaledwie 6% ogółu badanych, a projekt bardzo duży jest

na ponad 1500 projektów tylko jeden. Przeważają, jak pisze się na stronie 129, projekty małe i średnie (1349 obserwacji czyli 85%).

Pozytywne w przeprowadzonych badaniach jest to, że Doktorant przeprowadził pogłębione badania porównawcze stosowanych modeli. Uzyskane wyniki na pewno mogą być zastosowane do analiz szacunkowych a uzyskane rezultaty jednoznacznie wskazują na potrzebę kontynuacji badań.

Nie jest dla mnie zrozumiałe, dlaczego do oceny uzyskanych wyników równolegle są liczone zarówno współczynniki korelacji Pearsona jak i współczynniki korelacji Spearmana. W powszechnym użyciu są trzy typy korelacji: wg. Pearsona, Spearmana i Kendalla (ten ostatni dość rzadko) . Do zbadania siły liniowego związku używamy techniki korelacji Pearsona. Badając zależności nieliniowe, użyjemy korelacji Separmana. Autor używa obu. Dlaczego? To jest pytanie do Doktoranta o zasadności tego postępowania.

### **Podsumowanie oceny**

Zgodnie z ustawą o stopniach naukowych, ocena rozprawy doktorskiej winna zawierać ocenę:

1. stopnia oryginalności problemu naukowego,
2. ogólnej wiedzy teoretycznej kandydata do danej dyscypliny naukowej tu ekonomii,
3. umiejętności samodzielnego prowadzenia pracy naukowej przez kandydata.

Do ustosunkowanie się do tych zaleceń przeprowadziłem analizę jak te postulaty zostały spełnione w ocenianej pracy. W ocenie skoncentrowałem się szczególnie na stosowanej metodzie badawczej jak i poprawności uzyskanych wyników. Całość pracy oceniam pozytywnie. Praca dotyczy ważnego problemu teoretycznego i praktycznego jakim jest estymacja pracochłonności i czasu realizacji projektów informatycznych. Zastosowana metoda estymacji jest oryginalna. Pozytywnie również oceniam poziom wiedzy teoretycznej kandydata. Praca pod względem stosowanych narzędzi oraz obszaru zastosowań dotyczy nauk ekonomicznych. Pan mgr Przemysław Pospieszny w swojej rozprawie pt. Zastosowanie technik eksploracji danych do estymacji pracochłonności i czasu trwania projektów informatycznych, wykazał się umiejętności samodzielnego prowadzenia pracy naukowej. Moje uwagi mają na celu pogłębienie prowadzonej przez Doktoranta analizy problemu jak też pokazania dyskusyjnego charakteru niektórych przyjętych w pracy rozwiązań.

Uważam iż recenzowana praca spełniła wymagania Ustawy o stopniach naukowych i tytule naukowym z 14.03.2003 z późniejszymi zmianami. Moja opinia o recenzowanej rozprawie, jak już wcześniej zaznaczyłem, jest pozytywna.

Rekomenduję Radzie Kolegium Analiz Ekonomicznych, Szkoły Głównej Handlowej w Warszawie dopuszczenie Pana mgr. Przemysława Pospiesznego do następnych etapów procedury uzyskania stopnia doktora.

A handwritten signature in blue ink, appearing to read 'Przemysław Pospieszny', written in a cursive style.