

Lublin, 28 styczeń 2024 r.

Dr hab. inż. Łukasz Skowron
profesor Politechniki Lubelskiej
Słonecznikowa 5/2
20-827 Lublin

Recenzja pracy doktorskiej

mgr Grzegorza Miguta pt.
Identyfikacja optymalnej ścieżki budowy modeli *data mining* w obszarze retencji klientów

Podstawę niniejszej oceny stanowi decyzja Uchwały Rady Dyscypliny Naukowej Nauki o Zarządzaniu i Jakości Uniwersytetu Ekonomicznego w Krakowie nr RDC.600.20P.2.20233 z dnia 30 października 2023 roku w sprawie wyznaczenia recenzenta pracy doktorskiej Pana mgr Grzegorza Miguta oraz skierowane do mnie stosowne pismo w przedmiotowej sprawie.

Recenzja obejmuje:

- określenie strony formalnej i konstrukcji pracy,
- uwagi i ocenę natury merytorycznej,
- wniosek końcowy.

1. Strona formalna i konstrukcja pracy

Praca objętości 251 stron składa się ze wstępu, 5 rozdziałów, zakończenia, bibliografii, spisu tabel, rysunków oraz załącznika zawierającego opis zmiennych wykorzystanych w modelowaniu. W pracy Autor odwołał się do 172 pozycji literaturowych oraz zaprezentował 93 rysunki i 39 tabel.

Praca w swej konstrukcji nosi wszelkie znamiona dysertacji naukowej. Zawiera elementy metrykalne i metodyczne, zapisane we wprowadzeniu, w tym: precyzyjny opis problemu badawczego i przekonujące uzasadnienie jego podjęcia oraz cel główny badań w dekompozycji na 6 celów pobocznych. W części metodycznej brakuje jasno i precyzyjnie postawionej hipotezy głównej oraz hipotez cząstkowych, co jednocześnie ze względu na eksploracyjny charakter prowadzonych rozważań empirycznych (badania modelowe prowadzone z wykorzystaniem różnych technik i metod analitycznych) nie umniejsza jakości prowadzonego dyskursu naukowego. Metodyka badań została nakreślona precyzyjnie, w postaci dobrze opisanego i osadzonego w przestrzeni empirycznej procesu identyfikacji optymalnej ścieżki selekcji oraz

budowy modelu migracji klientów. Wszystkie te elementy oceniam pozytywnie w wymiarze merytorycznym.

Jedynym mankamentem pracy w jej warstwie formalnej i konstrukcyjnej jest brak zadbania przez Autora o to aby każdy rozdział rozpoczynał tekst wprowadzający Czytelnika w istotę podjętych w nim treści oraz aby kończył się podsumowaniem, które powinno zawierać syntezę treści rozdziału, wnioski z prowadzonych badań i analiz oraz refleksję natury ogólnej. Taki przyjęty układ pracy wzmacniałby spójny obraz prowadzonego dyskursu naukowego w obrębie poszczególnych jej rozdziałów, a te precyzyjnie osadzał w strukturze całego procesu badawczego. Niemniej jednak solidna strona edycyjna opracowania oraz widoczna dbałość Autora o formę i styl pisowni stanowią pozytywny element pracy i świadczą o dużym doświadczeniu Doktoranta w zakresie uprawiania piśmiennictwa naukowego.

2. Ocena merytoryczna

Rozprawa doktorska Pana Grzegorza Miguta jest swoistą próbą zmierzenia się z bardzo wyrafinowanym i wieloaspektowym zagadnieniem identyfikacji optymalnej ścieżki budowy modeli retencji klientów. Jak podkreśla sam Autor pracy poruszana problematyka z roku na rok nabiera coraz większego znaczenia ze względu na ciągły i wielce dynamiczny wzrost ilości danych generowanych przez wszystkie przedsiębiorstwa bez znaczenia na branże oraz sektory w jakich działają. Duże strumienie danych wymagają od kadr zarządzających przedsiębiorstw poszerzania swoich umiejętności w obszarze stosowanych metod modelowania zjawisk rynkowych z wykorzystaniem modeli ekonometrycznych oraz modeli uczenia maszynowego. Należy zaznaczyć, iż wraz ze wzrostem możliwości obliczeniowych komputerów rośnie również liczba możliwych do wykorzystania metod algorytmizacji i analizy zjawisk migracji i retencji klientów przy wykorzystaniu coraz bardziej zaawansowanych metod obliczeniowych. Głównym celem recenzowanej pracy było dokonanie swoistego przeglądu metod budowy oraz oceny jakości modeli retencji klientów, zwłaszcza w ujęciu wpływu wielu różnorodnych determinant jakości modelu na końcowy rezultat procesu modelowania. Autor w swojej dysertacji podjął się w ocenie recenzenta niezwykle trudnego zadania identyfikacji optymalnej strategii budowy modeli ukierunkowanej na uzyskanie przez badacza pożądanych rezultatów o wysokim poziomie aplikacyjności rynkowej.

We wstępie do pracy Autor w sposób poprawny przybliży główne założenia metodyczne dysertacji. Sposób prezentacji koncepcji badania dowodzi wysokich umiejętności Autora w obszarze przygotowania procesu badawczego i jego realizacji. Główne zastrzeżenie w mojej ocenie dotyczy braku postawienia przez Autora w pracy zarówno hipotez badawczych (brak jest

hipotezy głównej oraz hipotez cząstkowych) jak również pytań badawczych. Ujęte w pracy cele badawcze pokazują jedynie przyjęte kierunki rozważań naukowych Autora ale nie precyzują jego głównych wątpliwości naukowych, które w ramach realizowanej dysertacji powinny być wyjaśniane i rozwiewane. Takie spłylenie części metodycznej recenzowanej dysertacji utrudnia recenzentowi jasne oszacowanie „punktu wyjścia” zasobów wiedzy i założeń badawczych w układzie ex-ante, co w konsekwencji uniemożliwia ocenę poziomu zmiany myślenia doktora i jego krzywej nauki będącej efektem procesu realizacji badań pierwotnych oraz wtórnych w ramach omawianej pracy doktorskiej.

Bardzo pozytywnie należy ocenić część studialno-literaturową pracy, która w głównej mierze została ujęta w pierwszych trzech rozdziałach recenzowanej dysertacji. W pierwszym rozdziale Autor przedstawia bardzo ciekawe zestawienie koncepcji modelowania w badaniach marketingowych. Następnie treść rozdziału pierwszego przybliży czytelnikowi pojęcia oraz rolę: marketingu relacji w budowaniu postaw lojalnościowych klienta, definicji pojęcia lojalności klienta wraz z podziałem na jej rodzaje i poziomy oraz sposobów i metod jej pomiaru i modelowania. Zwłaszcza część dotycząca modelowania zjawiska lojalności zdaniem recenzenta zasługuje na szczególne uznanie. Autor w treści tego punktu ujmuje w sposób bardzo syntetyczny większość znanych z literatury przedmiotu i jednocześnie najbardziej popularnych metod i sposobów pomiaru i modelowania zjawiska lojalności klienta. W tym miejscu należy jednocześnie zaznaczyć, iż w ocenie recenzenta, omawiany fragment pracy powinien kończyć się pewnym rodzajem puenty zawierającej zagregowane zestawienie prezentowanych metod (w postaci tabelarycznej lub macierzowej), które powinno w swojej konstrukcji zawierać informację o obszarach spójnych i rozróżniających poszczególne omawiane koncepcje i struktury modelowe. Takie zestawienie, poprzez swoje nowatorskie podejście do syntezy omawianych koncepcji modelowych, stanowiłoby istotny wkład Autora w rozwój wiedzy poprzez konsolidację oceny poziomu podobieństwa i/lub odmienności poszczególnych koncepcji modelowania zjawisk lojalnościowych klientów. Na szczególną uwagę zdaniem recenzenta zasługuje również ostatni podrozdział ujęty w rozdziale pierwszym omawianej dysertacji, który został poświęcony tematyce „metodyki budowy modeli *data mining* na potrzeby retencji klientów”. Zarówno zakres prezentowanych w omawianej części dysertacji metod i koncepcji, jak również sposób i chronologia ich prezentacji w sposób jednoznaczny świadczą o głębokiej wiedzy i umiejętnościach Autora w obszarze omawianych zagadnień metodycznych. Jeśli chodzi o stronę formalną w rozdziale pierwszym recenzent dostrzegł jeden błąd polegający na

nieusunięciu z treści pracy komunikatu o „nieprawidłowym odsyłaczu do zakładki”, który znajduje się nad tabelą 3 na stronie 34.

W rozdziale drugim Autor koncentruje się na omówieniu zagadnień metod i technik przygotowania danych niezbędnych do wykorzystania w procesie budowy modeli retencji klientów. Zakres tematyczny omawianego rozdziału oraz prezentowane w nim koncepcje i metody pozwalają na bardzo wysoką ocenę wiedzy i potencjału merytorycznego Autora w obszarze: określania kluczowych parametrów procesu analitycznego; doboru metod walidacji i przygotowania danych do analiz (ocena ich wiarygodności, identyfikacja i imputacja braków); rozumienia konsekwencji i metod radzenia sobie z brakiem jednorodności w zbiorach danych oraz koncepcji transformacji i przygotowania zmiennych pochodnych. Ponadto Autor w rozdziale drugim poświęca sporo uwagi, często w sposób bezzasadny pomijanemu przez badaczy, problemowi niezbilansowanego rozkładu zmiennej zależnej i jego konsekwencjom zwłaszcza przy prognozowaniu zjawisk o charakterze rzadkim. Fakt, iż Autor poświęca szczególną uwagę omawianemu problemowi świadczy o jego bardzo wysokich umiejętnościach analitycznych i wysokim poziomie świadomości aplikacyjnej omawianych zagadnień do zastosowań rynkowych wraz z dbałością o rzetelność i celność wyciąganych wniosków natury zarządczo-kierowniczej. W ocenie recenzenta na szczególną uwagę, w treści rozdziału drugiego, zasługuje fragment poświęcony problematyce identyfikacji i imputacji braków w zbiorach danych. Autor w swoich rozważaniach przedstawia bardzo ciekawą i rzadko spotykaną w literaturze przedmiotu typologię klasyfikacji podejść do rozwiązania problemu imputacji braków danych. W tym miejscu należy również dodać, iż w opinii recenzenta, omawiany fragment pracy mógłby być dodatkowo wzmocniony o empiryczną analizę efektów stosowania różnych metod imputacji przedstawioną na realnych zbiorach danych. Ponadto prowadzone rozważania (zarówno natury teoretycznej jak również empirycznej) zdaniem recenzenta powinny zostać dodatkowo wzbogacone o bardziej szczegółowe omówienie najczęściej stosowanych w pragmatyce badawczej klasycznych metod imputacji braków (np. średnia, moda, mediana) wraz z ich oczywistymi ograniczeniami i negatywnymi skutkami (obecnie tym metodom i konsekwencjom ich stosowania poświęcono jedynie jeden krótki akapit). W treści tej części pracy brakuje również ustosunkowania się Autora do kwestii stosowania bardziej złożonych algorytmów imputacji danych jak na przykład algorytm EM, który, jak pokazuje praktyka badawczo-rynkowa, jest bardzo popularną metodą imputacji braków danych w badaniach społecznych realizowanych głównie na skalach całkowitych 1-5 lub 1-10.

Rozdział trzeci omawianej dysertacji w ocenie recenzenta jest jednym z najsilniejszych fragmentów analizowanej pracy. Autor przedstawia w nim między innymi bardzo wnikliwie merytoryczną procedurę wyboru zmiennych podczas budowy modelu pomiarowego oraz w sposób autorski prezentuje i opisuje przebieg procedury wyboru techniki modelowania jak również proces optymalizacji hiperparametrów. W poruszanych obszarach problemowych na szczególną uwagę w ocenie recenzenta zasługuje zwłaszcza fragment poświęcony analizie filtrów ukierunkowanych i nieukierunkowanych oraz opisana procedura wyboru techniki modelowania. Pomimo bardzo wysokiej oceny poziomu merytorycznego prowadzonego w rozdziale trzecim dyskursu naukowego recenzent dostrzega również, iż prowadzone rozważania i wykonana przez Autora praca pozwalałaby na zdecydowanie większą aplikacyjność środowiskową wniosków płynących ze zrealizowanych w pracy badań i analiz. Autor niestety, ale nie zakończył prowadzonych przez siebie rozważań w formie zbiorczego zestawienia, które mogłoby być swoistym podsumowaniem prowadzonych analiz w postaci syntetycznej prezentacji informacji różnicujących poszczególne opisane metody wyboru zmiennych oraz techniki modelowania. Ponadto brak jest w recenzowanej pracy podjęcia przez Autora próby opisu krokowego przebiegu algorytmu wyboru techniki modelowania, który mógłby zostać opisany na podstawie prowadzonych przez Autora rozważań, co w efekcie znacząco ułatwiłoby korzystanie z prezentowanych w dysertacji wniosków i doświadczeń badawczych, również przez mniej obeznanych w kwestiach analitycznych naukowców zajmujących się zagadnieniami analizy i badania procesów klienckich. Przedstawiony tu sposób prezentacji prowadzonych rozważań, choć cechuje się bardzo wysokim poziomem profesjonalizmu i biegłością Autora w doborze i wykorzystaniu różnych technik i metod obliczeniowych, jest mało przystępny i zrozumiały dla osób, które nie posiadają tak głębokiego warsztatu analityczno-statystycznego w obszarze metod i technik modelowania jaki prezentuje Autor dysertacji. W opinii recenzenta została utracona szansa na zbudowanie narzędzia o charakterze utylitarnym, które miałyby transdyscyplinarne pole zastosowania.

W treści rozdziału czwartego Autor podejmuje próbę omówienia zagadnienia walidacji i wdrażania modeli retencji klientów. W ramach prowadzonego dyskursu naukowego analizie i omówieniu zostają poddane cztery zasadnicze nurty rozważań, mianowicie: miary dobroci dopasowania modeli retencji klientów, strategie walidacji tychże modeli, metody określania optymalnego punktu odcięcia oraz zagadnienie wyboru klientów do strategii sprzedażowych i retencyjnych. Całość prowadzonych przez Autora rozważań w omawianym rozdziale dysertacji należy ocenić na bardzo wysokim poziomie merytorycznym i technicznym. Raz jeszcze

Autor udowadnia swoje wysokie umiejętności i wiedzę w obszarze obróbki statystycznej materiału empirycznego, tym razem uwidocznione w perspektywie poszukiwanych miar dopasowania i stabilności modeli oraz technik i miar ich walidacji. Prezentowane treści pokazują, że Autor nie tylko wie jak zastosować omawiane metody statystyczne i analityczne, ale również rozumie ich sens i kontekst aplikacyjno-rynkowy. Najlepszym tego potwierdzeniem jest poświęcenie sporej uwagi na kwestie budowania modeli *uplift* w modelowaniu strategii sprzedażowych i retencyjnych. Takie podejście pokazuje duże zrozumienie pragmatyki rynku w poszukiwaniu najlepszych i najbardziej efektywnych rozwiązań w kontekście realnych postaw rynkowych klientów, a nie koncentracji jedynie na czysto „laboratoryjnych” założeniach zachowań klienckich, które ze względu na swoją prostotę w doborze próby niejednokrotnie skutkuje problemami w aplikacji wyników prac o charakterze naukowym do realnych warunków i wymagań rynkowo-efektywnościowych. Ponadto w treści rozdziału czwartego zostało zaprezentowane syntetyczne zestawienie przeglądu miar jakości modelu dla macierzy błędnych klasyfikacji (tabela 9, strona 143). Tak kompleksowe a zarazem przystępne w formie i treści zestawienie ujmujące trudny, tudzież niezwykle istotny obszar wiedzy związany z oceną jakości budowanych modeli retencji klientów, należy widzieć jako istotny wkład Autora w rozwój wiedzy w obszarze poruszanej tematyki naukowej. Jest to zarazem w ocenie recenzenta największa wartość dodana omawianej dysertacji w piśmiennictwo naukowe poświęcone zagadnieniu modelowania zjawisk rynkowo-klienckich (recenzent dotychczas nie spotkał się w literaturze przedmiotu z tak kompleksowym a zarazem syntetycznym i klarownie zaprezentowanym zestawieniem stosowanych miar jakości modeli klienckich).

Ostatni rozdział omawianej dysertacji (rozdział piąty) stanowi w całości część empiryczną recenzowanej pracy. Autor prowadzone rozważania natury analitycznej koncentruje na zagadnieniach identyfikacji optymalnej ścieżki selekcji modelu migracji klientów, dzieląc je na sześć zasadniczych podrozdziałów. Na początku zostały określone cele prowadzonych analiz. Następnie Autor przedstawił w sposób szczegółowy zarówno charakterystykę wykorzystanych do analizy zbiorów danych jak również proces ich przygotowania do prowadzenia operacji statystycznych. I tak kolejno, dane zostały poddane ocenie poziomu ich jakości, nastąpiła segmentacja zbioru danych, dodatkowa kategoryzacja zawartych w zbiorze zmiennych o charakterze jakościowym, ich ostateczny podział na podzbiory, zdefiniowano i dokładnie opisano zmienne pochodne w zakładanym modelu pomiarowym, dokonano wstępnej selekcji zmiennych oraz przedstawiono proces budowy modeli za pomocą wybranych narzędzi analitycznych. Tak szczegółowy i metodycznie zaawansowany proces doboru, opisu i przygotowania danych do

prorowadzonych analiz potwierdza bardzo wysokie umiejętności i wiedzę Autora w obszarze przygotowania oraz realizacji procesu badawczo-analitycznego wraz z umiejętnościami praktycznymi realizacji poszczególnych etapów obliczeń natury statystycznej. Pomimo bardzo wysokiej oceny tej części rozdziału, w opinii recenzenta, Autor nie poinformował czytelnika w jaki sposób dokonano podziału i przypisania danych do poszczególnych podzbiorów: uczącego, testowego oraz walidacyjnego (na stronie 196 opisano jedynie wielkość poszczególnych frakcji w stosunku 70%/15%/15% bez precyzyjnego wskazania metod selekcji i podziału danych pomiędzy podzbiory – recenzent zakłada, że podział został dokonany w sposób losowy prosty). W dalszej kolejności Autor przedstawia cztery przykładowe klasy modeli retencji klienta opracowane na podstawie uprzednio przygotowanego zbioru danych, czyli modele: regresji logistycznej, drzew klasyfikacyjnych i regresyjnych CART, drzew wzmacnianych oraz sieci neuronowych. Zarówno sposób prezentacji poszczególnych metod i klas budowy modeli retencji jak również sam opis poszczególnych kroków w prowadzonych analizach oraz prezentacja uzyskanych wyników, pozwalają dostrzec w Autorze solidnego specjalistę w zakresie modelowania zjawisk rynkowych. Tak głębokie zrozumienie teorii i metodyki badań leżących u podstaw poszczególnych technik budowy modeli retencyjnych jest rzadko spotykanym zjawiskiem wśród osób zajmujących się w sposób praktyczny zagadnieniami modelowania zjawisk klienckich (jest to w zdecydowanej większości przypadków domena wiedzy osób wywodzących się wprost z nurtu nauk matematycznych lub ekonometrycznych). Autor w swoich rozważaniach nie tylko przedstawił w sposób klarowny wyniki prowadzonych analiz ale przeprowadził również dyskurs naukowy na temat wad i zalet każdej z omawianych metod wraz z ich ograniczeniami i płynącymi z nich limitacjami. Zdaniem recenzenta Autor mógł podjąć próbę przedstawienia przebiegu procesu budowy modelu pomiarowego, z zastosowaniem jednej z czterech omówionych w dysertacji metod, w postaci algorytmu krokowo-decyzyjnego, co umożliwiłoby skorzystanie z wyników tej pracy szerszemu gronu naukowców o mniejszym poziomie wiedzy i kompetencji z zakresu modelowania i analiz statystycznych. Niemniej jednak rozdział piąty recenzowanej dysertacji należy uznać za bardzo dobry i w pełni odpowiadający standardom metodyki prowadzenia badań naukowych w obszarze modelowania zjawisk retencji klientów.

W zakończeniu Doktorant rozlicza się z postawionego w dysertacji problemu i celów badawczych. W pracy wykazano osiągnięcie zarówno celu głównego jak również wszystkich celów szczegółowych. Recenzent zgadza się z Autorem pracy, iż omawiana dysertacja w zadowalającym stopniu wypełnia lukę badawczą w obszarze wieloaspektowej oceny jakości modeli

retencji. W ocenie ogólnej pracy wielce pozytywnie oceniam również fakt, że Doktorant zarówno w warstwie teoretyczno-studialnej, jak również w przedstawionych badaniach i analizach empirycznych jest w pełni świadomy ograniczeń płynących z poddania analizie tylko jednego zbioru danych. Recenzent popiera stanowisko Autora, iż dla pewności użytecznego charakteru uzyskanych wyników i płynących z nich wniosków natury aplikacyjnej, należy w przyszłości powtórzyć przyjętą procedurę badawczą również na innych zbiorach danych pochodzących z innych rynków i/lub sektorów.

3. Wniosek końcowy

Recenzowana praca prezentuje wysoki poziom naukowy i wnosi istotny wkład w rozwój wiedzy i nauki w obszarze badanego zagadnienia, to jest identyfikacji optymalnej ścieżki budowy modeli *data mining* w obszarze retencji klienta. Stwierdzam, że rozprawa doktorska Pana mgr Grzegorza Miguta spełnia wymogi ustawowe stawiane pracom doktorskim, określone w Ustawie z dnia 20 lipca 2018 roku *Prawo o szkolnictwie wyższym i nauce*.

W szczególności:

- stanowi samodzielne oryginalne rozwiązanie przez Doktoranta problemu naukowego,
- świadczy pozytywnie o poziomie ogólnej wiedzy teoretycznej Doktoranta w dyscyplinie nauki o zarządzaniu i jakości.
- dowodzi umiejętności samodzielnego prowadzenia przez Niego pracy naukowej.
- potwierdza wysokie umiejętności Autora w obszarze przygotowania i prowadzenia badań naukowych oraz analizy i interpretacji otrzymanych wyników empirycznych, wraz z ponadprzeciętnymi umiejętnościami modelowania zjawisk klienckich oraz wyciągania z otrzymanych wyników analiz modelowych wniosków natury rynkowej i aplikacyjnej.

Wnioskuje o dopuszczenie recenzowanej dysertacji do publicznej obrony.

Lukasz Skowron

dr hab. inż. Łukasz Skowron