

Prof. dr inż. Alicja Wakulicz-Deja  
Uniwersytet Śląski  
Wydział Nauk Ścisłych i Technicznych  
Instytut Informatyki

Dnia, 08.01.2024r.

**Recenzja osiągnięcia habilitacyjnego, dorobku naukowego, dydaktycznego  
i organizacyjnego dra Andrzeja Janusza w związku z postępowaniem o nadanie stopnia  
naukowego doktora habilitowanego w dziedzinie nauk inżynieryjno-technicznych,  
w dyscyplinie informatyka techniczna i telekomunikacja**

Niniejsza recenzja została przygotowana na zlecenie Dyrektora Instytutu Badań Systemowych Polskiej Akademii Nauk w związku z powołaniem Komisji Habilitacyjnej w postępowaniu w sprawie nadania stopnia doktora habilitowanego Panu dr Andrzejowi Januszowi.

Dr Andrzej Janusz jest adiunktem na Wydziale Matematyki, Informatyki i Mechaniki Uniwersytetu Warszawskiego. Habilitant uzyskał tytuł doktora nauk matematycznych w dziedzinie informatyki w 2014 roku, który został nadany przez Radę Wydziału Matematyki, Informatyki i Mechaniki Uniwersytetu Warszawskiego na podstawie rozprawy doktorskiej p.t. „Algorytmy uczenia się relacji podobieństwa z wielowymiarowych zbiorów danych”. Tytuł magistra matematyki Habilitant uzyskał w 2007 roku również na Uniwersytecie Warszawskim.

**Opis i ocena osiągnięcia naukowego**

Recenzja osiągnięcia naukowego p.t. „*Wybrane metody reprezentacji obiektów i pojęć w danych dla algorytmów uczenia maszynowego*” została wykonana na podstawie dokumentacji, w tym załączonego cyklu 11 prac opublikowanych w latach 2014-2023:

[O1] Andrzej Janusz, Sebastian Stawicki, Hung Son Nguyen: Adaptive Learning for Improving Semantic Tagging of Scientific Articles. in FedCSIS 2014: 27-34

[O2] Andrzej Janusz, Dominik Ślęzak: Rough Set Methods for Attribute Clustering and Selection. Appl. Artif. Intell. 28(3): 220-242 (2014)

- [O3] Andrzej Janusz, Dominik Ślęzak: Computation of Approximate Reducts with Dynamically Adjusted Approximation Threshold. ISMIS 2015: 19-28
- [O4] Andrzej Janusz, Marek Grzegorowski, Marcin Michalak, Łukasz Wróbel, Marek Sikora, Dominik Ślęzak: Predicting seismic events in coal mines based on underground sensor measurements. Eng. Appl. Artif. Intell. 64: 83-94 (2017)
- [O5] Sebastian Stawicki, Dominik Ślęzak, Andrzej Janusz, Sebastian Widz: Decision bireducts and decision reducts - a comparison. Int. J. Approx. Reason. 84: 75-109 (2017)
- [O6] Andrzej Janusz, Dominik Ślęzak: Investigating Similarity between Hearthstone Cards: Text Embeddings and Interchangeability Approaches. SMC 2018: 3421-3426
- [O7] Mateusz Przyborowski, Tomasz Tajmajer, Łukasz Grad, Andrzej Janusz, Piotr Biczuk, Dominik Ślęzak: Toward Machine Learning on Granulated Data - a Case of Compact Autoencoder-based Representations of Satellite Images. IEEE BigData 2018: 2657-2662
- [O8] Dominik Ślęzak, Marek Grzegorowski, Andrzej Janusz, Michał Kozielski, Sinh Hoa Nguyen, Marek Sikora, Sebastian Stawicki, Łukasz Wróbel: A framework for learning and embedding multi-sensor forecasting models into a decision support system: A case study of methane concentration in coal mines. Inf. Sci. 451-452: 112-133 (2018)
- [O9] Andrzej Janusz, Dominik Ślęzak: Analytics over Multi-sensor Time Series Data - A Case-Study on Prediction of Mining Hazards. Information Technology in Geo-Engineering. ICITG 2019: 815-830
- [O10] Andrzej Janusz, Daniel Kałuża, Maciej Matraszek, Łukasz Grad, Maciej Swiechowski, Dominik Ślęzak: Learning multimodal entity representations and their ensembles, with applications in a data-driven advisory framework for video game players. Inf. Sci., Volume 617: 193-210 (2022)
- [O11] Andrzej Janusz, Dominik Ślęzak, Sebastian Stawicki, Krzysztof Stencel: A Practical Study of Methods for Deriving Insightful Attribute Importance Rankings using Decision Bireducts. Inf. Sci., Volume 645, 119354, ISSN 0020-0255, <https://doi.org/10.1016/j.ins.2023.119354>. (2023)

Ocena punktowa cyklu przedstawionych publikacji wynosi 1170 punktów ministerialnych, co stanowi wysoki wskaźnik. Sześć prac cyklu [O2, O4, O5, O8, O10, O11] zostało opublikowanych w czasopiśmie z listy Journal Citation Reports (JCR). Na uwagę zasługuje fakt, że trzy prace z wyżej wymienionych [O8, O10, O11] zostały opublikowane w renomowanym czasopiśmie Information Sciences. Pozostałe prace [O1, O3, O6, O7, O9] zostały opublikowane w materiałach międzynarodowych konferencji. Dr Andrzej Janusz jest pierwszym autorem ośmiu publikacji z cyklu. Habilitant przy każdej z przedstawionych publikacji omawia swój merytoryczny wkład włożony w daną publikację. W większości prac

z cyklu wkład Habilitanta polegał na opracowaniu idei lub metody, implementacji zaproponowanych metod, zaprojektowaniu i przeprowadzeniu eksperymentów opisanych w pracy oraz pisaniu publikacji i nanoszeniu poprawek. Zatem wkład Habilitanta można uznać za znaczący. Do dokumentacji dołączono oświadczenia współautorów potwierdzające wkład Habilitanta. Prace te są powiązane tematycznie – zgodnie z art. 219 ust. 1 pkt 2 Ustawy z dnia 20 lipca 2018 r. – Prawo o szkolnictwie wyższym i nauce.

W autoreferacie dołączonym do dokumentacji przewodu habilitacyjnego zawarto wprowadzenie do tematyki przedstawionego cyklu publikacji. Przedmiotem badań Habilitanta była reprezentacja obiektów z wykorzystaniem wielowymiarowych wektorów cech w danych rzeczywistych. Rozważano na przykład zbiór dotyczący zgłoszeń oraz przebiegu akcji realizowanych przez Państwową Straż Pożarną czy przewidywania niebezpiecznych stężeń metanu i aktywności sejsmicznej rejestrowanej w polskich kopalniach węgla kamiennego. W swojej pracy Habilitant rozważał różne metody mające na celu poprawę wyników modeli predykcyjnych poprzez odpowiednią reprezentację obiektów. Wkład Habilitanta można podzielić na trzy główne części.

Pierwsza część pracy polegała na propozycji nowych i badaniu istniejących metod selekcji cech. Zastosowano w tym celu podejście polegające na redukcji wielowymiarowości wykorzystujące: reduktory aproksymacyjne, algorytm mRMR (minimum redundancy – maximum relevance) oraz bireduktory tablic decyzyjnych. W pracy [O3] Habilitant zaproponował algorytm obliczania reduktów aproksymacyjnych z dynamicznie określonym współczynnikiem aproksymacji (algorytm DAAR). Algorytm ten umożliwia znalezienie podzbioru atrybutów istotnych poprzez wykorzystanie permutacji wartości atrybutu do oceny znaczenia atrybutów. Nowością tego algorytmu jest dynamicznie określana wartość współczynnika aproksymacji dostosowana do zbioru danych oraz generowanie globalnych lub lokalnych permutacji testowanego atrybutu (co pozwala na rozważanie jedynie permutacji istotnych pod kątem wykrywania zależności między wartościami atrybutów). W pracy [O4] wykorzystano reduktory generowane z wykorzystaniem algorytmu DAAR do budowania zespołów modeli regresyjnych. Eksperymenty przeprowadzone na rzeczywistym zbiorze danych dotyczącym predykcji aktywności sejsmicznej rejestrowanej w polskich kopalniach węgla kamiennego wykazały wysoką skuteczność metody. W pracy [O8] wykorzystano również analogiczne podejście zespołu modeli regresyjnych do przewidywania niebezpiecznych stężeń metanu w kopalni węgla. Habilitant zaproponował również modyfikację algorytmu mRMR poprzez zamianę podejścia ze statycznie ustalonym parametrem określającym warunek stopu dla wyboru atrybutów istotnych na dynamiczne kryterium oparte na permutacyjnym teście statystycznym. W pracach Habilitanta badana była również skuteczność podejścia łączącego selekcję cech z algorytmami grupowania, głównie hierarchicznymi. W pracy [O2] dr Andrzej Janusz zaproponował ciekawą miarę podobieństwa atrybutów opartą na rozróżnialności obiektów. Z kolei w pracy [O9] Habilitant zaproponował miarę podobieństwa atrybutów opartą na współwystępowaniu atrybutów w reduktach

decyzyjnych. Te podejścia w połączeniu z algorytmem grupowania pozwoliły na wyznaczenie grup atrybutów, które mogą być stosowane zamiennie w reduktach decyzyjnych bez utraty jakości generowanych reguł decyzyjnych. Jak określa to Habilitant w ten sposób mierzony jest stopień wymienialności dwóch atrybutów. Jest to bardzo ciekawe podejście i ten wkład Habilitanta w rozwój dyscypliny informatyka techniczna uważam za znaczący, gdyż pozwala na radzenie sobie po pierwsze w przypadku gdy mamy mało obiektów, a znaczącą liczbę atrybutów w tablicy decyzyjnych, a po drugie w przypadku występowania braków w zbiorze danych. Habilitant miał również swój wkład w zaproponowanie efektywnego algorytmu do wyznaczania bireduktów decyzyjnych (praca [O5]). Powyżej opisany wkład Habilitanta uznaję za bardzo ważny. Jednak moim zdaniem należałoby szarzej porównać proponowane podejścia oparte na zespołach predykcyjnych z modelami zespołowymi znanymi z literatury. Również interesujące byłoby wykorzystanie różnych metod fuzji w połączeniu z modelami budowanymi z wykorzystaniem algorytmu DAAR.

Kolejno Habilitant skupił się na metodach ekstrakcji cech szeregów czasowych, których efektywność zweryfikował z wykorzystaniem rzeczywistych zbiorów danych złożonych z odczytów sensorów monitorujących bezpieczeństwo w kopalniach. W pracach [O4, O8, O9] dostosowano znany z literatury model Map-Reduce, którego zadaniem jest przekształcenie danych sensorycznych w wektory cech odzwierciedlające zależności pomiędzy wartościami różnych szeregów czasowych. Badania nad odpowiednią reprezentacją i ekstrakcją cech prowadzone były również w ramach pracy [O11], gdzie Habilitant zaproponował odpowiednią taksonomię i reprezentację ofert pracy oraz kandydatów w systemie wspierającym rekruterów w wyszukiwaniu potencjalnych kandydatów dla ofert pracy.

Trzeci obszar prac Habilitanta obejmował modelowanie zanurzeń w przestrzeni wektorowej – definiowanie przekształceń zbioru uniwersum w  $k$ -wymiarową przestrzeń nad zbiorem liczb rzeczywistych, które zachowuje relacje podobieństwa między obiektami. Pierwsze prace nad definiowaniem takich zanurzeń zawarte zostały w publikacji [O1] i polegały na zaproponowaniu modyfikacji znanej z literatury metody reprezentacji tekstów: ESA (explicit semantic analysis). Tutaj Habilitant zaproponował metodę dynamicznego dostosowywania parametrów modelu do zbioru danych. Następnie w celu modelowania zanurzeń zbioru obiektów wykorzystano sieci neuronowe. W pracy [O7] zaproponowano architekturę sieci neuronowych dla kompresji zdjęć satelitarnych, która znacząco zmniejszyła zajętość danych przy zachowaniu praktycznie niezmięnionej jakości klasyfikacji modelu. Ostatecznie Habilitant wykorzystał metody zaczerpnięte z dziedziny przetwarzania języka naturalnego (m.in. technikę word2vec) w celu wyznaczenia reprezentacji obiektów różnego typu tj.: ofert specjalistów z firmy rekrutacyjnej [O11] czy reprezentacji kart i talii z gry Hearthstone [O6]. W pracy [O6] wykorzystano również podejście oparte na sieciach neuronowych do uczenia się zanurzeń dla talii kart. Tak wyznaczone reprezentacje obiektów różnego typu zostały następnie wykorzystane przy tworzeniu modeli predykcyjnych. Wykazano, że modele te generują predykcje o wysokiej jakości. W pracy [O10] zastosowano nową architekturę sieci

neuronowych do uczenia się zanurzeń obiektów oraz wykorzystano dane pochodzące z gier wideo. Ostatecznie na podstawie przeprowadzonych eksperymentów dla różnych typów obiektów i zbiorów danych Habilitant potwierdził, że wykorzystując jednocześnie wiele różnych metod do reprezentacji cech obiektów i budując zespoły predykcyjne osiągnięta jest lepsza jakość modelu. Myślę, że ten wkład Habilitanta również jest istotny, niemniej jednak prowadzone badania w moim odczuciu należałoby potwierdzić na innych i różnorodnych zbiorach danych. Należałoby również uwzględnić różne problemy występujące w zbiorach danych takie jak niepewność czy braki danych.

Oceniane osiągnięcie naukowe dotyczy w ogólności reprezentacji obiektów i pojęć z wykorzystaniem czy to zbioru atrybutów istotnych czy też pewnych charakterystyk przedstawionych w postaci wielowymiarowych wektorów. W swojej pracy dr Andrzej Janusz wykorzystywał dane rzeczywiste dotyczące przewidywania zagrożeń w kopalni węgla kamiennego, wspierania rekrutów w dobieraniu ofert pracy do specjalistów w danej dziedzinie czy też danych wideo, zbioru zdjęć satelitarnych oraz reprezentacji talii kart do gry. Przedstawiony zbiór publikacji stanowi spójny materiał. Habilitant w swojej pracy wykazał użyteczność proponowanych rozwiązań.

W moim przekonaniu najistotniejszy wkład dra Andrzeja Janusza polega na zaproponowaniu modelu selekcji cech opartego na adaptacyjnym ustalaniu progu aproksymacji w reduktach aproksymacyjnych (algorytm DAAR) oraz wykorzystaniu tego podejścia do konstrukcji efektywnych zespołów predykcyjnych. Za kolejny bardzo istotny wkład Habilitanta uważam zaproponowanie dwóch różnych miar oceny atrybutów oraz połączenie sposobu oceny atrybutów z algorytmem grupowania hierarchicznego. W ten sposób wyznaczane są grupy atrybutów, które mogą być wykorzystywane zamiennie podczas generowania reguł decyzyjnych bez znaczącej utraty jakości tych reguł. Użyteczność tych podejść jest istotna w rzeczywistych zastosowaniach. Potwierdza to również znacząca liczba cytowań publikacji, w których zostały zaproponowane te podejścia jak również duże zainteresowanie ogólnodostępnym kodem źródłowym proponowanych metod.

Podsumowując uważam, że Habilitant uzyskał interesujące wyniki oraz wniósł istotny wkład w rozwój dziedziny nauk inżynierjno-technicznych, w dyscyplinie informatyka techniczna i telekomunikacja. Moja ocena dokonania habilitacyjnego dra Andrzeja Janusza jest pozytywna. Dokonanie to spełnia w moim przekonaniu warunki stawiane w przewodzie habilitacyjnym przez obowiązujące przepisy.

## Ocena dorobku naukowego i dydaktycznego

Działalność naukowa dra Andrzeja Janusza koncentrowała się na reprezentacji różnego typu obiektów i pojęć w postaci zbioru istotnych atrybutów lub w postaci wektorów z  $k$ -wymiarowej przestrzeni liczb rzeczywistych. Dorobek publikacyjny Habilitanta w dniu złożenia wniosku o wszczęcie postępowania habilitacyjnego obejmował łącznie 45 publikacji (opublikowanych po uzyskaniu tytułu doktora), w tym cykl 11 prac stanowiących wskazane osiągnięcie naukowe. Prace tego cyklu wydają się reprezentatywne dla wszystkich publikacji powstałych po uzyskaniu stopnia doktora przez Habilitanta. Należy też wspomnieć, że Habilitant organizował liczne międzynarodowe konkursy eksploracji danych. Łącznie wykazał siedemnaście konkursów organizowanych w ramach znanych międzynarodowych konferencji, dodatkowo organizował liczne konkursy pomniejsze. Wiele publikacji Habilitanta dotyczy właśnie wyników uzyskanych z wykorzystaniem zbiorów danych konkursowych. Ta aktywność wpłynęła również pozytywnie na współpracę międzynarodową oraz współpracę z otoczeniem Habilitanta.

W dniu złożenia wniosku o wszczęcie postępowania habilitacyjnego liczba cytowań dra Andrzeja Janusza wynosiła 1197 wg Google Scholar, 891 wg Scopus oraz 530 wg Web of Science; indeks Hirscha wynosił 18 wg Google Scholar, 17 wg Scopus oraz 12 wg Web of Science. Wskaźnik cytowań uważam za bardzo dobry.

Habilitant brał udział w komitetach organizacyjnych oraz w komitetach programowych siedmiu konferencji międzynarodowych oraz organizował sesje specjalne na międzynarodowych konferencjach. Po uzyskaniu stopnia doktora Habilitant brał aktywny udział w jedenastu konferencjach międzynarodowych, na których wygłosił referaty.

Dr Andrzej Janusz był kierownikiem naukowym dwóch projektów „SENSEI: System wspomagający rozwój umiejętności graczy” oraz „BrightBox: Narzędzia klasy explainable AI służącego do poprawy interpretowalności i przewidywalności działania metod uczących oraz diagnostyki poprawności działania wyuczonych modeli AI/ML”. Projekty te były finansowane przez NCBR. Ponadto Habilitant bierze udział jako wykonawca badań w pięciu innych projektach.

Habilitant ukończył staż naukowy na University of Halifax w Kanadzie gdzie pracował w grupie badawczej Prof. Stana Matwina nad metodami uczenia się reprezentacji obiektów w zbiorach danych. Dr Andrzej Janusz aktywnie współpracuje z otoczeniem społecznym i gospodarczym oraz angażuje się we współpracę z firmami przy realizacji projektów badawczo-rozwojowych. W trakcie swojej kariery zawodowej nawiązał współpracę z renomowanymi przedsiębiorstwami, takimi jak AdgaM Solutions, Zoined Oy i Silver Bullet Labs oraz QED Software.

W ramach działalności dydaktycznej Habilitant prowadzi zajęcia na Wydziale Matematyki, Informatyki i Mechaniki Uniwersytetu Warszawskiego. Ponadto Habilitant był promotorem

jedenastu prac dyplomowych oraz pełnił rolę promotora pomocniczego jednej rozprawy doktorskiej, która została obroniona. Obecnie Habilitant pełni rolę promotora pomocniczego dwóch doktorantów z Szkoły Doktorskiej Nauk Ścisłych i Przyrodniczych Uniwersytetu Warszawskiego.

W mojej opinii dorobek naukowy Habilitanta powstały po uzyskaniu stopnia doktora i jego aktywność naukowa oraz działalność dydaktyczna i popularyzatorska spełniają wymagania ustawowe stawiane doktorowi habilitowanemu.

### **Konkluzja**

W mojej ocenie osiągnięcie habilitacyjne, dorobek naukowy, dydaktyczny i organizacyjny dra Andrzeja Janusza, spełniają wymagania Ustawy z dnia 20 lipca 2018 r, Prawo o szkolnictwie wyższym i nauce odnośnie nadania stopnia naukowego doktora habilitacyjnego. Całość oceniam pozytywnie. Uzyskane wyniki są wartościowe i oryginalne, stanowią istotny wkład w rozwój dyscypliny informatyka techniczna i telekomunikacja, a w szczególności w rozwój metod reprezentacji różnorodnych obiektów i pojęć.

Wniosuję o dopuszczenie dra Andrzeja Janusza do dalszych etapów postępowania habilitacyjnego w celu nadania stopnia naukowego doktora habilitowanego nauk inżyniersko-technicznych w dyscyplinie informatyka techniczna i telekomunikacja.

.....

Alicja Wakulicz-Deja