

Lublin, 19.07.2024

dr hab. Leszek Gawrysiak, prof. UMCS
Katedra Geologii, Gleboznawstwa i Geoinformacji
Instytut Nauk o Ziemi i Środowisku
Wydział Nauk o Ziemi i Gospodarki Przestrzennej
Uniwersytet Marii Curie-Skłodowskiej w Lublinie

Recenzja rozprawy doktorskiej
magistra Krzysztofa Dyby

Interpretation of geospatial data using explainable machine learning methods

Interpretacja danych geoprzestrzennych przy użyciu wyjaśnialnych metod uczenia maszynowego

przygotowanej w Szkole Doktorskiej Nauk Przyrodniczych

na Wydziale Nauk Geograficznych i Geologicznych

Uniwersytetu Adama Mickiewicza w Poznaniu

Promotor prof. UAM dr hab. Jarosław Jasiewicz

Promotor pomocniczy prof. UAM dr hab. inż. Cezary Kaźmierowski

I. Uwagi ogólne

Przedstawiona recenzja dotyczy manuskryptu rozprawy doktorskiej Pana mgr Krzysztofa Dyby w formie oprawionego, jednostronnego wydruku komputerowego, którego integralną część stanowią trzy załączniki w postaci kopii trzech opublikowanych prac naukowych. Autoreferat napisany jest w języku angielskim i zawiera 28 stron, na które składają się: strona tytułowa, podziękowania, spis treści, wprowadzenie, cel, materiały i metody, podsumowanie artykułów, dyskusja, wnioski oraz spis literatury. Po spisie publikacji znajdują się załączniki, w których wymieniono osiągnięcia akademickie: uczestnictwo w konferencjach, udział w grantach, nauczanie i inne osiągnięcia.

Wykaz literatury zawiera 67 anglojęzycznych pozycji, wszystkie cytowane są w tekście autoreferatu a także wszystkie cytowane prace znajdują się w spisie.

Do autoreferatu dołączone jest streszczenie w języku polskim liczące 1 stronę, w którym bardzo zwięźle opisany jest sposób realizacji celu głównego i celów szczegółowych oraz wnioski, bez rycin, tabel i odwołań do literatury.

Pod względem edytorskim tekst autoreferatu przygotowany jest rzetelnie.

Należy tutaj podkreślić, że autoreferat jest jedynie komentarzem/przewodnikiem do dysertacji, na którą składają się artykuły naukowe i to one podlegają recenzji jako całość.

Po spisie literatury znajdują się załączniki (3) zawierające kopie współautorskich artykułów opublikowanych w czasopismach: Remote Sensing, Geomorphology i Scientific Reports.

Załącznik 1:

Dyba, K., Ermida, S., Ptak, M., Piekarczyk, J., & Sojka, M. (2022). Evaluation of methods for estimating lake surface water temperature using Landsat 8. Remote Sensing, 14(15), 3839.
<https://doi.org/10.3390/rs14153839>

Autorzy udostępniłi w internecie, na stronie wydawnictwa MDPI, materiały uzupełniające do tego artykułu.

Jest to najobszerniejszy z artykułów, liczy 21 stron, zilustrowany jest 11-ma rycinami i zawiera 2 tabele. Literatura zawiera 86 pozycji.

Załącznik 2:

Dyba, K., & Jasiewicz, J. (2022). Toward geomorphometry of plains - Country-level unsupervised classification of low-relief areas (Poland). Geomorphology, 413, 108373.
<https://doi.org/10.1016/j.geomorph.2022.108373>

Autorzy udostępniłi w internecie, na stronach wydawnictwa MDPI, czasopisma Geomorphology oraz serwisu Zenodo, materiały uzupełniające do tego artykułu.

Artykuł liczy 18 stron, zilustrowany jest 12-ma rycinami 2-ma tabelami. Literatura liczy 98 pozycji.

Załącznik 3

Dyba, K. (2024). Explanation of the influence of geomorphometric variables on the landform classification based on selected areas in Poland. Scientific Reports, 14(1), 5447.
<https://doi.org/10.1038/s41598-024-56066-6>

Autor udostępnił w internecie, na stronie wydawnictwa czasopisma materiały uzupełniające do tego artykułu.

Artykuł liczy 13 stron, zilustrowany jest 8-ma rycinami i zawiera 3 tabele. Literatura zawiera 62 pozycje.

W autoreferacie nie ma deklaracji wkładu własnego w opublikowanych, współautorskich pracach naukowych złożonych w ramach pracy doktorskiej.

Wszystkie artykuły zostały opublikowane w 3 renomowanych czasopismach naukowych z listy Journal Citation Report (JCR) indeksowanych w bazie Web of Science (WoS) i mających wskaźniki wpływu IF (za 5 lat) w zakresie od 3.6 do 4.9. Czasopisma lokują się w pierwszym lub drugim kwartyle rankingu Journal Impact Factor. Wspomniane czasopisma znajdują się na „liście A” rankingu czasopism naukowych MNiSW z 2024 roku i punktowane są w zakresie 100-140 pkt. Według Web of Science czasopisma te należą do następujących kategorii: geosciences, multidisciplinary geosciences oraz multidisciplinary sciences. Według listy MNiSW wszystkie należą do dziedziny nauk o Ziemi i środowisku. Sumaryczny Impact Factor wszystkich publikacji wynosi 12,8, natomiast łączna suma punktów (wg list MNiSW) wynosi 340. Dwie prace (1 i 2), wchodzące w skład dysertacji, dotychczas zostały zacytowane łącznie 8 razy (04.07.2024), większą liczbę cytowań (6) ma praca pierwsza.

Dwa z przedłożonych artykułów są rezultatem prac zespołowych, prowadzonych w zespole dwu- (2 prace) lub pięcioosobowym (praca 1). Praca trzecia jest samodzielna. W pracach 1 i 2 Doktorant jest pierwszym autorem. Według oświadczeń Autorów rola Doktoranta obejmowała w pierwszej pracy: metodologię, oprogramowanie, walidację, analizy formalne, zarządzanie danymi, opracowanie tekstu, przegląd, edycję i wizualizację. W pracy drugiej Doktorant odpowiadał za koncepcję, dane, analizy formalne, badania, oprogramowanie, walidację, pisanie tekstu, wizualizację i pozyskanie funduszy.

II. Opis i merytoryczna ocena rozprawy

Recenzowana dysertacja pt. „Interpretation of geospatial data using explainable machine learning methods” wpisuje się w nowy nurt badań geograficznych, z wykorzystaniem technik uczenia maszynowego. Można nawet pokusić się o stwierdzenie, że Doktorant wkracza na obszary dotychczas mało eksplorowane.

Jak stwierdza Doktorant w pierwszym zdaniu Autoreferatu (s.1) *„In scientific research, the transparency of the data processing is crucial, both for the interpretation of the results and the comparability of the results achieved using different methods.”*. Co można przetłumaczyć „W badaniach naukowych przejrzystość przetwarzania danych ma kluczowe znaczenie zarówno dla interpretacji wyników, jak i porównywalności wyników uzyskanych przy użyciu różnych metod”. Jest to niezwykle istotne, wręcz kluczowe stwierdzenie wskazujące na warunek konieczny poprawnego posługiwania się narzędziami (geo)informatycznymi, jakim jest precyzyjny opis użytej procedury na każdym etapie postępowania badawczego. Dzięki temu, jak pisze Doktorant, możliwe jest powtórzenie procesu jak i jego krytyczna analiza.

Jednym z zadań recenzenta jest przyjrzenie się, czy Doktorant konsekwentnie stosuje się do tej zasady w pracach, wchodzących w skład rozprawy doktorskiej.

W części wstępnej autoreferatu Doktorant zwraca uwagę na gwałtowny rozwój technologii geoinformatycznych i powiązany z nim przyrost dostępnych danych przestrzennych. Wskazuje tu na technologie zdalne wykorzystujące sensory różnego typu. Doprowadziło to do nowej sytuacji, w której dotychczasowe metody eksperckie są niewystarczające i konieczności sięgania po metody uczenia maszynowego do analizy danych przestrzennych.

W dalszej części objaśnia czym jest uczenie maszynowe, wskazując, że jego istotą jest tworzenie modeli (narzędzi) do efektywne rozpoznawanie wzorów i zależności w nowych (niezależnych) zestawach danych. Biorąc pod uwagę naturę procesu uczenia wyróżnia uczenie nadzorowane i nienadzorowane. Tym co je odróżnia jest obecność zmiennej zależnej (referencyjnej) w klasyfikacji nadzorowanej i zestawu zmiennych niezależnych (obiektów) w kl. nienadzorowanej i brak założeń co do spodziewanego wyniku. Zwykle ten drugi proces klasyfikacji kończy się interpretacją i nadaniem nazw wydzielonym zbiorom. W drugim wymienionym podziale metod badawczych wymienia czarne i białe skrzynki. Do czarnych zalicza sieci neuronowe, losowy las (random forest) i wektory nośne (support vector machines). Mają one wysoką dokładność predykcji, ale nie jest jawne metoda ich działania. Obecnie w praktyce badawczej obserwuje się wzrost zainteresowania ich użyciem. Z kolei modele typu białe skrzynki (regresja liniowa, regresja logistyczna, drzewa decyzyjne i ogólne modele liniowe) mają mniejszą dokładność, ale mechanizmy ich działania są znane. Dalej wskazuje, że obecnie luką badawczą jest brak interpretacji wpływu zmiennych objaśniających na wyniki modelowania danych przestrzennych z wykorzystaniem uczenia maszynowego, brak oceny skuteczności takich rozwiązań oraz brak oceny ich kompatybilności z eksperckimi schematami postępowania. Podkreśla, że wynik

modelowania powinien być powtarzalny, zrozumiały i użyteczny dla potencjalnych użytkowników poprzez wsparcie procesu interpretacji danych przestrzennych.

Dalej Doktorant zaznacza, że pomimo dużej liczby publikacji, w których wykorzystano uczenie maszynowe, w teledetekcji i innych dziedzinach, podejście interpretacyjne nie było szeroko stosowane, na co wielokrotnie zwracano uwagę. Podejmowano takie próby w naukach o Ziemi i środowisku. Istnieje zatem luka, która zachęciła Doktoranta do podjęcia badań w tym zakresie.

Doktorant jako cel główny definiuje **wykazanie, czy wyjaśnialne metody uczenia maszynowego mogą być narzędziem wspierającym proces modelowania danych geoprzestrzennych, gdzie miarą sukcesu nie jest wyszkolenie skutecznego modelu, ale zdolność do uzyskania dodatkowej wiedzy na temat czynników prowadzących do wyniku**. Osiągnięcie tego celu realizuje za pomocą trzech eksperymentów badawczych, obejmujących różne metody uczenia maszynowego, w tym analizy regresji, klasyfikacji nadzorowanej i nienadzorowanej i wybranych objaśnialnych metod uczenia maszynowego dla globalnego modelu interpretacji.

Takie sformułowanie celu głównego wydaje się zasadne a dobór problemów badawczych stwarza możliwość przetestowania szeregu metod i dostarcza materiał do próby odpowiedzi na główne pytanie badawcze.

Omówienie prac wchodzących w skład rozprawy doktorskiej.

Praca pierwsza *Evaluation of methods for estimating lake surface water temperature using Landsat 8* (Dyba et al. 2022) dotyczy modelowania temperatury powierzchni jezior. Wykorzystano w niej ogólnodostępne dane pochodzące z monitoring 38 jezior, prowadzonego przez IMGW w okresie 2013-2020 oraz obrazy satelitarne Landsat 8 z NASA i USG Survey. Jako główny cel badawczy Autorzy stawiają porównanie 4 metod ustalania temperatury wód w jeziorach w Polsce w oparciu o zobrażenia termalne. Celem szczegółowym jest identyfikacja ograniczeń zastosowanych metod oraz analiza przestrzennego zróżnicowania temperatury w obrębie wybranych zbiorników.

Do predykcji temperatury jezior wykorzystano dwa różne modele regresji – prostą regresję wielokrotną (LM) i model regresji oparty na losowym lesie (RF - random forest). Jako zmiennych objaśniających w metodzie LM użyto wartości z 2 kanałów termalnych zobrażeń lub 12 parametrów w metodzie RF, a modelowaną zmienną jest temperatura. Pozostałe dwie metody (LST i LST-L2) dostarczają gotowe produkty będące wynikiem analizy rozkładu wartości kanałów termalnych zobrażeń satelitarnych.

Cel główny pracy został osiągnięty, przetestowano i porównano wszystkie metody a wyniki pozwalają stwierdzić, że metoda RF daje wynik najbardziej zbliżony do referencyjnego. Druga pod względem dokładności jest metoda LM. Stwierdzono również, że wartości uzyskiwane metodą LST-L2 obarczone są stałym błędem, w związku z czym rekomenduje się kalibrację tych danych przed wykorzystaniem w badaniach związanych z hydrologią czy ekologią jezior. Przestrzenne zróżnicowanie temperatury w wybranych zbiornikach przeanalizowano i zaprezentowano na przykładzie 2 jezior o odmiennej cyrkulacji. Stwierdzono, że kluczową rolę odgrywają tu parametry morfologiczne misy jeziora bądź procesy hydrologiczne.

Jak stwierdzają Autorzy jest to pierwsze opracowanie, w którym wykorzystano satelitarne dane termalne do modelowania i monitoringu temperatury jezior w Polsce.

UWAGI:

Jezióra do próby treningowej i sprawdzającej zostały dobrane przypadkowo. Czy sprawdzono czy te jeziora miały podobne parametry warunkujące rozkład temperatury? Jest o tym wspomniane w części 3.1 (s.12), ale odnosi się to do rozkładu wartości RMSE wyników modelowania, brak jest porównania pod tym względem obydwu grup zbiorów. Na Fig. 8 pokazano ich rozmieszczenie i jeziora z próby sprawdzającej nie są rozmieszczone równomiernie, w większości położone są w części NE.

Celem drugiej pracy ***Toward geomorphometry of plains - Country-level unsupervised classification of low-relief areas (Poland)*** (Dyba, Jasiewicz 2022) było rozpoznanie typów powierzchni (terenu) na obszarze Polski z wykorzystaniem klasyfikacji nienadzorowanej, z naciskiem na obszary nizinne stanowiące przeszło 80% powierzchni kraju. Analizy grupowania wykonano w oparciu o przetworzony Numeryczny Model Terenu DTED2 i jego 9 pochodnych, z wykorzystaniem modelu mieszanin rozkładów Gaussa (GMM - *Gaussian Mixture Model*). Uzyskane klastry zostały następnie zaetykietowane i pogrupowane hierarchicznie w oparciu o wyniki analiz głównych składowych (PCA – *principal components analysis*) i rozkłady wartości zmiennych geomorfometrycznych. Ich efekt zaprezentowano na schemacie oraz szeregu map przedstawiających sklasyfikowane typy i subtypy terenu (*land surface types and subtypes*), w kilku wariantach liczebności klastrów. Autorzy dużo uwagi poświęcili równinom, które, jak zaznaczają, dotychczas nie były wystarczająco różnicowane. Spośród ostatecznie wydzielonych 20 typów powierzchni aż 15 stanowią różnego typu równiny, co pozwala bardzo szczegółowo sklasyfikować ich morfologię. Zwrócono również uwagę na niepewność klasyfikacji poszczególnych typów i w związku z tym występowanie stref przejściowych pomiędzy wydzielonymi jednostkami.

UWAGI:

Tytuł pracy sugeruje, że analizy dotyczą obszarów nizinnych, jednak jest to opracowanie obejmujące obszar całego kraju. Należy to uznać za zaletę opracowania gdyż dają możliwość porównania obszarów o różnym typie rzeźby.

Termin DEM (*digital elevation model*) używany jest często w literaturze w odniesieniu do wysokości terenu. Autorzy stosują go właśnie w taki sposób, podczas gdy w sensie dosłownym jest to cyfrowy model wysokości, bez podania wysokości czego. Zgodnie ze specyfikacją INSPIRE (*Data Specification on Elevation – Technical Guidelines*, 2013, s. 26) termin cyfrowy/numeryczny model terenu (Digital Terrain Model – DTM) należy stosować w odniesieniu do wysokości nagiej powierzchni terenu (*bare surface*), natomiast termin Digital Elevation Model (DEM) odnosi się do dwóch typów danych – wspomnianego DTM (w nomenklaturze GUGiK to numeryczny model terenu, NMT) oraz DSM (*Digital Surface Model* – numeryczny model pokrycia terenu, NMPT wg GUGiK).

Rycina 1B przedstawia hipsometrię obszaru Polski a nie rzeźbę. W pracy trzeciej taka mapa podpisana jest jako *elevation map*.

W pracy nie wyjaśniono w jaki sposób wytypowano reprezentatywną próbkę z całego zbioru (Rozdział 2.4, s.5) by na niej wykonać modelowanie GMM. A jest kluczowe dla przeprowadzonych analiz.

W Tabeli 2 suma procent w kolumnie *Coverage* wynosi 100.1

Rycina 9 byłaby kompletna gdyby uwzględniono na niej przykład z gór, skoro jest przykład z wyżyn. Podobnie Rycina 12.

Poprawna nazwa wskaźnika TPI to Topographic Position Index (Weiss 2001) a nie Terrain Position Index (s.12). Rycina 11A nie przedstawia rozkładu wartości TPI a mapę TPI Landforms.

W pracy nie wyjaśniono dlaczego do wygenerowania map umieszczonych na Rycinie 12 użyto parametrów domyślnych. Mogą one być zupełnie przypadkowe, wcale nie optymalne. Metody użyte do utworzenia map TPI Landforms i geomorphons (Fig. 12B i Fig. 12D) dają możliwość manipulacji parametrami wejściowymi i w konsekwencji uzyskania istotnie różniących się map.

Celem pracy trzeciej ***Explanation of the influence of geomorphometric variables on the landform classification based on selected areas in Poland*** (Dyba 2024) było wykonanie klasyfikacji nadzorowanej form rzeźby za pomocą trzech algorytmów uczenia maszynowego i oceny użyteczności użytych algorytmów do kartowania geomorfologicznego. Jako drugi cel Autor wskazuje interpretację wykonanej klasyfikacji, a w szczególności wyjaśnienie, które ze zmiennych geomorfometrycznych są najważniejsze i jak wpływają na wynik klasyfikacji. W opracowaniu wykorzystano 8 dostępnych arkuszy Cyfrowej Mapy Geomorfologicznej Polski, odpowiadających dokładnością opracowaniom w skali 1:100 000 oraz model DTED2 i jego 15, ostatecznie, 9 wybranych pochodnych. Użyto wybrane metody klasyfikacji nadzorowanej (Random Forest, XGBoost i LightGBM) do predykcji form rzeźby w obszarach testowych, a następnie wykonano ocenę wydajności klasyfikacyjnej modeli. W kolejnym etapie zgłębiono analizę zależności pomiędzy wykorzystanymi zmiennymi a prawdopodobieństwem wystąpienia określonych form, na przykładzie czterech reprezentatywnych elementów rzeźby. W ten sposób wskazano najistotniejsze ze zmiennych oraz dokonano oceny ich wpływu na wynik klasyfikacji z wykorzystaniem metody interpretowalności uczenia maszynowego ALE. Wykazano, że model XGBoost charakteryzuje się największą dokładnością, kolejno są Random Forest i LightGBM.

UWAGI:

Kwestia formalna – W tekście Autor często w narracji używa pierwszej osoby w liczbie mnogiej (we divided, we removed, we first assessed itp.). Nie jestem specjalistą od języka angielskiego, ale nie wydaje mi się aby to była forma właściwa w przypadku tekstu mającego jednego Autora.

Pixels czy cells – w odniesieniu do danych rastrowych właściwszym terminem wydaje się komórka (cell).

Nie istnieje nic takiego jak *geomorphological forms* poprawnym terminem jest *landforms*.

Zmienne geomorfometryczne wygenerowano z użyciem parametrów domyślnych. Czy ta kwestia była jakoś analizowana? Czy były to parametry optymalne? Nic nie ma o tym w tekście.

Wskaźnik TPI był zapewne liczony w innym oprogramowaniu. W wersji GRASS GIS 7.8.0, użytej w opracowaniu, go nie ma. Jakich parametrów użyto?

W tym opracowaniu użyto jako powierzchni uczących formy z obszaru całej Polski, ze wszystkich stref morfogenetycznych razem. Jak wiadomo np. stoki mogą się między sobą istotnie różnić pod względem morfologicznym. Zupełnie inaczej wygląda stok tatrzański, inaczej wyżynny a jeszcze inny jest w obszarze staro czy młodoglacjalnym. Może właściwsze było by ich rozdzielenie form według stref morfogenetycznych.

Wnioski i dyskusja w autoreferacie

W dyskusji Doktorant zwraca uwagę na szybki rozwój wyjaśnialnych metod uczenia maszynowego i ich ciągle nieugruntowaną pozycję w nauce. Zaznacza, że ciągle trwa dyskusja na temat pojmowania pojęć „wyjaśnialność” i „interpretowalność” w kontekście rozumienia działania metod uczenia maszynowego. Wskazuje, że część badaczy rekomenduje posługiwanie się modelami typu white-box z uwagą na ich powtarzalność i możliwość rozumienia w jaki sposób uzyskuje się wynik. Przy czym istnieją

tu też kwestie problematyczne, do których zalicza fakt, że wyjaśnienie jest tylko przybliżeniem i/lub uproszczeniem działania modelu i może być podatne na błędy danych wejściowych. Dalej stwierdza, że w świetle uzyskanych wyników zastosowane w dysertacji metody przyczyniły się do zwiększenia wiedzy o procesach geoprzestrzennych. W pracy pierwszej zidentyfikowano najważniejsze zmienne objaśniające, ale nie wskazano w jaki sposób zmieniają one wynik predykcji. W pracy drugiej ujawniono jak użyte zmienne geomorfometryczne wpływają na prawdopodobieństwo klasyfikacji form terenu i jakie są interakcje pomiędzy predykcją i zmiennymi. W tej pracy podkreślono wagę interpretacji uzyskanych wyników i ich właściwego opisu. W pracy trzeciej wykazano wpływ zmiennych geomorfometrycznych na wynik klasyfikacji nadzorowanej oraz wydzielono zmienne mające największy na nią wpływ.

We wnioskach Doktorant stwierdza, że użycie metod wyjaśnialnych do analizy danych przestrzennych wnosi nową wartość w kontekście interpretacji modeli black-box wykorzystywanych do rozwiązywania problemów klasyfikacyjnych i regresji. Rozumienie wpływu badanych zmiennych na wynik predykcji i procesu decyzyjnego modeli przyczynia się do wzrostu zaufania do uzyskiwanych wyników oraz poszerzenia wiedzy o wzorcach i procesach przestrzennych. Z kolei konfrontacja wiedzy z modelowaniem z ekspercką może przyczynić się do poprawy samego modelu. Doktorant zwraca jeszcze uwagę na wpływ jakości danych wejściowych, właściwy dobór zmiennych oraz samego modelu do celu może mieć istotny wpływ na ostateczny wynik. Jako końcową konkluzję wymienia postulat upowszechniania i dalszy rozwój metod wyjaśnialnych w badaniach wykorzystujących uczenie maszynowe w kontekście przestrzennym.

III. Dorobek naukowy i dydaktyczny Doktoranta

Magister Krzysztof Dyba, według informacji zamieszczonych na portalach Google Scholar i Research Gate, swoje zainteresowania naukowe definiuje słowami kluczowymi: teledetekcja, geoinformatyka (GIS) i nauka o danych przestrzennych.

Na Jego dorobek naukowy składa się 5 prac wymienionych w autereferacie, z czego 4 są indeksowane w Web of Science. Trzy z nich stanowią podstawę recenzowanej rozprawy doktorskiej. Prace te ukazały się w okresie 2021-2024 i lokują się w następujących dziedzinach: geosciences, multidisciplinary sciences i environmental sciences. Według informacji opublikowanych w serwisie Google Scholar Doktorant ma w dorobku 5 prac, cytowanych 24 razy. Według serwisu Research Gate Doktorant ma w dorobku 5 prac cytowanych 18 razy. Indeks Hirscha Doktoranta, wg bazy WoS, wynosi 2, wg Google Scholar ma wartość 3 a wg Resarch Gate $h=3$ (15.07.2024).

Ponadto w dorobku naukowym doktorant wykazuje udział w 7 konferencjach naukowych, których był aktywnym uczestnikiem. Brał udział w realizacji 2 grantów finansowanych przez NCN oraz był głównym wykonawcą we własnym grantcie.

Jako inne osiągnięcie wymienia autorstwo pakietu oprogramowania służącego do wyszukiwania i pobierania z internetu danych geoprzestrzennych, które zostało nagrodzone przez GUGiK w 2020 roku.

W sferze kształcenia Doktorant wykazuje prowadzenie zajęć z teledetekcji i analiz przestrzennych oraz udział w workshopach dotyczących GIS i programowania w języku R. Dwukrotnie był zapraszany jako wykładowca na szkoły letnie OpenGeoHub.

Podsumowując cały dorobek naukowy Doktoranta należy uznać za bardzo dobry, spełniający wymagania stawiane w tym zakresie kandydatom do stopnia naukowego doktora.

IV. Wniosek końcowy

Recenzowana rozprawa doktorska „**Interpretation of geospatial data using explainable machine learning methods**” autorstwa mgr Krzysztofa Dyby, przygotowana pod opieką dra hab. Jarosława Jasiewicza oraz dra hab. inż. Cezarego Kaźmierowskiego, stanowi oryginalne opracowanie naukowe wnoszące istotne treści w zakresie badań geograficznych z wykorzystaniem metod uczenia maszynowego. Uzyskane wyniki ukazują duży potencjał użytych metod badawczych do rozwiązywania problemów badawczych z różnych dziedzin nauk o Ziemi. Autor wykazał się dużą wiedzą w zakresie technik geoinformatycznych a także odpowiednią wiedzą w zakresie nauk o Ziemi oraz dowiódł umiejętności planowania i prowadzenia badań naukowych, na wszystkich etapach ich realizacji.

Tym samym przedstawiona dysertacja wyczerpuje wymagania ustawy *Prawo o szkolnictwie wyższym i nauce* z dnia 20 lipca 2018 r. oraz z ustawy z dnia 3 lipca 2018 r. *Przepisy wprowadzające ustawę – Prawo o szkolnictwie wyższym i nauce*. W związku z tym wnoszę do Rady naukowej dyscypliny nauki o Ziemi i środowisku Uniwersytetu Adama Mickiewicza w Poznaniu o dopuszczenie Pana mgra Krzysztofa Dybę do dalszych etapów postępowania o nadanie stopnia naukowego doktora.

dr hab. Leszek Gawrysiak

