

Nazwa przedmiotu	<b>Analiza danych z wykorzystaniem narzędzi sztucznej inteligencji</b>
Nazwa jednostki prowadzącej przedmiot	Instytut Katalizy i Fizykochemii Powierzchni im. Jerzego Habera Polskiej Akademii Nauk przy współudziale przedstawiciela firmy StatSoft
Język przedmiotu	polski
Efekty kształcenia dla przedmiotu ujęte w kategoriach: wiedzy, umiejętności i kompetencji społecznych	<p>Po zakończeniu zajęć student:</p> <p><b><u>w zakresie wiedzy:</u></b></p> <ul style="list-style-type: none"> <li>- posiada wiedzę ogólną z dyscyplin reprezentowanych przez strony umowy (EK2-W) w szczególności: <ul style="list-style-type: none"> <li>• - wie jakie klasy problemów można rozwiązywać za pomocą metod machine learning;</li> <li>• - wie jakie są różnice i podobieństwa pomiędzy statystyczną analizą danych a machine learning;</li> <li>• - zna popularne metody machine learning: sieci neuronowe, drzewa decyzyjne, k najbliższych sąsiadów, k-średnich;</li> </ul> </li> </ul> <p><b><u>w zakresie umiejętności:</u></b></p> <ul style="list-style-type: none"> <li>- jest w stanie biegle komunikować się w zakresie swej specjalności z innymi badaczami, min. dokonać krytycznej analizy i oceny rezultatów badań, prezentować swe wyniki i poglądy międzynarodowej społeczności naukowej oraz szerszemu otoczeniu społecznemu na spotkaniach krajowych i międzynarodowych (EK6-U) w szczególności: <ul style="list-style-type: none"> <li>• - potrafi dokonać operacjonalizacji własnego problemu badawczego na optymalną strategię uczenia maszynowego;</li> <li>• - umie przeprowadzić proces doboru hiperparametrów modelu (np. funkcje aktywacji, liczba neuronów);</li> <li>• - umie zbudować modele predykcyjne oraz dokonać wyboru najlepszego z nich;</li> <li>• - potrafi zastosować zbudowany model na nowych danych;</li> </ul> </li> </ul> <p><b><u>w zakresie kompetencji społecznych:</u></b></p> <ul style="list-style-type: none"> <li>- jest gotów do myślenia i działania w sposób przedsiębiorczy, związany w szczególności z komercjalizacją i transferem technologii (EK11-KS);</li> <li>- jest gotów wykazać się odpowiedzialnością za własny rozwój naukowy.</li> <li>- jest gotów do podtrzymywania i rozwijania etosu środowisk badawczych, w tym do prowadzenia badania w sposób niezależny. (EK12-KS).</li> </ul>

Typ przedmiotu (obowiązkowy/fakultatywny)	Do wyboru
Semestr/rok	II semestr / II rok
Imię i nazwisko osoby/osób prowadzącej/prowadzących przedmiot	mgr Grzegorz Migut – prowadzący przedmiot dr hab. Maciej Szaleniec prof. IKiFP – współprowadzący
Imię i nazwisko osoby/osób egzaminującej/egzaminujących bądź udzielającej zaliczenia, w przypadku gdy nie jest to osoba prowadząca dany przedmiot	mgr Grzegorz Migut - koordynator przedmiotu
Sposób realizacji	Zajęcia wymagające bezpośredniego udziału nauczyciela akademickiego i słuchaczy.
Wymagania wstępne i dodatkowe	Podstawowa wiedza z zakresu metodologii prowadzenia badań naukowych.  Znajomość podstawowych pojęć rachunku prawdopodobieństwa (prawdopodobieństwo, miary tendencji centralnej i rozproszenia).
Liczba punktów ECTS przypisana przedmiotowi	2 ECTS
Bilans punktów ECTS	Rozliczenie dla 2 ECTS w semestrze: 18 godz. – udział w wykładach i warsztatach (godziny kontaktowe) 15 godz. – przygotowanie do wykładów/seminariów 10 godz. – wykonanie zadań zleconych do zaliczenia przedmiotu ----- 43 godz. – łączny nakład pracy doktoranta
Stosowane metody dydaktyczne	<ul style="list-style-type: none"> <li>• Wykład podający, prezentacja przykładów z omówieniem</li> <li>• Warsztaty – rozwiązywanie postawionych problemów badawczych z wykorzystaniem oprogramowania statystycznego</li> <li>• Dyskusja problemowa</li> <li>• Samodzielne rozwiązywanie problemów badawczych.</li> </ul>
Metody sprawdzania i oceny efektów kształcenia uzyskanych przez doktorantów	Egzamin końcowy ma formę pisemną. Warunkiem zaliczenia egzaminu jest uzyskanie minimum 60% z maksymalnej liczby punktów.
Forma i warunki zaliczenia przedmiotu, w tym zasady dopuszczenia do egzaminu, zaliczenia, a także forma i warunki zaliczenia przedmiotu	<ol style="list-style-type: none"> <li>1. Zajęcia odbywają się w formie wykładów i warsztatów.</li> <li>2. Student zobowiązany jest do uczestnictwa w co najmniej 80% zajęć przewidzianych harmonogramem zajęć. W przypadku uzasadnionej nieobecności na większej liczbie godzin student, po uzyskaniu zgody prowadzącego, zobowiązany jest zaliczyć odpowiednią część materiału w formie wskazanej przez prowadzącego.</li> </ol>

	<p>Ocena końcowa obliczana jest na podstawie wyniku oceny wykonania zadania końcowego.</p> <p>Kryteria oceny końcowej:</p> <p style="padding-left: 40px;">60,0% do 67,9% = dostateczny</p> <p style="padding-left: 40px;">68,0% - 75,9% = dostateczny plus</p> <p style="padding-left: 40px;">76,0% do 83,9% = dobry</p> <p style="padding-left: 40px;">84,0% do 91,9% = dobry plus</p> <p style="padding-left: 40px;">92,0% do 100% = bardzo dobry</p> <p>Zaliczenie w drugim terminie odbywa się w sesji poprawkowej, ma formę pisemną. Warunkiem zaliczenia egzaminu jest uzyskanie minimum 60% z maksymalnej liczby punktów. Kryteria oceny końcowej jak wyżej.</p>
<p>Treści przedmiotu*</p>	<p>Treści realizowane w czasie zajęć:</p> <ol style="list-style-type: none"> <li>1. Machine Learning\Data Mining – wprowadzenie       <ol style="list-style-type: none"> <li>a. Co to jest Machine Learning</li> <li>b. Strategie uczenia maszynowego</li> <li>c. Machine Learning a statystyka</li> </ol> </li> <li>2. Sieci neuronowe       <ol style="list-style-type: none"> <li>a. Podstawy sieci neuronowych           <ol style="list-style-type: none"> <li>i. Budowa sztucznego neuronu</li> <li>ii. Dobór struktury i rozmiaru sieci</li> <li>iii. Metody uczenia sieci</li> </ol> </li> <li>b. Budowa modeli neuronowych           <ol style="list-style-type: none"> <li>i. Wstępna analiza danych</li> <li>ii. Definiowanie sieci neuronowych</li> <li>iii. Neuronowe modele regresyjne i klasyfikacyjne</li> <li>iv. Strategie doboru hiperparametrów</li> <li>v. Ocena jakości uzyskanych modeli sieci</li> </ol> </li> <li>c. Praktyczne przykłady</li> </ol> </li> <li>3. Drzewa decyzyjne       <ol style="list-style-type: none"> <li>a. Wprowadzenie</li> <li>b. Własności drzew</li> <li>c. Praktyczny przykład</li> <li>d. Drzewa vs sieci neuronowe – dyskusja</li> </ol> </li> <li>4. K- najbliższych sąsiadów i Naiwny Klasyfikator Bayesa       <ol style="list-style-type: none"> <li>a. Wprowadzenie</li> <li>b. Praktyczny przykład</li> </ol> </li> <li>5. Analiza skupień       <ol style="list-style-type: none"> <li>a. Wprowadzenie</li> <li>b. Sieci neuronowe realizujące analizę skupień           <ol style="list-style-type: none"> <li>i. Określanie wymiaru sieci Kohonena</li> <li>ii. Wizualizacja i interpretacja wyników</li> </ol> </li> <li>c. Metoda k-średnich</li> <li>d. Praktyczny przykład</li> </ol> </li> </ol>

<p>Wykaz literatury podstawowej i uzupełniającej*</p>	<p><u>Literatura podstawowa:</u></p> <ol style="list-style-type: none"><li>1. A. Stanisławski „Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny.” Tom 1- 3. StatSoft, Kraków, 2007.</li><li>2. C. Watała „BIOSTATYSTYKA – wykorzystanie metod statystycznych w pracy badawczej w naukach biomedycznych”. Alfa-medica Press, Bielsko-Biała, 2002.</li></ol> <p><u>Literatura uzupełniająca:</u> Materiały wskazane przez koordynatora prace przeglądowe i oryginalne.</p>
---	--

\* W szczególnie uzasadnionych przypadkach można podać informację ogólną.