

Politechnika Białostocka
Wydział Informatyki

Program studiów podyplomowych

Data Science

edycja 2022/23

Sylwetka absolwenta

Studia poświęcone są zagadnieniom dotyczącym uczenia maszynowego (ang. *machine learning*), w szczególności sztucznym sieciom neuronowym, zarówno z punktu widzenia programisty jak i analityka danych oraz biznesowym aspektem i otoczeniu działania systemów wykorzystujących przetwarzanie dużych ilości danych - big data.

Realizacja procesu dydaktycznego jest ukierunkowana na przygotowanie absolwentów do wykorzystania zdobytej wiedzy na temat metod uczenia maszynowego oraz technologii w zakresie przetwarzania big data. Wiedza i umiejętności nabyte w trakcie studiów powinny umożliwić absolwentom zrozumienie problemów związanych z przetwarzania danych big data, budowę i wykorzystywanie rozwiązań opartych na języku Python i technologiach Hadoop, w zakresie programowania systemów produkcyjnych oraz projektowania i opracowywania analiz i raportów.

Absolwenci obu specjalności, wspólnie zdobywają na pierwszym semestrze wiedzę niezbędną do zrozumienia mechanizmów składowania i przetwarzania dużej ilości danych, statystycznej i wielowymiarowej analizy danych, problemów biznesowych i rozwiązań big data oraz wiedzę na temat metod i narzędzi uczenia maszynowego.

Uczestnik na drugim semestrze specjalności *Developer* zdobędzie wiedzę w zakresie wykorzystania sztucznych sieci neuronowych, deep learningu oraz umiejętności wykorzystania tej wiedzy do implementacji algorytmów przetwarzania danych w środowisku produkcyjnym Hadoop/Spark oraz języka R/Python. Natomiast uczestnik specjalności *Analityk danych* zdobędzie wiedzę w zakresie biznesowych modeli danych (finansowe, e-commerce, medyczne i social media), big data, przetwarzania wstępnego i transformacji danych, a także rozwiązań wykorzystujących Data Science. Dopelnieniem będą umiejętności pracy z wykorzystaniem języka R, raportowania i wizualizacji danych oraz wprowadzenie do wybranych środowisk analitycznych.

Do podstawowych celów studiów należy pomoc w zdobyciu kwalifikacji i nowej wiedzy nawet dla osób, które dotychczas nie miały styczności z ww. tematami. Dlatego też program studiów zawiera zagadnienia teoretyczne związane z przetwarzaniem big data oraz narzędziami uczenia maszynowego. Chcemy nauczyć nie tylko korzystania z narzędzi i programowania skryptów przetwarzających dane, ale też uświadomić słuchaczom co kryje się za poszczególnymi etapami przetwarzania danych i poleceniami, które wydają.

Dzięki partnerom merytorycznym z rynku, słuchacz otrzyma duży materiał dotyczący zagadnień przetwarzania danych i big data, problemów biznesowych i rozwiązań wykorzystujących big data, z uwzględnieniem wymagań rynku, modeli biznesowych oraz realiów biznesowych.

Uczestnikiem studiów podyplomowych może być osoba, która posiada kwalifikację pełną co najmniej na poziomie 6 PRK uzyskaną w systemie szkolnictwa wyższego i nauki.

Kandydaci ubiegający się o przyjęcie na studia podyplomowe powinni mieć podstawową wiedzę i umiejętności z zakresu statystyki matematycznej (statystyki opisowe, miary statystyczne, prawdopodobieństwo i rozkład normalny) oraz znajomość podstawowych zagadnień związanych z technologiami informacyjnymi, w szczególności związane z obsługą systemów operacyjnych. Znajomość metod statystycznych będzie ułatwieniem dla słuchacza, ale nie jest konieczna. Wiedzę tę będzie mógł uczestnik przyswoić sobie na studiach, gdyż na I semestrze są prowadzone zajęcia mające na celu uzupełnienie brakującej wiedzy. Będzie to jednak wymagało od słuchacza dodatkowych nakładów pracy.

Dodatkowo dla specjalności „Developer” konieczne jest posiadanie co najmniej średniej wiedzy z programowania obiektowego, tj. w szczególności w co najmniej jednym języku wysokiego poziomu, np. C++, CSharp, Java, Python.

PROGRAM STUDIÓW PODYPLOMOWYCH

DATA SCIENCE

Zajęcia na studiach obejmują łącznie **268 godzin**, w tym 136 godzin w semestrze pierwszym oraz 132 godziny w semestrze drugim, zgodnie z poniżej przedstawionym planem studiów.

Semestr 1 - Blok przedmiotów wspólnych

Kod	Nazwa przedmiotu	Liczba godzin					Punkty ECTS
		W	CW	PS	S	Suma	
SPDS01	Statystyczna analiza danych (Z, Z)	16	0	8	0	24	4
SPDS02	Przetwarzanie danych - bazy, hurtownie i big data (Z)	8	0	0	0	8	1
SPDS03	Wprowadzenie do Uczenia Maszynowego (Z)	16	0	0	0	16	3
SPDS05	Wybrane metody matematyczne w analizie danych (Z, Z)	8	0	8	0	16	2
SPDS07	Data Science z językiem Python 1 (Z)	0	0	16	0	16	2
SPDS25	Technologie i środowisko Hadoop Spark - Wstęp (Z, Z)	4	0	12	0	16	1
SPDS08	Język R w przetwarzaniu danych (Z)	0	0	16	0	16	3
SPDS10	Podstawowe narzędzia raportowania - Arkusz kalkulacyjny i wprowadzenie do SQL (Z)	0	0	20	0	20	3
SPDS26	Seminarium - Prezentacje projektów (Z)	0	0	0	4	4	1
Razem		52	0	80	4	136	20

Semestr 2 - Blok przedmiotów wspólnych

Kod	Nazwa przedmiotu	Liczba godzin					Punkty ECTS
		W	CW	PS	S	Suma	
SPDS04	Problemy biznesowe i rozwiązania wykorzystujące data science (Z)	12	0	0	0	12	2
SPDS06	Przegląd metod Uczenia Maszynowego (Z)	16	0	0	0	16	4
SPDS29	Seminarium - Prezentacje projektów (Z)	0	0	0	4	4	1
Razem		28	0	0	4	32	7

Semestr 2 - Blok przedmiotów - specjalność DEVELOPER

Kod	Nazwa przedmiotu	Liczba godzin					Punkty ECTS
		W	CW	PS	S	Suma	
SPDS11	Data Science z językiem Python 2 (Z)	0	0	24	0	24	3
SPDS14	Technologie i środowisko Hadoop (Z, Z)	8	0	16	0	24	4
SPDS25	Sztuczne Sieci Neuronowe i Deep Learning (Z, Z)	16	0	20	0	36	4
SPDS15	Kontenery w nowoczesnych środowiskach programis-tycznych i produkcyjnych - Docker, Kubernetes, Kafka (Z)	0	0	8	0	8	2
SPDS16	Hadoop Machine Learning Frameworks (Z)	0	0	8	0	8	2
Razem		24	0	76	0	100	15
Godziny wspólne semestr 2 + DEVELOPER		56	0	76	0	132	22

Semestr 2 - Blok przedmiotów - specjalność ANALITYK DANYCH

Kod	Nazwa przedmiotu	Liczba godzin					Punkty ECTS
		W	CW	PS	S	Suma	
SPDS19	Rynek badań, analiz i data science w Polsce i na świecie (Z)	8	0	0	0	8	1
SPDS21	Język R jako narzędzie wizualizacji i raportowania (Z)	0	0	16	0	16	2
SPDS27	Interfejsy API w przetwarzaniu danych - rozpoznawanie obrazu i przetwarzanie języka naturalnego (Z, Z)	8	0	16	0	24	4
SPDS23	Zaawansowane narzędzia analityczne (Z)	0	0	20	0	20	3
SPDS24	Język SQL dla Data Science (Z, Z)	8	0	16	0	24	4
SPDS28	Dedykowane narzędzia analityczne (Z)	0	0	8	0	8	1
Razem		24	0	76	0	100	15
Godziny wspólne semestr 2 + ANALITYK DANYCH		56	0	76	0	132	22

Oznaczenia

Formy zajęć: W - wykład, CW - ćwiczenia, PS - pracownia specjalistyczna, S - seminarium

Formy zaliczenia: Z - zaliczenie z oceną (w przypadku przedmiotów z kilkoma formami zajęć pierwsza litera dotyczy wykładu, druga pracowni specjalistycznej).

Zestawienie efektów uczenia się

Zestawienie tabelaryczne kierunkowych efektów uczenia się odnoszących się do charakterystyk drugiego stopnia (poziom 6 PRK), określonych na podstawie Ustawy z dnia 22 grudnia 2015 r. o Zintegrowanym Systemie Kwalifikacji.

Objaśnienia oznaczeń:

P6 – poziom 6 PRK (Polska Rama Kwalifikacji)

S – charakterystyka typowa dla kwalifikacji uzyskiwanych w ramach szkolnictwa wyższego

W – wiedza

T – teorie, zasady

Z – zjawiska i procesy

O – organizacja pracy

G – głębia i zakres

K – kontekst

N – narzędzia i materiały

U – umiejętności

I – informacje

W – wykorzystanie wiedzy

K – komunikowanie się

O – organizacja pracy

U – uczenie się i rozwój zawodowy

N – narzędzia i materiały

K – kompetencje społeczne

K – krytyczna ocena

O – odpowiedzialność

R – rola zawodowa

P – przestrzeganie reguł

W – współpraca

DS – Data Science

1, 2, 3 i kolejne – numery efektu uczenia się

Załącznik nr 1 do „Wytyczne do tworzenia programów studiów podyplomowych”

Symbol	Efekty uczenia się dla studiów podyplomowych	Odniesienie do charakterystyk drugiego stopnia określonych na podstawie art. 7 ust. 3 Ustawy z dnia 22 grudnia 2015 r. o Zintegrowanym Systemie Kwalifikacji na poziomie 6 PRK	Odniesienie do charakterystyk drugiego stopnia określonych na podstawie art. 7 ust. 4 Ustawy z dnia 22 grudnia 2015 r. o Zintegrowanym Systemie Kwalifikacji na poziomie 6 PRK
WIEDZA: absolwent zna i rozumie			
DS_W01	podstawowe pojęcia i metody dotyczące Uczenia Maszynowego (Machine Learning)	P6S_WG, P6S_WK	P6Z_WT, P6Z_WZ, P6Z_WO
DS_W02	podstawowe pojęcia statystycznej analizy danych oraz wybrane metody matematyczne w analizie danych	P6S_WG, P6S_WK	P6Z_WT, P6Z_WZ, P6Z_WO
DS_W03	pojęcia dotyczące idei i działania Sztucznych Sieci Neuronowych oraz metod i narzędzi Deep Learning	P6S_WG, P6S_WK	P6Z_WT, P6Z_WZ, P6Z_WO
DS_W04	struktury danych, instrukcje, kod w języku Python i R	P6S_WG, P6S_WK	P6Z_WO
DS_W05	metody gromadzenia danych oraz wybrane narzędzia zaawansowanej analizy danych, wizualizacji i analizy danych oraz danych z sieci społecznych	P6S_WG, P6S_WK	P6Z_WT, P6Z_WZ, P6Z_WO
DS_W06	ideę, sposób działania i wybrane frameworki ekosystemu Hadoop oraz narzędzia budowania nowoczesnych środowisk programistycznych i produkcyjnych dla celów przetwarzania danych	P6S_WG, P6S_WK	P6Z_WZ, P6Z_WO
UMIEJĘTNOŚCI: absolwent potrafi			
DS_U01	wykorzystać podstawowe pojęcia i metody dotyczące Uczenia Maszynowego w projektowaniu procesu przygotowania i analizy danych w samodzielnej pracy lub w projekcie zespołowym	P6S_UW, P6S_UK, P6S_UU, P6S_UO	P6Z_UI, P6Z_UN, P6Z_UO
DS_U02	przeprowadzić podstawowe operacje statystycznej analizy danych oraz zastosować wybrane metody matematyczne w analizie danych w samodzielnej pracy lub w projekcie zespołowym	P6S_UW, P6S_UK, P6S_UU, P6S_UO	P6Z_UI, P6Z_UN, P6Z_UO
DS_U03	wykorzystać podstawowe narzędzia i frameworki Sztucznych Sieci Neuronowych oraz metod i narzędzi Deep Learning	P6S_UW, P6S_UU	P6Z_UI, P6Z_UN, P6Z_UO
DS_U04	operować na poziomie podstawowym językiem Python i R w przygotowaniu, przetwarzaniu i wizualizacji danych oraz wykorzystać frameworki Machine Learning w analizie danych	P6S_UW	P6Z_UI, P6Z_UN, P6Z_UO
DS_U05	wykorzystać wybrane narzędzia zaawansowanej analizy danych oraz sieci społecznych w analizie danych	P6S_UW, P6S_UU	P6Z_UI, P6Z_UN, P6Z_UO
DS_U06	skonfigurować środowisko przetwarzania danych Hadoop z wykorzystaniem wybranych frameworków oraz produkcyjne środowisko pracy dla celów przetwarzania danych	P6S_UW, P6S_UU	P6Z_UI, P6Z_UN, P6Z_UO

KOMPETENCJE SPOŁECZNE: absolwent jest gotów do			
DS_S01	korzystania z wiedzy ekspertów odnośnie specjalizowanych narzędzi w pracy samodzielnej i zespołowej	P6S_KK, P6S_KO, P6S_KR	P6Z_KP, P6Z_KO
DS_S02	do wejścia na rynek pracy jako Junior Data Scientist i dbania o dorobek w zakresie Data Science	P6S_KR, P6S_KK, P6S_KO	P6Z_KP, P6Z_KO, P6Z_KW
DS_S03	myślenia i działania w sposób przedsiębiorczy w swojej pracy zawodowej	P6S_KO	P6Z_KO

Zasady oceniania i warunki uzyskania świadectwa ukończenia studiów

Dokumentem stwierdzającym ukończenie studiów podyplomowych jest Świadectwo ukończenia studiów podyplomowych. Wpisuje się do niego ostateczny wynik studiów, którym jest średnia ważona obliczana według wzoru:

$$\text{średnia ocen} = \frac{\sum(\text{ocena z przedmiotu} \times \text{punkty zaliczeniowe ECTS})}{\sum(\text{punktów zaliczeniowych ECTS})}$$

wyrównana do pełnej oceny wg zasady:

- a) do 3,25 – dostateczny,
- b) od 3,26 do 3,75 – dostateczny plus,
- c) od 3,76 do 4,25 – dobry,
- d) od 4,26 do 4,50 – dobry plus,
- e) od 4,51 do 5,00 – bardzo dobry.

Z każdego przedmiotu zostanie wystawiona jedna ocena. Warunkiem uzyskania świadectwa jest uzyskanie wszystkich zaliczeń i egzaminów przewidzianych w planie studiów. Przy zaliczeniach stosuje się następującą skalę ocen:

- a) bardzo dobry – 5,0
- b) dobry plus – 4,5
- c) dobry – 4,0
- d) dostateczny plus – 3,5
- e) dostateczny – 3,0
- f) niedostateczny – 2,0

Nie przewiduje się wykonania pracy końcowej.

Ramowe programy przedmiotów

Wydział Informatyki									
Kierunek studiów	Data Science						Poziom i forma studiów	studia podyplomowe	
Specjalność / ścieżka dyplomowania							Profil kształcenia		
Nazwa przedmiotu	Statystyczna analiza danych						Kod przedmiotu	SPDS01	
							Rodzaj przedmiotu	obowiązkowy	
Formy zajęć i liczba godzin	W	C	L	P	Ps	T	S	Semestr	1
	16				8			Punkty ECTS	4
Przedmioty wprowadzające									
Cele przedmiotu	Celem przedmiotu jest zapoznanie się oraz nabycie praktycznej umiejętności używania i wykorzystywania pojęć i metod statystycznych do wykonania statystyki opisowej, przeprowadzenia weryfikacji hipotez, analizy wariancji i regresji								
Treści programowe	<p>W: Podstawowe pojęcia w statystyce. Elementy kombinatoryki. Zdarzenia. Prawdopodobieństwo - rozkład prawdopodobieństwa; prawdopodobieństwo całkowite; warunkowe; zdarzenia niezależne; wzór Bayesa. Zmienne losowe - typy zmiennych; rozkłady zmiennych losowych; niezależność zmiennych; rozkłady brzegowe i warunkowe. Zmienne losowe dyskretne i ciągłe. Charakterystyki i funkcje zmiennych losowych. Zmienne dwuwymiarowe. Prawa wielkich liczb. Twierdzenia graniczne. Estymacja: estymacja punktowa i przedziałowa; przedziały ufności. Weryfikacja hipotez - testy zgodności, testy istotności, testy niezależności. Analiza wariancji. Analiza regresji.</p> <p>Ps: Czyszczenie danych. Statystyka opisowa. Estymacja punktowa i przedziałowa; przedziały ufności. Weryfikacja hipotez - testy zgodności, testy istotności, testy niezależności. Analiza wariancji. Analiza regresji.</p>								
Metody dydaktyczne	1. prezentacja multimedialna; 2. analiza wybranych przypadków; 3. samodzielne opracowanie wskazanych zagadnień; 4. ćwiczenia indywidualne, ćwiczenia grupowe; 5. przygotowanie przez słuchacza samodzielnej koncepcji projektowej; 8. dyskusje dydaktyczne.								
Forma zaliczenia	W: Zaliczenie ustne lub test pisemny Ps: Projekty częściowe na zajęciach, sprawozdania z realizowanych analiz								
Symbol efektu uczenia się	Zakładane efekty uczenia się						Odniesienie do kierunkowych efektów uczenia się		

EU1	Zna wybrane metody statystyczne	DS_W02	
EU2	Zna pojęcia statystyki opisowej, metody weryfikacji hipotez, analizę wariancji i regresji	DS_W02	
EU3	Potrafi używać metod statystycznych w analizie danych	DS_U02, DS_U04	
EU4	Potrafi wykonywać podstawowe analizy statystyczne	DS_U02, DS_U04	
Symbol efektu uczenia się	Sposoby weryfikacji efektów uczenia się	Forma zajęć, na której zachodzi weryfikacja	
EU1	Zaliczenie ustne lub test pisemny	W	
EU2	Zaliczenie ustne lub test pisemny	W	
EU3	projekty częściowe, sprawozdania z realizowanych analiz	Ps	
EU4	projekty częściowe, sprawozdania z realizowanych analiz	Ps	
Bilans nakładu pracy studenta (w godzinach)		Liczba godz.	
Wyliczenie	Udział w wykładach	16	
	Udział w zajęciach pracowni specjalistycznej	8	
	Przygotowanie do pracowni specjalistycznej	16	
	Wykonanie projektów i raportów częściowych	40	
	Udział w konsultacjach	2	
	Przygotowanie do zaliczenia i obecność na nim	18	
	RAZEM:	100	
Wskaźniki ilościowe		GODZINY	ECTS
Nakład pracy studenta związany z zajęciami wymagającymi bezpośredniego udziału nauczyciela		26	1
Nakład pracy studenta związany z zajęciami o charakterze praktycznym		66	3
Literatura podstawowa	<ol style="list-style-type: none"> W. Krysicki: Rachunek prawdopodobieństwa i statystyka matematyczna w zadaniach, cz. 1 i 2. PWN 1998 J. Koronacki, J. Mielniczuk: Statystyka. WNT 2004 J. Józwiak, J. Podgórski: Statystyka od podstaw. PWE 1994 		
Literatura uzupełniająca	<ol style="list-style-type: none"> A.D. Aczel: Statystyka w zarządzaniu. Wydawnictwo Naukowe PWN 2010 W. Niemiro: Rachunek prawdopodobieństwa i statystyka matematyczna. SNS 1999 		
Jednostka realizująca	Wydział Informatyki	Data opracowania programu	
Program opracował	dr inż. Magdalena Topczewska	14.03.2019	

Wydział Informatyki										
Kierunek studiów	Data Science							Poziom i forma studiów	studia podyplomowe	
Specjalność / ścieżka dyplomowania								Profil kształcenia		
Nazwa przedmiotu	Przetwarzanie danych - bazy, hurtownie i big data							Kod przedmiotu	SPDS02	
								Rodzaj przedmiotu	obowiązkowy	
Formy zajęć i liczba godzin	W	C	L	P	Ps	T	S	Semestr	1	
	8							Punkty ECTS	1	
Przedmioty wprowadzające										
Cele przedmiotu	Celem przedmiotu jest nabycie wiedzy na temat systemów baz danych, w szczególności zagadnień związanych z architekturą i funkcjonalnością systemów relacyjnych. Pokazane zostaną różnice pomiędzy klasyczną bazą danych, hurtownią danych a systemem przetwarzania big data									
Treści programowe	Business intelligence, data science and big data What makes big data "big"; Understanding how big data impacts consumers, businesses, and research Exploring the intersection of data science and big data; Facing big data's ethical challenges; Understanding the sources and structures of big data; Storing big data; Prepping big data for analysis									
Metody dydaktyczne	1. prezentacja multimedialna; 2. analiza wybranych przypadków; 3. samodzielne opracowanie wskazanych zagadnień; 4. ćwiczenia indywidualne, ćwiczenia grupowe; 5. przygotowanie przez słuchacza samodzielnej koncepcji projektowej; 8. dyskusje dydaktyczne.									
Forma zaliczenia	Zaliczenie ustne lub test pisemny									
Symbol efektu uczenia się	Zakładane efekty uczenia się							Odniesienie do kierunkowych efektów uczenia się		
EU1	Zna podstawowe pojęcia z zakresu baz danych, projektowania i hurtownie danych							DS_W06		
EU2	Ma wiedzę na temat zastosowań baz danych oraz systemów hurtowni danych i big data							DS_W01, DS_W06		
EU3	Wie czym różni baza danych od hurtowni oraz systemu analizy big data							DS_W06		
EU4	Wie jaka architektura bazy danych najbardziej odpowiada konkretnemu celowi przetwarzania danych							DS_W06		
Symbol efektu	Sposoby weryfikacji efektów uczenia się							Forma zajęć, na której		

uczenia się		zachodzi weryfikacja	
EU1	Zaliczenie ustne lub test pisemny	W	
EU2	Zaliczenie ustne lub test pisemny	W	
EU3	Zaliczenie ustne lub test pisemny	W	
EU4	Zaliczenie ustne lub test pisemny	W	
Bilans nakładu pracy studenta (w godzinach)		Liczba godz.	
Wyliczenie	Udział w wykładach	8	
	Przygotowanie do wykładu	10	
	Przygotowanie do zaliczenia	10	
	RAZEM:	28	
Wskaźniki ilościowe		GODZINY	ECTS
Nakład pracy studenta związany z zajęciami wymagającymi bezpośredniego udziału nauczyciela		8	0,25
Nakład pracy studenta związany z zajęciami o charakterze praktycznym		20	0,75
Literatura podstawowa	<ol style="list-style-type: none"> 1. Elmasri R., Navathe S.B., Wprowadzenie do systemów baz danych, Wyd. Helion, Seria „Kanon Informatyki”, Gliwice, 2005 2. Garcia-Molina H., Ullman J.D., Widom J., Systemy baz danych. Pełny wykład, WNT, Seria „Klasyka Informatyki”, Warszawa, 2006 3. Królikowski Z.: Hurtownie danych - logiczne i fizyczne struktury danych. Wydawnictwo Politechniki Poznańskiej, 2007 		
Literatura uzupełniająca	<ol style="list-style-type: none"> 1. Jarke M., Lenzerini M., Vassiliou Y., Vassiliadis P.: Fundamentals of Data Warehouses. Springer Verlag, 2010 2. Golfarelli M., Rizzi S.: Data Warehouse Design: Modern Principles and Methodologies. McGraw-Hill Osborne, 2009 3. Wrembel R., Bębel B., Oracle - Projektowanie rozproszonych baz danych, HELION Publisher, 2003, ISBN 83-7197-951-7 		
Jednostka realizująca	Wydział Informatyki Politechniki Białostockiej	Data opracowania programu	
Program opracował	dr inż. Paweł Tadejko	14.03.2019	

Wydział Informatyki										
Kierunek studiów	Data Science							Poziom i forma studiów	studia podyplomowe	
Specjalność / ścieżka dyplomowania								Profil kształcenia		
Nazwa przedmiotu	Wprowadzenie do Uczenia Maszynowego							Kod przedmiotu	SPDS03	
								Rodzaj przedmiotu	obowiązkowy	
Formy zajęć i liczba godzin	W	C	L	P	Ps	T	S	Semestr	1	
	16							Punkty ECTS	3	
Przedmioty wprowadzające										
Cele przedmiotu	Nabycie wiedzy, w jaki sposób prawidłowo przygotować dane do analizy, metod uczenia maszynowego, sposobów wyboru modelu odpowiedniego do postawionego zadania badawczego, doboru parametrów oraz przeprowadzenia procesu uczenia									
Treści programowe	Wprowadzenie do analizy danych z przestrzeni wielowymiarowych. Wielowymiarowy rozkład normalny. Problem przekleństwa wielowymiarowości. Wstępne przygotowanie danych do analizy (braki w danych, obserwacje odstające, niezbalansowanie klas, wybielanie danych, ang. data binning). Podstawowe metody wizualizacji danych, w tym analiza składowych głównych oraz skalowanie wielowymiarowe. Poznanie pojęć uczenia nienadzorowanego oraz nadzorowanego. Podstawowe metody klasyfikacji i grupowania danych. Metody ewaluacji modelu. Analiza składowych niezależnych. Bayesowska teoria decyzji. Metoda Maximum likelihood estimation.									
Metody dydaktyczne	1. prezentacja multimedialna; 2. analiza wybranych przypadków; 3. dyskusje dydaktyczne									
Forma zaliczenia	Test pisemny									
Symbol efektu uczenia się	Zakładane efekty uczenia się							Odniesienie do kierunkowych efektów uczenia się		
EU1	Zna modele badawcze							DS_W01, DS_W02		
EU2	Zna podstawowe zagadnienia uczenia maszynowego							DS_W01, DS_W02		
EU3	Posiada wiedzę na temat prawidłowego przygotowania danych do analizy							DS_W01, DS_W02		
EU4	Posiada wiedzę na temat podejść do wyboru odpowiedniego modelu badawczego							DS_W01, DS_S01		

Symbol efektu uczenia się	Sposoby weryfikacji efektów uczenia się	Forma zajęć, na której zachodzi weryfikacja	
EU1	Test pisemny	W	
EU2	Test pisemny	W	
EU3	Test pisemny	W	
EU4	Test pisemny	W	
Bilans nakładu pracy studenta (w godzinach)		Liczba godz.	
Wyliczenie	Udział w wykładach	16	
	Udział w konsultacjach	5	
	Przygotowanie do wykładu	24	
	Przygotowanie do zaliczenia	30	
	RAZEM:	75	
Wskaźniki ilościowe		GODZINY	ECTS
Nakład pracy studenta związany z zajęciami wymagającymi bezpośredniego udziału nauczyciela		21	0,84
Nakład pracy studenta związany z zajęciami o charakterze praktycznym		59	2,36
Literatura podstawowa	<ol style="list-style-type: none"> 1. T. Mitchell: Machine learning. McGraw Hill, 1997 2. R.T. Clemen: Making hard decisions. An Introduction to decision analysis. Duxbury Press 1996 3. P. Cichosz: Systemy uczące się. WNT 2009 		
Literatura uzupełniająca	<ol style="list-style-type: none"> 1. R.O. Duda, P.E. Hart, D.G. Stork: Pattern classification. Wiley, 2000 		
Jednostka realizująca	Wydział Informatyki Politechniki Białostockiej	Data opracowania programu	
Program opracował	<i>dr inż. Paweł Tadejko, dr inż. Magdalena Topczewska</i>	14.03.2019	

Wydział Informatyki										
Kierunek studiów	Data Science							Poziom i forma studiów	studia podyplomowe	
Specjalność / ścieżka dyplomowania								Profil kształcenia		
Nazwa przedmiotu	Wybrane metody matematyczne w analizie danych							Kod przedmiotu	SPDS05	
								Rodzaj przedmiotu	obowiązkowy	
Formy zajęć i liczba godzin	W	C	L	P	Ps	T	S	Semestr	1	
	8				8			Punkty ECTS	2	
Przedmioty wprowadzające										
Cele przedmiotu	Nabycie wiedzy i zrozumienia efektywności algebry liniowej przy wykorzystaniu w data science. Zrozumienie podstawowych pojęć matematycznych algorytmów stosowanych w data science. Wykorzystanie pakietu NumPy i odpowiednich metod matematycznych przekształceń danych i obliczeń									
Treści programowe	W: Algebra liniowa - wprowadzenie i operacje macierzowe. Metody numeryczne. Równania różniczkowe, pochodne cząstkowe. Funkcje. Optymalizacja. Metody stochastyczne. Metoda PCA, metoda SVD.									
	Ps: Operacje na macierzach. Metody optymalizacji. Równania różniczkowe. Metoda PCA, SVD.									
Metody dydaktyczne	1. prezentacja multimedialna; 2. analiza wybranych przypadków; 3. samodzielne opracowanie wskazanych zagadnień; 4. ćwiczenia indywidualne, ćwiczenia grupowe; 5. przygotowanie przez słuchacza samodzielnej koncepcji projektowej; 8. dyskusje dydaktyczne.									
Forma zaliczenia	W: Test pisemny									
	Ps: Projekt końcowy lub projekty cząstkowe na zajęciach, praca domowa									
Symbol efektu uczenia się	Zakładane efekty uczenia się							Odniesienie do kierunkowych efektów uczenia się		
EU1	Zna modele badawcze							DS_W01, DS_W02, DS_W04		
EU2	Posiada wiedzę na temat podstawowych zagadnień uczenia maszynowego							DS_W01, DS_W02		
EU3	Potrafi prawidłowo przygotować dane do analizy							DS_U01		
EU4	Potrafi dobrać odpowiedni model badawczy							DS_U01, DS_U02, DS_U04		
Symbol efektu uczenia się	Sposoby weryfikacji efektów uczenia się							Forma zajęć, na której zachodzi weryfikacja		
EU1	Test pisemny							W		

EU2	Test pisemny	W	
EU3	Praca domowa, projekt końcowy lub projekty cząstkowe	Ps	
EU4	Praca domowa, projekt końcowy lub projekty cząstkowe	Ps	
Bilans nakładu pracy studenta (w godzinach)		Liczba godz.	
Wyliczenie	Udział w wykładach	8	
	Udział w pracowni specjalistycznej	8	
	Przygotowanie do pracowni specjalistycznej	2	
	Opracowanie sprawozdań z pracowni i/lub wykonanie zadań domowych	16	
	Udział w konsultacjach	2	
	Realizacja programistycznych zadań domowych	10	
	Przygotowanie do zaliczenia	10	
	RAZEM:		56
Wskaźniki ilościowe		GODZINY	ECTS
Nakład pracy studenta związany z zajęciami wymagającymi bezpośredniego udziału nauczyciela		25	1
Nakład pracy studenta związany z zajęciami o charakterze praktycznym		38	1,5
Literatura podstawowa	1. Numpy Beginner's Guide by Ivan Idris, Packt Publishing, 2015 2. Coding the Matrix: Linear Algebra through Applications to Computer Science 1st Edition by Philip N. Klein, Newtonian Press, 2013		
Literatura uzupełniająca	1. Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython 1st Edition by Wes McKinney, O'Reilly Media, Inc, 2012		
Jednostka realizująca	Wydział Informatyki Politechniki Białostockiej	Data opracowania programu	
Program opracował	<i>dr inż. Dmitry Ziolkovskiy</i>	14.03.2019	

Wydział Informatyki									
Kierunek studiów	Data Science						Poziom i forma studiów	studia podyplomowe	
Specjalność / ścieżka dyplomowania							Profil kształcenia		
Nazwa przedmiotu	Data Science z językiem Python 1						Kod przedmiotu	SPDS07	
							Rodzaj przedmiotu	obowiązkowy	
Formy zajęć i liczba godzin	W	Ć	L	P	Ps	T	S	Semestr	1
					16			Punkty ECTS	2
Przedmioty wprowadzające									
Cele przedmiotu	Zapoznanie słuchaczy ze składnią i możliwościami języka Python na poziomie podstawowym oraz nabycie umiejętności podstawowego programowania w tym języku								
Treści programowe	Wprowadzenie do języka Python; Zaawansowany Python. Sphinx, docstrings, unittest, py.test, coverage.py. Użycie bibliotek Python do stworzenia data processing workflow: przetwarzanie z Numpy; praca ze strukturami danych w Pandas; podstawy wizualizacji danych z Matplotlib								
Metody dydaktyczne	1. prezentacja multimedialna; 2. analiza wybranych przypadków; 3. samodzielne opracowanie wskazanych zagadnień; 4. ćwiczenia indywidualne, ćwiczenia grupowe; 5. przygotowanie przez słuchacza samodzielnej koncepcji projektowej; 8. dyskusje dydaktyczne.								
Forma zaliczenia	Projekt końcowy lub projekty cząstkowe na zajęciach, praca domowa								
Symbol efektu uczenia się	Zakładane efekty uczenia się						Odniesienie do kierunkowych efektów uczenia się		
EU1	Zna i stosuje podstawy języka Python, składni i jego podstawowych bibliotek						DS_W04, DS_U04		
EU2	Zna i stosuje różne sposoby wykorzystania Pythona pod kątem przetwarzania danych						DS_W04, DS_U04		
EU3	Potrafi posługiwać się językiem Python na poziomie podstawowym						DS_U04		
EU4	Potrafi wykorzystać podstawowe biblioteki języka Python w przetwarzaniu danych						DS_U04		
Symbol efektu uczenia się	Sposoby weryfikacji efektów uczenia się						Forma zajęć, na której zachodzi weryfikacja		
EU1	Projekt końcowy lub projekty cząstkowe na zajęciach						Ps		

EU2	Projekt końcowy lub projekty cząstkowe na zajęciach	Ps	
EU3	Praca domowa, projekty cząstkowe na zajęciach	Ps	
EU4	Praca domowa, projekty cząstkowe na zajęciach	Ps	
Bilans nakładu pracy studenta (w godzinach)		Liczba godz.	
Wyliczenie	Udział w pracowni specjalistycznej	16	
	Opracowanie sprawozdań z pracowni	10	
	Realizacja zadań domowych	20	
	Udział w konsultacjach z wiązanych z pracownią specjalistyczną	5	
	RAZEM:	51	
Wskaźniki ilościowe		GODZINY	ECTS
Nakład pracy studenta związany z zajęciami wymagającymi bezpośredniego udziału nauczyciela		21	0,82
Nakład pracy studenta związany z zajęciami o charakterze praktycznym		51	2
Literatura podstawowa	<ol style="list-style-type: none"> Python. Receptury, Beazley David, Jones K. Brian, Wydawnictwo Helion, 2014 Python. Podstawy nauki o danych. Alberto Boschetti, Luca Massaron, Helion, 2017 		
Literatura uzupełniająca	<ol style="list-style-type: none"> Introduction to Python for Data Science edX : https://www.edx.org/course/introduction-python-data-science-microsoft-dat208x-5 Introduction to Data Science in Python Coursera : https://www.coursera.org/learn/python-data-analysis 		
Jednostka realizująca	Wydział Informatyki Politechniki Białostockiej	Data opracowania programu	
Program opracował	<i>mgr inż. Łukasz Prusiel, mgr inż. Piotr Tynecki</i>	14.03.2019	

Wydział Informatyki									
Kierunek studiów	Data Science						Poziom i forma studiów	studia podyplomowe	
Specjalność / ścieżka dyplomowania							Profil kształcenia		
Nazwa przedmiotu	Technologie i środowisko Hadoop Spark - Wstęp						Kod przedmiotu	SPDS25	
							Rodzaj przedmiotu	obowiązkowy	
Formy zajęć i liczba godzin	W	C	L	P	Ps	T	S	Semestr	1
	4				12			Punkty ECTS	1
Przedmioty wprowadzające									
Cele przedmiotu	Celem zajęć jest nabycie wiedzy oraz umiejętności obsługi i instalacji technologii i środowiska Hadoop i Spark oraz wybranych frameworków do składowania i przetwarzania danych								
Treści programowe	<p>W: Ekosystem Hadoop oraz Spark. Podstawowe operacje na systemie plików HDFS.</p> <p>Ps: Ekosystem Hadoop oraz Spark. Poznanie podstawowych operacji na systemie plików HDFS. Umiejętność analizy i przetwarzania dużych ilości danych (Big Data) z wykorzystaniem silnika obliczeniowego Spark. Tworzenie i przetwarzanie Spark DataFrame przy wykorzystaniu języka Python (PySpark) oraz SQL (Spark SQL). Analiza, interpretacja i wizualizacja zbiorów danych.</p>								
Metody dydaktyczne	1. prezentacja multimedialna; 2. analiza wybranych przypadków; 3. samodzielne opracowanie wskazanych zagadnień; 4. ćwiczenia indywidualne, ćwiczenia grupowe; 5. przygotowanie przez słuchacza samodzielnej koncepcji projektowej; 8. dyskusje dydaktyczne.								
Forma zaliczenia	<p>W: Test pisemny</p> <p>Ps: Projekty cząstkowe na zajęciach, praca domowa</p>								
Symbol efektu uczenia się	Zakładane efekty uczenia się						Odniesienie do kierunkowych efektów uczenia się		
EU1	Zna podstawową architekturę systemów Apache Hadoop i Spark						DS_W06		
EU2	Zna mechanizmy składowania i przetwarzania danych wielkoskalowych za pomocą systemie plików HDFS i silnika Apache Spark						DS_W06		
EU3	Potrafi wykorzystać narzędzia Apache Spark do wczytywania, przetwarzania i analizy danych						DS_U06, DS_S02		
EU4	Potrafi zainstalować środowiska Apache Spark i wybrane frameworki						DS_U06, DS_S02		
Symbol efektu	Sposoby weryfikacji efektów uczenia się						Forma zajęć, na której		

uczenia się		zachodzi weryfikacja	
EU1	Test pisemny	W	
EU2	Test pisemny	W	
EU3	Praca domowa i projekty cząstkowe na zajęciach	Ps	
EU4	Praca domowa i projekty cząstkowe na zajęciach	Ps	
Bilans nakładu pracy studenta (w godzinach)		Liczba godz.	
Wyliczenie	Udział w wykładach	4	
	Udział w pracowni specjalistycznej	12	
	Wykonanie zadań domowych	12	
	Przygotowanie do zaliczenia	5	
	RAZEM:	28	
Wskaźniki ilościowe		GODZINY	ECTS
Nakład pracy studenta związany z zajęciami wymagającymi bezpośredniego udziału nauczyciela		16	0,5
Nakład pracy studenta związany z zajęciami o charakterze praktycznym		28	1
Literatura podstawowa	1. Holden Karau, Krishna Sankar, Fast Data Processing with Spark 2, 3rd Edition, Packt Publishing 2017 2. https://spark.apache.org/docs/latest/ 3. https://hadoop.apache.org/docs/r2.7.2/hadoop-project-dist/hadoop-hdfs/HDFSCommands.html		
Literatura uzupełniająca	1. M. Frampton: Mastering Apache Spark. Packt Publishing 2015		
Jednostka realizująca	Wydział Informatyki Politechniki Białostockiej	Data opracowania programu	
Program opracował	<i>mgr inż. Daniel Reska</i> <i>mgr inż. Michał Wawrzyniak</i>	18.03.2019	

Wydział Informatyki									
Kierunek studiów	Data Science							Poziom i forma studiów	studia podyplomowe
Specjalność / ścieżka dyplomowania								Profil kształcenia	
Nazwa przedmiotu	Język R w przetwarzaniu danych							Kod przedmiotu	SPDS08
								Rodzaj przedmiotu	obowiązkowy
Formy zajęć i liczba godzin	W	Ć	L	P	Ps	T	S	Semestr	1
					16			Punkty ECTS	3
Przedmioty wprowadzające									
Cele przedmiotu	Nabywanie umiejętności programowania w R w głównej mierze na tym, co w powszechnej opinii środowiska data science zajmuje największą część pracy badacza, czyli - wydobywaniu, czyszczeniu, transformacji danych. Drugi ważny moduł to z kolei eksploracja danych								
Treści programowe	Wprowadzenie do języka R i środowiska RStudio. Przetwarzanie danych w R i data munging(pakiety dplyr, data.table, reshape2, tidyr). Praca z R w środowisku analitycznym (R i Hadoop, R i SQL, R i web scraping). Wizualizacja danych w R (ggplot2, ggvis, base graphics, lattice). Case study: Facebook i R								
Metody dydaktyczne	1. prezentacja multimedialna; 2. analiza wybranych przypadków; 3. samodzielne opracowanie wskazanych zagadnień; 4. ćwiczenia indywidualne, ćwiczenia grupowe; 5. przygotowanie przez słuchacza samodzielnej koncepcji projektowej; 8. dyskusje dydaktyczne.								
Forma zaliczenia	Projekt końcowy lub projekty cząstkowe na zajęciach, praca domowa								
Symbol efektu uczenia się	Zakładane efekty uczenia się							Odniesienie do kierunkowych efektów uczenia się	
EU1	Zna i stosuje podstawy języka R, składni i jego podstawowe biblioteki							DS_W04, DS_U04	
EU2	Posiada wiedzę i umie stosować R pod kątem przetwarzania danych							DS_W04, DS_U04	
EU3	Potrafi przetwarzać dane i przygotowywać je do analizy w R							DS_U04	
EU4	Potrafi wyciągać wnioski na podstawie wyników analizy danych w R							DS_U04	
Symbol efektu uczenia się	Sposoby weryfikacji efektów uczenia się							Forma zajęć, na której zachodzi weryfikacja	
EU1	Projekt końcowy lub projekty cząstkowe na zajęciach							Ps	

EU2	Projekt końcowy lub projekty cząstkowe na zajęciach	Ps	
EU3	Praca domowa, projekty cząstkowe na zajęciach	Ps	
EU4	Praca domowa, projekty cząstkowe na zajęciach	Ps	
Bilans nakładu pracy studenta (w godzinach)		Liczba godz.	
Wyliczenie	Udział w pracowni specjalistycznej	16	
	Przygotowanie do pracowni specjalistycznej	20	
	Opracowanie sprawozdań z pracowni	20	
	Realizacja zadań domowych	20	
	Przygotowanie do zaliczenia pracowni	10	
	RAZEM:	86	
Wskaźniki ilościowe		GODZINY	ECTS
Nakład pracy studenta związany z zajęciami wymagającymi bezpośredniego udziału nauczyciela		16	0,5
Nakład pracy studenta związany z zajęciami o charakterze praktycznym		86	3
Literatura podstawowa	1. N. Zumel, J. Mount: Practical Data Science with R. Manning Publications 2014 2. T. Fischetti: Data Analysis with R. Packt Publishing 2015		
Literatura uzupełniająca	1. R. Heimann, N. Danneman: Social Media Mining with R: Deploy cutting-edge sentiment analysis techniques to real-world social media data using R. Packt Publishing 2014 2. H. Wickham: ggplot2: Elegant Graphics for Data Analysis (Use R!). Springer 2009		
Jednostka realizująca	Wydział Informatyki Politechniki Białostockiej	Data opracowania programu	
Program opracował	<i>mgr inż. Filip Cyprowski</i>	14.03.2019	

Wydział Informatyki									
Kierunek studiów	Data Science							Poziom i forma studiów	studia podyplomowe
Specjalność / ścieżka dyplomowania								Profil kształcenia	
Nazwa przedmiotu	Podstawowe narzędzia raportowania - Arkusz kalkulacyjny i wprowadzenie do SQL							Kod przedmiotu	SPDS10
								Rodzaj przedmiotu	obowiązkowy
Formy zajęć i liczba godzin	W	C	L	P	Ps	T	S	Semestr	1
					20			Punkty ECTS	3
Przedmioty wprowadzające									
Cele przedmiotu	Celem zajęć jest nabycie umiejętności przygotowywania, weryfikowania i formatowania raportów przy pomocy odpowiednich narzędzi (głównie Excel, również bazy danych, Access)								
Treści programowe	Ps: przygotowywanie, weryfikowanie i formatowanie raportów przy pomocy odpowiednich narzędzi (Excel, Access). Zapytania SQL.								
Metody dydaktyczne	1. prezentacja multimedialna; 2. analiza wybranych przypadków; 3. samodzielne opracowanie wskazanych zagadnień; 4. ćwiczenia indywidualne, ćwiczenia grupowe; 5. przygotowanie przez słuchacza samodzielnej koncepcji projektowej; 8. dyskusje dydaktyczne.								
Forma zaliczenia	Projekt końcowy lub projekty cząstkowe na zajęciach, praca domowa								
Symbol efektu uczenia się	Zakładane efekty uczenia się							Odniesienie do kierunkowych efektów uczenia się	
EU1	Zna i stosuje podstawy raportowania							DS_W02, DS_U02	
EU2	Zna i stosuje różne formy reprezentacji danych w postaci raportów							DS_W02, DS_W05, DS_U02	
EU3	Potrafi wykorzystać w analizie i raportowaniu narzędzie Microsoft Excel							DS_U02	
EU4	Potrafi budować zapytania SQL							DS_U02	
Symbol efektu uczenia się	Sposoby weryfikacji efektów uczenia się							Forma zajęć, na której zachodzi weryfikacja	
EU1	Projekt końcowy lub projekty cząstkowe na zajęciach							Ps	
EU2	Projekt końcowy lub projekty cząstkowe na zajęciach							Ps	

EU3	Praca domowa, projekty cząstkowe na zajęciach	Ps	
EU4	Praca domowa, projekty cząstkowe na zajęciach	Ps	
Bilans nakładu pracy studenta (w godzinach)		Liczba godz.	
Wyliczenie	Udział w pracowni specjalistycznej	20	
	Przygotowanie do pracowni specjalistycznej	20	
	Opracowanie sprawozdań z pracowni	25	
	Realizacja zadań domowych	25	
	RAZEM:	90	
Wskaźniki ilościowe		GODZINY	ECTS
Nakład pracy studenta związany z zajęciami wymagającymi bezpośredniego udziału nauczyciela		20	0,75
Nakład pracy studenta związany z zajęciami o charakterze praktycznym		90	3
Literatura podstawowa	<ol style="list-style-type: none"> 1. J. Scott: Social Network Analysis Newbury Park CA: Sage, 2000 2. R. A. Hanneman, M. Riddle: Introduction to Social Network Methods, 2005 3. S. Wasserman, K. Faust: Social Network Analysis: Methods and Applications Cambridge University Press 1994 		
Literatura uzupełniająca	<ol style="list-style-type: none"> 1. W. de Nooy, A. Mrvar i Vladimir Batagelj: Exploratory Social Network Analysis with Pajek Cambridge University Press 2005 		
Jednostka realizująca	Wydział Informatyki Politechniki Białostockiej	Data opracowania programu	
Program opracował	<i>Mateusz Aniserowicz</i>	14.03.2019	

Wydział Informatyki									
Kierunek studiów	Data Science						Poziom i forma studiów	studia podyplomowe	
Specjalność / ścieżka dyplomowania							Profil kształcenia		
Nazwa przedmiotu	Seminarium - Prezentacje projektów						Kod przedmiotu	SPDS26	
							Rodzaj przedmiotu	obowiązkowy	
Formy zajęć i liczba godzin	W	C	L	P	Ps	T	S	Semestr	1
							4	Punkty ECTS	1
Przedmioty wprowadzające									
Cele przedmiotu	Celem przedmiotu jest weryfikacja pracy oraz doskonalenie techniki prezentacji poprzez realizację projektu, przedstawienie danych, wstępne postawienie hipotez i ich dyskusja, postawienie i przedstawienie zadań do realizacji, metod i algorytmów, które mogą zostać użyte oraz doskonalenie techniki prezentacji.								
Treści programowe	Wstępne wiadomości o celu seminarium. Czym należy kierować się wybierając temat projektu końcowego. Omówienie przez studentów przykładowych projektów. Ustalenie wstępnego planu projektu.								
Metody dydaktyczne	prezentacja multimedialna, samodzielne opracowanie wskazanych zagadnień, przygotowanie przez słuchacza samodzielnej koncepcji projektowej, dyskusje								
Forma zaliczenia	Ocena prezentacji realizowanego projektu								
Symbol efektu uczenia się	Zakładane efekty uczenia się						Odniesienie do kierunkowych efektów uczenia się		
EU1	Posiada podstawową wiedzę z dziedziny, do której należy temat projektu						DS_W01, DS_W02		
EU2	Potrafi przygotować prezentację ustną						DS_S01		
EU3	Potrafi analizować wymagania dotyczące wybranego tematu projektu						DS_U01, DS_U02, DS_S01		
EU4	Potrafi wybrać narzędzia i zaplanować ich użycie w poszczególnych etapach samodzielnej pracy						DS_U01, DS_U03, DS_W05		
Symbol efektu uczenia się	Sposoby weryfikacji efektów uczenia się						Forma zajęć, na której zachodzi weryfikacja		
EU1	Ocena prezentacji realizowanego projektu						S		

EU2	Ocena prezentacji realizowanego projektu	S	
EU3	Ocena prezentacji realizowanego projektu	S	
EU4	Ocena prezentacji realizowanego projektu	S	
Bilans nakładu pracy studenta (w godzinach)		Liczba godz.	
Wyliczenie	Udział w seminarium	4	
	Udział w konsultacjach	5	
	Przygotowanie do wystąpienia	20	
	RAZEM:	29	
Wskaźniki ilościowe		GODZINY	ECTS
Nakład pracy studenta związany z zajęciami wymagającymi bezpośredniego udziału nauczyciela		9	0,3
Nakład pracy studenta związany z zajęciami o charakterze praktycznym		25	1
Literatura podstawowa	1. Regulamin studiów PB		
Literatura uzupełniająca	1. Wytoczne i zalecenia w sprawie przygotowywania prac dyplomowych na Wydziale Informatyki PB		
Jednostka realizująca	Wydział Informatyki Politechniki Białostockiej	Data opracowania programu	
Program opracował	<i>dr inż. Magdalena Topczewska</i>	14.03.2019	

Wydział Informatyki									
Kierunek studiów	Data Science						Poziom i forma studiów	studia podyplomowe	
Specjalność / ścieżka dyplomowania							Profil kształcenia		
Nazwa przedmiotu	Problemy biznesowe i rozwiązania wykorzystujące data science						Kod przedmiotu	SPDS04	
							Rodzaj przedmiotu	obowiązkowy	
Formy zajęć i liczba godzin	W	Ć	L	P	Ps	T	S	Semestr	2
	12							Punkty ECTS	2
Przedmioty wprowadzające									
Cele przedmiotu	Nabycie wiedzy jak wykorzystanie data science pozwala wspierać firmy i organizacje mające do czynienia z dużą ilością danych w podejmowaniu decyzji biznesowych								
Treści programowe	Specyfika pracy badacza, analityka i data scientist. Logika analityki danych biznesowych. Inżynieria i podejmowanie decyzji bazujące na danych. Dane i analityka jako element strategii. Od problemu biznesowego do eksploracji danych								
Metody dydaktyczne	1. prezentacja multimedialna; 2. analiza wybranych przypadków; 3. samodzielne opracowanie wskazanych zagadnień; 4. ćwiczenia indywidualne, ćwiczenia grupowe; 5. przygotowanie przez słuchacza samodzielnej koncepcji projektowej; 8. dyskusje dydaktyczne.								
Forma zaliczenia	Zaliczenie ustne lub test pisemny								
Symbol efektu uczenia się	Zakładane efekty uczenia się						Odniesienie do kierunkowych efektów uczenia się		
EU1	Zna przykładowe klasy problemów, które mogą być rozwiązywane przy pomocy data science						DS_W01, DS_W05		
EU2	Zna podstawowe techniki data science wykorzystywane w problemach biznesowych						DS_W01, DS_W05		
EU3	Posiada wiedzę na temat określenia czy problem należy do kategorii, dla których może pomóc wykorzystanie narzędzi data science						DS_W01, DS_S01		
EU4	Posiada wiedzę na temat wybranych gałęzi biznesowych, gdzie wykorzystanie data science daje mierzalne efekty						DS_W01, DS_S01		
Symbol efektu uczenia się	Sposoby weryfikacji efektów uczenia się						Forma zajęć, na której zachodzi weryfikacja		

EU1	Zaliczenie ustne lub test pisemny	W	
EU2	Zaliczenie ustne lub test pisemny	W	
EU3	Zaliczenie ustne lub test pisemny	W	
EU4	Zaliczenie ustne lub test pisemny	W	
Bilans nakładu pracy studenta (w godzinach)		Liczba godz.	
Wyliczenie	Udział w wykładach	12	
	Przygotowanie do wykładu	20	
	Przygotowanie zaliczenia	20	
	RAZEM:	52	
Wskaźniki ilościowe		GODZINY	ECTS
Nakład pracy studenta związany z zajęciami wymagającymi bezpośredniego udziału nauczyciela		12	0,5
Nakład pracy studenta związany z zajęciami o charakterze praktycznym		52	2
Literatura podstawowa	1. F. Provost, T. Fawcett: Analiza danych w biznesie. Sztuka podejmowania skutecznych decyzji. Onepress, 2014 2. E. Frątczak, U.Gach-Ciepiela, H. Babiker: Analiza historii zdarzeń. Teoria, przykłady zastosowań z wykorzystaniem programów: SAS, TDA, STATA. Oficyna Wydawnicza SGH, Warszawa 2005 3. Materiały do nauki pakietu / środowiska / programu R : http://smarterpoland.pl/		
Literatura uzupełniająca	1. J. Wieczorkowski „Wykorzystanie koncepcji big data w administracji publicznej”, Roczniki Kolegium Analiz Ekonomicznych 33/2014, SGH 2014 2. Advanced Analytics & Data Science SmarterPoland.pl : http://smarterpoland.pl		
Jednostka realizująca	Wydział Informatyki Politechniki Białostockiej	Data opracowania programu	
Program opracował	dr inż. Paweł Tadejko	14.03.2019	

Wydział Informatyki										
Kierunek studiów	Data Science							Poziom i forma studiów	studia podyplomowe	
Specjalność / ścieżka dyplomowania								Profil kształcenia		
Nazwa przedmiotu	Przegląd metod Uczenia Maszynowego							Kod przedmiotu	SPDS06	
								Rodzaj przedmiotu	obowiązkowy	
Formy zajęć i liczba godzin	W	Ć	L	P	Ps	T	S	Semestr	2	
	16							Punkty ECTS	4	
Przedmioty wprowadzające										
Cele przedmiotu	Nabywanie wiedzy odnośnie wybranych technik uczenia maszynowego. Wprowadzenie słuchaczy w teoretyczne oraz praktyczne aspekty dziedziny uczenia maszynowego. Zadaniem przedmiotu jest wprowadzenie do problematyki tworzenia systemów posiadających zdolności uczenia się na przykładach i w rezultacie doskonalenia swojego działania.									
Treści programowe	<p>Techniki nieparametryczne (density estimation, Parzen windows, K- Nearest Neighbor estimation and rule, similarity metrics). Liniowe funkcje dyskryminacyjne (two- and multicategory cases for linear discrim. functions and decision surfaces, Minimum squared error, Fisher linear discriminant)</p> <p>-special section: Kernel methods and Support Vector Machines. Algorytmy nieparametryczne (additive models, Trees and related methods (ID3,C4.5, CART, MARS)). Boosting and additive trees.</p> <p>Analiza skupień (densities estimation, criterion functions for clustering, agglomerative clustering, EM with Fisher discrimination criterion, unknown number of clusters, SOM). NLP - Natural Language Processing</p>									
Metody dydaktyczne	1. prezentacja multimedialna; 2. analiza wybranych przypadków; 3. samodzielne opracowanie wskazanych zagadnień; 4. ćwiczenia indywidualne, ćwiczenia grupowe; 5. przygotowanie przez słuchacza samodzielnej koncepcji projektowej; 8. dyskusje dydaktyczne.									
Forma zaliczenia	Zaliczenie ustne lub test pisemny									
Symbol efektu uczenia się	Zakładane efekty uczenia się							Odniesienie do kierunkowych efektów uczenia się		
EU1	Zna pojęcia podstawowe z zakresu uczenia maszynowego							DS_W01, DS_W03, DS_W05		
EU2	Zna typowe zastosowania metod uczenia maszynowego							DS_W01, DS_W03, DS_W05		
EU3	Zna wykorzystanie gotowych systemów zawierających implementacje algorytmów nadzorowanego uczenia							DS_W01		
EU4	Zna metodologię wstępnej oceny przydatności metody uczenia do							DS_W01		

	konkretnego problemu obliczeniowego	
Symbol efektu uczenia się	Sposoby weryfikacji efektów uczenia się	Forma zajęć, na której zachodzi weryfikacja
EU1	Zaliczenie ustne lub test pisemny	W
EU2	Zaliczenie ustne lub test pisemny	W
EU3	Zaliczenie ustne lub test pisemny	W
EU4	Zaliczenie ustne lub test pisemny	W
Bilans nakładu pracy studenta (w godzinach)		Liczba godz.
Wyliczenie	Udział w wykładach	16
	Przygotowanie do wykładu	35
	Udział w konsultacjach związanych z wykładem	20
	Przygotowanie do zaliczenia	30
	RAZEM:	101
Wskaźniki ilościowe		GODZINY ECTS
Nakład pracy studenta związany z zajęciami wymagającymi bezpośredniego udziału nauczyciela		36 1,42
Nakład pracy studenta związany z zajęciami o charakterze praktycznym		85 3,36
Literatura podstawowa	1. K. Krawiec K., J. Stefanowski: Uczenie maszynowe i sieci neuronowe. Wydawnictwo Politechniki Poznańskiej, Poznań, 2003 (2 wydanie 2004) 2. P. Cichosz: Systemy uczące się. WNT, Warszawa, 2000 3. I. Kononenko, M. Kukar: Machine Learning and Data Mining. Horwood Publ., 2007.	
Literatura uzupełniająca	1. T. Mitchell T.: Machine Learning. McGraw Hill, 1997 2. J. Koronacki J., J. Cwik: Statystyczne systemy uczące się. Warszawa, 2008 3. R. Tadeusiewicz: Sieci neuronowe. Akademicka Oficyna Wydawnicza, Warszawa, 1993	
Jednostka realizująca	Wydział Informatyki Politechniki Białostockiej	Data opracowania programu
Program opracował	<i>mgr inż. Mariusz Ferenc</i>	14.03.2019

Wydział Informatyki									
Kierunek studiów	Data Science						Poziom i forma studiów	studia podyplomowe	
Specjalność / ścieżka dyplomowania							Profil kształcenia		
Nazwa przedmiotu	Seminarium - Prezentacje projektów						Kod przedmiotu	SPDS29	
							Rodzaj przedmiotu	obowiązkowy	
Formy zajęć i liczba godzin	W	C	L	P	Ps	T	S	Semestr	2
							4	Punkty ECTS	1
Przedmioty wprowadzające									
Cele przedmiotu	Celem przedmiotu jest prezentacja, dyskusja i weryfikacja przygotowanego projektu zgodnie z wytycznymi, realizowanie harmonogramu oraz doskonalenie techniki prezentacji								
Treści programowe	Omówienie przez studentów zrealizowanych projektów - harmonogramu prac, postawionych hipotez badawczych, wykorzystanych metod i algorytmów, wykorzystanego oprogramowania oraz dyskusja otrzymanych wyników								
Metody dydaktyczne	prezentacja multimedialna, pokaz								
Forma zaliczenia	Ocena prezentacji projektu przedstawionej w trakcie seminarium								
Symbol efektu uczenia się	Zakładane efekty uczenia się							Odniesienie do kierunkowych efektów uczenia się	
EU1	Posiada podstawową wiedzę z dziedziny, do której należy temat projektu							DS_W01, DS_W02	
EU2	Potrafi przygotować prezentację ustną							DS_S01, DS_S02	
EU3	Potrafi znaleźć rozwiązanie realizowanego zadania projektowego							DS_U01, DS_U02, DS_U04	
EU4	Potrafi wyciągnąć wnioski odnośnie użytych narzędzi i zaproponować nowe rozwiązania							DS_W05, DS_U03	
Symbol efektu uczenia się	Sposoby weryfikacji efektów uczenia się							Forma zajęć, na której zachodzi weryfikacja	
EU1	Ocena prezentacji projektu							S	
EU2	Ocena prezentacji projektu							S	

EU3	Ocena prezentacji projektu	S	
EU4	Ocena prezentacji projektu	S	
Bilans nakładu pracy studenta (w godzinach)		Liczba godz.	
Wyliczenie	Udział w seminarium	4	
	Udział w konsultacjach związanych z seminarium	4	
	Przygotowanie do prezentacji	20	
	RAZEM:	28	
Wskaźniki ilościowe		GODZINY	ECTS
Nakład pracy studenta związany z zajęciami wymagającymi bezpośredniego udziału nauczyciela		8	0,3
Nakład pracy studenta związany z zajęciami o charakterze praktycznym		28	1
Literatura podstawowa	1. Regulamin studiów PB		
Literatura uzupełniająca	1. Wytoczne i zalecenia w sprawie przygotowywania prac dyplomowych na Wydziale Informatyki PB		
Jednostka realizująca	Wydział Informatyki Politechniki Białostockiej	Data opracowania programu	
Program opracował	<i>dr inż. Magdalena Topczewska</i>	14.03.2019	

Wydział Informatyki									
Kierunek studiów	Data Science						Poziom i forma studiów	studia podyplomowe	
Specjalność / ścieżka dyplomowania	Developer						Profil kształcenia		
Nazwa przedmiotu	Data Science z językiem Python 2						Kod przedmiotu	SPDS11	
							Rodzaj przedmiotu	obowiązkowy	
Formy zajęć i liczba godzin	W	C	L	P	Ps	T	S	Semestr	2
					24			Punkty ECTS	3
Przedmioty wprowadzające									
Cele przedmiotu	Celem zajęć jest nabycie umiejętności zastosowania metod znanych z przedmiotu Wprowadzenie do Uczenia Maszynowego przy wykorzystaniu języka Python oraz ich zastosowania w praktycznych, rzeczywistych przykładach								
Treści programowe	Praktyczne przykłady opracowania metod znanych z przedmiotu Wprowadzenie do Uczenia Maszynowego: - biblioteka SciKit-learn i Tensorflow; - podbudowa ML-modelowej z ich zastosowaniem								
Metody dydaktyczne	1. prezentacja multimedialna; 2. analiza wybranych przypadków; 3. samodzielne opracowanie wskazanych zagadnień; 4. ćwiczenia indywidualne, ćwiczenia grupowe; 5. przygotowanie przez słuchacza samodzielnej koncepcji projektowej; 8. dyskusje dydaktyczne.								
Forma zaliczenia	Projekt końcowy lub projekty cząstkowe na zajęciach, praca domowa								
Symbol efektu uczenia się	Zakładane efekty uczenia się						Odniesienie do kierunkowych efektów uczenia się		
EU1	Zna i stosuje podstawy języka Python, składni i jego podstawowe biblioteki						DS_W04, DS_U04		
EU2	Posiada i stosuje wiedzę o wykorzystaniu Pythona pod kątem przetwarzania danych						DS_W04, DS_U04		
EU3	Potrafi posługiwać się językiem Python na poziomie podstawowym						DS_U04		
EU4	Potrafi wykorzystać podstawowe biblioteki języka Python w przetwarzaniu danych						DS_U04		
Symbol efektu uczenia się	Sposoby weryfikacji efektów uczenia się						Forma zajęć, na której zachodzi weryfikacja		

EU1	Projekt końcowy lub projekty cząstkowe na zajęciach	Ps	
EU2	Projekt końcowy lub projekty cząstkowe na zajęciach	Ps	
EU3	Praca domowa, projekty cząstkowe na zajęciach	Ps	
EU4	Praca domowa, projekty cząstkowe na zajęciach	Ps	
Bilans nakładu pracy studenta (w godzinach)		Liczba godz.	
Wyliczenie	Udział w pracowni specjalistycznej	24	
	Przygotowanie do pracowni specjalistycznej	20	
	Opracowanie sprawozdań z pracowni	10	
	Realizacja zadań domowych	24	
	RAZEM:	78	
Wskaźniki ilościowe		GODZINY	ECTS
Nakład pracy studenta związany z zajęciami wymagającymi bezpośredniego udziału nauczyciela		24	0,9
Nakład pracy studenta związany z zajęciami o charakterze praktycznym		78	3
Literatura podstawowa	1. D. Beazley, K. B. Jones: Python. Receptury. Wydawnictwo Helion, 2014 2. A. Boschetti, L. Massaron: Python. Podstawy nauki o danych., Helion, 2017		
Literatura uzupełniająca	1. Introduction to Python for Data Science edX: https://www.edx.org/course/introduction-python-data-science-microsoft-dat208x-5 2. Introduction to Data Science in Python Coursera: https://www.coursera.org/learn/python-data-analysis		
Jednostka realizująca	Wydział Informatyki Politechniki Białostockiej	Data opracowania programu	
Program opracował	<i>Łukasz Prusiel (T.J.Software/OpenTopic)</i> <i>Mateusz Rogowski (OwnedOutcomes)</i>	14.03.2019	

Wydział Informatyki									
Kierunek studiów	Data Science						Poziom i forma studiów	studia podyplomowe	
Specjalność / ścieżka dyplomowania	Developer						Profil kształcenia		
Nazwa przedmiotu	Technologie i środowisko Hadoop						Kod przedmiotu	SPDS14	
							Rodzaj przedmiotu	obowiązkowy	
Formy zajęć i liczba godzin	W	C	L	P	Ps	T	S	Semestr	2
	8				16			Punkty ECTS	4
Przedmioty wprowadzające									
Cele przedmiotu	Celem zajęć jest nabycie wiedzy oraz umiejętności wykorzystania technologii i środowiska Hadoop oraz wybranych frameworków do składowania i przetwarzania danych								
Treści programowe	<p>W: Hadoop and ecosystem; MapReduce, HDFS, Hive, Presto, Kafka, Spark (distributed file systems, programming, data warehousing and analysis, streaming).</p> <p>Ps: Najważniejsze narzędzia ekosystemu Hadoop, służące do składowania, przetwarzania i analizy ogromnych ilości danych (Big Data). Na zajęciach przedstawione zostaną podstawowe komponenty platformy Apache Hadoop (obliczenia w modelu MapReduce, system plików HDFS, system zarządzający Apache Ambari), główne narzędzia do budowy i analizy hurtowni danych (Apache Hive, Presto), narzędzia do strumieniowego przetwarzania danych (Apache Kafka, Spark Streaming) oraz zaawansowane możliwości silnika Apache Spark.</p>								
Metody dydaktyczne	1. prezentacja multimedialna; 2. analiza wybranych przypadków; 3. samodzielne opracowanie wskazanych zagadnień; 4. ćwiczenia indywidualne, ćwiczenia grupowe; 5. przygotowanie przez słuchacza samodzielnej koncepcji projektowej; 8. dyskusje dydaktyczne.								
Forma zaliczenia	<p>W: Zaliczenie pisemne</p> <p>Ps: Projekt końcowy lub projekty cząstkowe na zajęciach, praca domowa</p>								
Symbol efektu uczenia się	Zakładane efekty uczenia się						Odniesienie do kierunkowych efektów uczenia się		
EU1	Zna podstawowe własności sieci: atrybuty węzłów i relacji, gęstość, stopień						DS_W06		
EU2	Zna sposób reprezentacji danych sieciowych						DS_W06		
EU3	Potrafi wykorzystać w analizie narzędzia Hadoop, wczytywanie danych, tworzenie obiektów						DS_U06, DS_S02		
EU4	Potrafi przeprowadzić instalację środowiska Hadoop i wybranych						DS_U06, DS_S02		

	frameworków		
Symbol efektu uczenia się	Sposoby weryfikacji efektów uczenia się	Forma zajęć, na której zachodzi weryfikacja	
EU1	Zaliczenie pisemne	W	
EU2	Zaliczenie pisemne	W	
EU3	Praca domowa, projekt końcowy lub projekty cząstkowe	Ps	
EU4	Praca domowa, projekt końcowy lub projekty cząstkowe	Ps	
Bilans nakładu pracy studenta (w godzinach)		Liczba godz.	
Wyliczenie	Udział w wykładach	8	
	Udział w pracowni specjalistycznej	16	
	Przygotowanie do pracowni specjalistycznej	20	
	Udział w konsultacjach	10	
	Realizacja zadań domowych	40	
	Przygotowanie do zaliczenia	20	
	RAZEM:	114	
Wskaźniki ilościowe		GODZINY	ECTS
Nakład pracy studenta związany z zajęciami wymagającymi bezpośredniego udziału nauczyciela		24	0,9
Nakład pracy studenta związany z zajęciami o charakterze praktycznym		104	4
Literatura podstawowa	<ol style="list-style-type: none"> 1. R. Jurney: Zwinna analiza danych. Apache Hadoop dla każdego, Helion 2015. 2. A. Rajaraman, WalmartLabs, J. David Ullman: Mining of Massive Datasets., Stanford University, California 3. http://atbrox.com/2011/11/09/mapreduce-hadoop-algorithms-in-academic-papers-5th-update-%E2%80%93-nov-2011/ 		
Literatura uzupełniająca	<ol style="list-style-type: none"> 1. J. Lin, C. Dyer Data-Intensive Text Processing with MapReduce., Morgan & Claypool Publishers, 2010 		
Jednostka realizująca	Wydział Informatyki Politechniki Białostockiej	Data opracowania programu	
Program opracował	<i>mgr inż. Daniel Reska</i>	14.03.2019	

Wydział Informatyki									
Kierunek studiów	Data Science						Poziom i forma studiów	studia podyplomowe	
Specjalność / ścieżka dyplomowania	Developer						Profil kształcenia		
Nazwa przedmiotu	Sztuczne Sieci Neuronowe i Deep Learning						Kod przedmiotu	SPDS25	
							Rodzaj przedmiotu	obowiązkowy	
Formy zajęć i liczba godzin	W	Ć	L	P	Ps	T	S	Semestr	2
	16				20			Punkty ECTS	4
Przedmioty wprowadzające									
Cele przedmiotu	Celem zajęć jest nabycie wiedzy na temat kolejnych zagadnień związanych ze sztucznymi sieciami neuronowymi oraz nabycie umiejętności wykorzystania i zastosowania technik analizy i wizualizacji								
Treści programowe	<p>W: Sieci rekurencyjne: sieć Hopfielda; Sieci samoorganizujące się na zasadzie współzawodnictwa; Gaz neuronowy; Sieci samoorganizujące się typu Hebb'a; Logika rozmyta; Sieci neuronowe o logice rozmytej; Wykorzystanie sieci neuronowych w analizie przeżyć</p> <p>Ps: Perceptron; Sieci rekurencyjne: sieć Hopfielda; Sieci samoorganizujące się na zasadzie współzawodnictwa; Gaz neuronowy; Sieci samoorganizujące się typu Hebb'a; Sieci neuronowe o logice rozmytej; Wykorzystanie sieci neuronowych w analizie przeżyć</p>								
Metody dydaktyczne	1. prezentacja multimedialna; 2. analiza wybranych przypadków; 3. samodzielne opracowanie wskazanych zagadnień; 4. ćwiczenia indywidualne, ćwiczenia grupowe; 5. przygotowanie przez słuchacza samodzielnej koncepcji projektowej; 8. dyskusje dydaktyczne.								
Forma zaliczenia	<p>W: Zaliczenie pisemne</p> <p>Ps: Projekt końcowy lub projekty cząstkowe na zajęciach, sprawozdania, praca domowa</p>								
Symbol efektu uczenia się	Zakładane efekty uczenia się						Odniesienie do kierunkowych efektów uczenia się		
EU1	Zna różne modele sieci neuronowych i sposoby ich działania						DS_W03, DS_S03		
EU2	Zna ograniczenia związane w wykorzystaniem sieci neuronowych						DS_W03, DS_S03		
EU3	Potrafi przeprowadzić analizę danych (klasyfikacja, regresja) przy użyciu SSN						DS_U03, DS_U05, DS_S03		
EU4	Potrafi dokonać właściwego doboru typu i architektury sieci						DS_U03, DS_S03		
Symbol efektu uczenia się	Sposoby weryfikacji efektów uczenia się						Forma zajęć, na której zachodzi weryfikacja		

EU1	Zaliczenie pisemne	W	
EU2	Zaliczenie pisemne	W	
EU3	Praca domowa, projekt końcowy lub projekty cząstkowe, sprawozdania	Ps	
EU4	Praca domowa, projekt końcowy lub projekty cząstkowe, sprawozdania	Ps	
Bilans nakładu pracy studenta (w godzinach)		Liczba godz.	
Wyliczenie	Udział w wykładach	16	
	Udział w pracowni specjalistycznej	20	
	Przygotowanie do pracowni specjalistycznej	10	
	Opracowanie sprawozdań z pracowni	20	
	Realizacja zadań domowych	25	
	Przygotowanie do zaliczenia	10	
	RAZEM:		101
Wskaźniki ilościowe		GODZINY	ECTS
Nakład pracy studenta związany z zajęciami wymagającymi bezpośredniego udziału nauczyciela		36	1,25
Nakład pracy studenta związany z zajęciami o charakterze praktycznym		101	4
Literatura podstawowa	<ol style="list-style-type: none"> 1. R. Tadeusiewicz: Sieci neuronowe, Akademicka Oficyna Wydawnicza RM, Warszawa 1993 2. J. Żurada, M. Barski, W. Jędruch: Sztuczne sieci neuronowe, Wydawnictwo Naukowe PWN, Warszawa 1996 3. S. Osowski: Sieci neuronowe w ujęciu algorytmicznym, Wydawnictwo Naukowo-Techniczne, Warszawa 1996 		
Literatura uzupełniająca	<ol style="list-style-type: none"> 1. J. Hertz, A. Krogh, R. G. Palmer: Wstęp do teorii obliczeń neuronowych, Wydawnictwa Naukowo-Techniczne, Warszawa, 1993 2. J. Korbicz, A. Obuchowicz, D. Uciński: Sztuczne sieci neuronowe. Podstawy i zastosowania, Akademicka Oficyna Wydawnicza PLJ, Warszawa 1994 		
Jednostka realizująca	Wydział Informatyki Politechniki Białostockiej	Data opracowania programu	
Program opracował	dr inż. Małgorzata Krętowska	14.03.2019	

Wydział Informatyki									
Kierunek studiów	Data Science						Poziom i forma studiów	studia podyplomowe	
Specjalność / ścieżka dyplomowania	Developer						Profil kształcenia		
Nazwa przedmiotu	Kontenery w nowoczesnych środowiskach programistycznych i produkcyjnych - Docker, Kubernetes, Kafka						Kod przedmiotu	SPDS15	
							Rodzaj przedmiotu	obowiązkowy	
Formy zajęć i liczba godzin	W	C	L	P	Ps	T	S	Semestr	2
					8			Punkty ECTS	2
Przedmioty wprowadzające									
Cele przedmiotu	Nabycie umiejętności pracy w nowoczesnym środowisku programistycznym. Rozumienie zaawansowanych zagadnień procesu wytwarzania oprogramowania. Umiejętność efektywnego zarządzania logami aplikacji. Umiejętność dostarczania oprogramowania w trybie Continuous Delivery.								
Treści programowe	<ul style="list-style-type: none"> - Change Management (Gerrit) - Static Code Analysis (Sonar) - Configuration Management (Salt, Ansible) - Log Management (Elastic, Logstash, Kibana) - Continuous Delivery - Continuous Deployment 								
Metody dydaktyczne	1. prezentacja multimedialna; 2. analiza wybranych przypadków; 3. samodzielne opracowanie wskazanych zagadnień; 4. ćwiczenia indywidualne, ćwiczenia grupowe; 5. przygotowanie przez słuchacza samodzielnej koncepcji projektowej; 8. dyskusje dydaktyczne.								
Forma zaliczenia	Projekt końcowy lub projekty cząstkowe na zajęciach, sprawozdania, praca domowa								
Symbol efektu uczenia się	Zakładane efekty uczenia się						Odniesienie do kierunkowych efektów uczenia się		
EU1	Zna i stosuje różne sposoby przeprowadzania obliczeń równoległych i konfiguracji nowoczesnych środowisk programistycznych						DS_W06, DS_U06		
EU2	Zna i stosuje metody dostarczania oprogramowania w trybie Continuous Delivery						DS_W06, DS_U06		
EU3	Potrafi efektywnie zarządzać konfiguracją środowiska programistycznego						DS_U06, DS_S03		

EU4	Potrafi analizować logi i dostarczać oprogramowanie w trybie Continuous Delivery	DS_U06, DS_S03	
Symbol efektu uczenia się	Sposoby weryfikacji efektów uczenia się	Forma zajęć, na której zachodzi weryfikacja	
EU1	Projekt końcowy lub projekty cząstkowe na zajęciach	Ps	
EU2	Projekt końcowy lub projekty cząstkowe na zajęciach	Ps	
EU3	Praca domowa, projekty cząstkowe na zajęciach	Ps	
EU4	Praca domowa, projekty cząstkowe na zajęciach	Ps	
Bilans nakładu pracy studenta (w godzinach)		Liczba godz.	
Wyliczenie	Udział w pracowni specjalistycznej	8	
	Przygotowanie do pracowni specjalistycznej	15	
	Wykonanie sprawozdań	15	
	Realizacja zadań domowych	15	
	RAZEM:	53	
Wskaźniki ilościowe		GODZINY	ECTS
Nakład pracy studenta związany z zajęciami wymagającymi bezpośredniego udziału nauczyciela		8	0,3
Nakład pracy studenta związany z zajęciami o charakterze praktycznym		53	2
Literatura podstawowa	1. M. Serafin, Wirtualizacja w Praktyce, Helion, 2011. 2. Dokumentacja VirtualBox. https://www.virtualbox.org 3. Dokumentacja Docker. https://www.docker.com		
Literatura uzupełniająca	1. Dokumentacja FUSE. http://fuse.sourceforge.net 2. Dokumentacja LXC. https://linuxcontainers.org 3. Dokumentacja KVM. http://www.linux-kvm.org		
Jednostka realizująca	Wydział Informatyki Politechniki Białostockiej	Data opracowania programu	
Program opracował	Piotr Szwed	14.03.2019	

Wydział Informatyki									
Kierunek studiów	Data Science						Poziom i forma studiów	studia podyplomowe	
Specjalność / ścieżka dyplomowania	Developer						Profil kształcenia		
Nazwa przedmiotu	Hadoop Machine Learning Frameworks						Kod przedmiotu	SPDS16	
							Rodzaj przedmiotu	obowiązkowy	
Formy zajęć i liczba godzin	W	C	L	P	Ps	T	S	Semestr	2
					8			Punkty ECTS	2
Przedmioty wprowadzające									
Cele przedmiotu	Nabywanie wiedzy oraz umiejętności stosowania i wykorzystania najważniejszych frameworków do uczenia maszynowego, wykorzystywanych w ekosystemie Hadoop, służącym do przetwarzania i analizy ogromnych ilości danych (Big Data)								
Treści programowe	W ramach zajęć przewidziane jest przedstawienie narzędzi jak Apache Spark MLib, Apache Mahout oraz Oryx, a także praktyczne zastosowanie wybranych bibliotek w ramach pracowni specjalistycznej								
Metody dydaktyczne	1. prezentacja multimedialna; 2. analiza wybranych przypadków; 3. samodzielne opracowanie wskazanych zagadnień; 4. ćwiczenia indywidualne, ćwiczenia grupowe; 5. przygotowanie przez słuchacza samodzielnej koncepcji projektowej; 8. dyskusje dydaktyczne.								
Forma zaliczenia	Projekt końcowy lub projekty cząstkowe na zajęciach, praca domowa								
Symbol efektu uczenia się	Zakładane efekty uczenia się						Odniesienie do kierunkowych efektów uczenia się		
EU1	Zna i stosuje różne sposoby przeprowadzania obliczeń równoległych (CPU / CUDA)						DS_W06, DS_U06		
EU2	Zna i stosuje metody optymalizacji kodu wynikowego						DS_W06, DS_U06		
EU3	Potrafi stosować współbieżność (MPI) oraz przeprowadzać obliczenia równoległe (CPU / CUDA)						DS_U06, DS_S03		
EU4	Potrafi przeprowadzić analizę zużycia zasobów (RAM, CPU).						DS_U06, DS_S03		
Symbol efektu uczenia się	Sposoby weryfikacji efektów uczenia się						Forma zajęć, na której zachodzi weryfikacja		
EU1	Projekt końcowy lub projekty cząstkowe na zajęciach						Ps		
EU2	Projekt końcowy lub projekty cząstkowe na zajęciach						Ps		

EU3	Praca domowa, projekty cząstkowe na zajęciach	Ps	
EU4	Praca domowa, projekty cząstkowe na zajęciach	Ps	
Bilans nakładu pracy studenta (w godzinach)		Liczba godz.	
Wyliczenie	Udział w pracowni specjalistycznej	8	
	Przygotowanie do pracowni specjalistycznej	15	
	Wykonanie sprawozdań	15	
	Realizacja zadań domowych	20	
	RAZEM:	58	
Wskaźniki ilościowe		GODZINY	ECTS
Nakład pracy studenta związany z zajęciami wymagającymi bezpośredniego udziału nauczyciela		8	0,3
Nakład pracy studenta związany z zajęciami o charakterze praktycznym		58	2
Literatura podstawowa	<ol style="list-style-type: none"> 1. R. Jurney: Zwinna analiza danych. Apache Hadoop dla każdego, Helion 2015 2. A. Rajaraman, WalmartLabs, J. David Ullman: Mining of Massive Datasets., Stanford University, California 3. http://atbrox.com/2011/11/09/mapreduce-hadoop-algorithms-in-academic-papers-5th-update-%E2%80%93-nov-2011/ 		
Literatura uzupełniająca	<ol style="list-style-type: none"> 1. J. Lin, C. Dyer: Data-Intensive Text Processing with MapReduce. Morgan & Claypool Publishers, 2010 		
Jednostka realizująca	Wydział Informatyki Politechniki Białostockiej	Data opracowania programu	
Program opracował	<i>Daniel Reska</i> <i>Paweł Kucharski</i>		14.03.2019

Wydział Informatyki										
Kierunek studiów	Data Science							Poziom i forma studiów	studia podyplomowe	
Specjalność / ścieżka dyplomowania	Analityk danych							Profil kształcenia		
Nazwa przedmiotu	Rynek badań, analiz i data science w Polsce i na świecie							Kod przedmiotu	SPDS19	
								Rodzaj przedmiotu	obowiązkowy	
Formy zajęć i liczba godzin	W	Ć	L	P	Ps	T	S	Semestr	2	
	8							Punkty ECTS	1	
Przedmioty wprowadzające										
Cele przedmiotu	Przedstawienie rynku big data i otoczenia biznesowego									
Treści programowe	Wielkość i charakterystyka rynków. Modele biznesowe występujące na rynku badań i analiz. Sprzedaż i odbiorcy rozwiązań analitycznych. Przykłady rozwiązań analitycznych wykorzystywanych w marketingu i sprzedaży. Praca dla analityków w polskich realia: oczekiwania, warunki, stawki									
Metody dydaktyczne	1. prezentacja multimedialna; 2. analiza wybranych przypadków; 3. samodzielne opracowanie wskazanych zagadnień; 4. ćwiczenia indywidualne, ćwiczenia grupowe; 5. przygotowanie przez słuchacza samodzielnej koncepcji projektowej; 8. dyskusje dydaktyczne.									
Forma zaliczenia	Zaliczenie ustne lub test pisemny									
Symbol efektu uczenia się	Zakładane efekty uczenia się							Odniesienie do kierunkowych efektów uczenia się		
EU1	Zna podstawowe modele biznesowe stosowane na rynku usług Big Data							DS_W02, DS_W05		
EU2	Zna sposoby świadczenia usługi charakterystyce rynku							DS_W02, DS_W05		
EU3	Posiada wiedzę o sprzedaży i odbiorcy rozwiązań analitycznych przy wprowadzaniu nowego produktu							DS_W02, DS_W05, DS_S01		
EU4	Posiada wiedzę o doborze narzędzi analitycznych wykorzystywanych w marketingu i sprzedaży							DS_W02, DS_W05		
Symbol efektu uczenia się	Sposoby weryfikacji efektów uczenia się							Forma zajęć, na której zachodzi weryfikacja		
EU1	Zaliczenie ustne lub test pisemny							W		

EU2	Zaliczenie ustne lub test pisemny	W	
EU3	Zaliczenie ustne lub test pisemny	W	
EU4	Zaliczenie ustne lub test pisemny	W	
Bilans nakładu pracy studenta (w godzinach)		Liczba godz.	
Wyliczenie	Udział w wykładach	8	
	Przygotowanie do wykładu	10	
	Przygotowanie do zaliczenia	10	
	RAZEM:	28	
Wskaźniki ilościowe		GODZINY	ECTS
Nakład pracy studenta związany z zajęciami wymagającymi bezpośredniego udziału nauczyciela		8	0,3
Nakład pracy studenta związany z zajęciami o charakterze praktycznym		28	1
Literatura podstawowa	<ol style="list-style-type: none"> Specjaliści data science SAS : https://www.sas.com/pl_pl/insights/analytics/specjalisci-data-science.html Tracking down answers to your questions about data scientists SAS : https://www.sas.com/en_us/insights/articles/analytics/tracking-down-answers-to-your-questions-about-data-scientists.html Platformy Data Science najszybciej rozwijającym się segmentem analityki biznesowej – ccnews.pl : http://ccnews.pl/2017/04/11/platformy-data-science-najszybciej-rozwijajacym-sie-segmentem-analityki-biznesowej/ 		
Literatura uzupełniająca	<ol style="list-style-type: none"> Preparing a new generation for leadership in a big data world SAS : https://www.sas.com/en_us/insights/articles/analytics/preparing-a-new-generation-for-leadership-in-a-big-data-world.html 4 Major Trends Disrupting the Data Science Market : http://www.kdnuggets.com/2016/07/4-trends-disrupting-data-science-market.html 		
Jednostka realizująca	Wydział Informatyki Politechniki Białostockiej	Data opracowania programu	
Program opracował	Jan Zając	14.03.2019	

Wydział Informatyki									
Kierunek studiów	Data Science						Poziom i forma studiów	studia podyplomowe	
Specjalność / ścieżka dyplomowania	Analityk danych						Profil kształcenia		
Nazwa przedmiotu	Język R jako narzędzie wizualizacji i raportowania						Kod przedmiotu	SPDS21	
							Rodzaj przedmiotu	obowiązkowy	
Formy zajęć i liczba godzin	W	C	L	P	Ps	T	S	Semestr	2
					16			Punkty ECTS	2
Przedmioty wprowadzające									
Cele przedmiotu	Nabycie umiejętności wykorzystania kodowania oraz specjalizowanych bibliotek języka R do raportowania i sposobów i narzędzi do wizualizacji								
Treści programowe	1) Biblioteka do R – Shiny 2) Apache Zeppelin on Spark Możliwości wizualizacji w Zeppelin. Zeppelin a SQL (SparkSQL) i inne języki.								
Metody dydaktyczne	1. prezentacja multimedialna; 2. analiza wybranych przypadków; 3. samodzielne opracowanie wskazanych zagadnień; 4. ćwiczenia indywidualne, ćwiczenia grupowe; 5. przygotowanie przez słuchacza samodzielnej koncepcji projektowej; 8. dyskusje dydaktyczne.								
Forma zaliczenia	Projekt końcowy lub projekty cząstkowe na zajęciach, praca domowa								
Symbol efektu uczenia się	Zakładane efekty uczenia się						Odniesienie do kierunkowych efektów uczenia się		
EU1	Zna i stosuje podstawowe własności danych wielorelacyjnych: atrybuty węzłów i relacji, gęstość, stopień						DS_W02, DS_W05, DS_U02		
EU2	Zna i stosuje sposoby reprezentacji danych wielorelacyjnych i sieciowych						DS_W02, DS_W05, DS_U02		
EU3	Potrafi wykorzystać w analizie sieci pakietów R, wczytywać dane, tworzyć obiekty						DS_U02, DS_U05, DS_S01		
EU4	Potrafi wizualizować dane i sieci						DS_U02, DS_U05, DS_S01		
Symbol efektu uczenia się	Sposoby weryfikacji efektów uczenia się						Forma zajęć, na której zachodzi weryfikacja		

EU1	Projekt końcowy lub projekty cząstkowe na zajęciach	Ps	
EU2	Projekt końcowy lub projekty cząstkowe na zajęciach.	Ps	
EU3	Praca domowa, projekty cząstkowe	Ps	
EU4	Praca domowa, projekty cząstkowe	Ps	
Bilans nakładu pracy studenta (w godzinach)		Liczba godz.	
Wyliczenie	Udział w pracowni specjalistycznej	16	
	Opracowanie sprawozdań z pracowni i/lub wykonanie zadań domowych (prac domowych)	10	
	Przygotowanie projektu semestralnego	20	
	Realizacja zadań domowych	10	
	RAZEM:	56	
Wskaźniki ilościowe		GODZINY	ECTS
Nakład pracy studenta związany z zajęciami wymagającymi bezpośredniego udziału nauczyciela		16	0,5
Nakład pracy studenta związany z zajęciami o charakterze praktycznym		56	2
Literatura podstawowa	<ol style="list-style-type: none"> 1. J. Scott: Social Network Analysis Newbury Park CA: Sage, 2000 2. R. A. Hanneman, M. Riddle: Introduction to Social Network Methods, 2005 3. S. Wasserman, K. Faust: Social Network Analysis: Methods and Applications. Cambridge University Press. 1994 		
Literatura uzupełniająca	<ol style="list-style-type: none"> 1. W. de Nooy, A. Mrvar, V. Batagelj: Exploratory Social Network Analysis with Pajek. Cambridge University Press, 2005 		
Jednostka realizująca	Wydział Informatyki Politechniki Białostockiej	Data opracowania programu	
Program opracował	<i>Dominik Batorski</i>	14.03.2019	

Wydział Informatyki									
Kierunek studiów	Data Science						Poziom i forma studiów	studia podyplomowe	
Specjalność / ścieżka dyplomowania	Analityk danych						Profil kształcenia		
Nazwa przedmiotu	Interfejsy API w przetwarzaniu danych - rozpoznawanie obrazu i przetwarzanie języka naturalnego						Kod przedmiotu	SPDS27	
							Rodzaj przedmiotu	obowiązkowy	
Formy zajęć i liczba godzin	W	C	L	P	Ps	T	S	Semestr	2
	8				16			Punkty ECTS	4
Przedmioty wprowadzające									
Cele przedmiotu	Celem zajęć jest nabycie wiedzy na temat pracy z API i możliwości i umiejętność wykorzystania we współczesnym modelu tworzenia aplikacji oraz przedstawienie możliwości na wybranym przykładzie								
Treści programowe	<p>W: Coraz częściej w analizie danych wykorzystywane są różnego rodzaju interfejsy programistyczne aplikacji tzw. API. Pozwalają one na dostęp do danych, ale również ich przetwarzanie. Celem tego kursu jest zaznajomienie uczestników i nauczanie pracy z różnymi API, które każdy data scientist powinien znać. Lista zagadnień poruszanych na wykładzie i warsztatach obejmuje: Wprowadzenie do API; Dostęp do danych na przykładzie API Quandl, Twitter, Facebook; Wykorzystanie API w rozpoznawaniu obrazu. Wykorzystanie API w przetwarzaniu języka naturalnego. API w uczeniu maszynowym</p> <p>Ps: Wprowadzenie do API; Dostęp do danych na przykładzie API Quandl, Twitter, Facebook; Wykorzystanie API w rozpoznawaniu obrazu. Wykorzystanie API w przetwarzaniu języka naturalnego. API w uczeniu maszynowym</p>								
Metody dydaktyczne	1. prezentacja multimedialna; 2. analiza wybranych przypadków; 3. samodzielne opracowanie wskazanych zagadnień; 4. ćwiczenia indywidualne, ćwiczenia grupowe; 5. przygotowanie przez słuchacza samodzielnej koncepcji projektowej; 8. dyskusje dydaktyczne.								
Forma zaliczenia	<p>W: Zaliczenie ustne lub test pisemny</p> <p>Ps: Projekt końcowy lub projekty cząstkowe na zajęciach, praca domowa</p>								
Symbol efektu uczenia się	Zakładane efekty uczenia się						Odniesienie do kierunkowych efektów uczenia się		
EU1	Zna podstawowe zastosowania API i model budowania rozwiązań API Economy						DS_W05, DS_S02		
EU2	Zna możliwości wybranych API w przetwarzaniu obrazów i języka naturalnego						DS_W05, DS_S02		
EU3	Potrafi wykorzystać wybrane API do implementacji podstawowych						DS_U05, DS_S02		

	operacji na obrazach	
EU4	Potrafi wykorzystać wybrane API do implementacji podstawowych operacji z NLP	DS_U05, DS_S02
Symbol efektu uczenia się	Sposoby weryfikacji efektów uczenia się	Forma zajęć, na której zachodzi weryfikacja
EU1	Zaliczenie ustne lub test pisemny	W
EU2	Zaliczenie ustne lub test pisemny	W
EU3	Praca domowa, projekt końcowy lub projekty cząstkowe	Ps
EU4	Praca domowa, projekt końcowy lub projekty cząstkowe	Ps
Bilans nakładu pracy studenta (w godzinach)		Liczba godz.
Wyliczenie	Udział w wykładach	8
	Udział w pracowni specjalistycznej	16
	Przygotowanie do pracowni specjalistycznej	20
	Opracowanie sprawozdań z pracowni	20
	Przygotowanie projektu semestralnego	20
	Realizacja zadań domowych	17
	RAZEM:	101
Wskaźniki ilościowe		GODZINY
Nakład pracy studenta związany z zajęciami wymagającymi bezpośredniego udziału nauczyciela		24
Nakład pracy studenta związany z zajęciami o charakterze praktycznym		101
Literatura podstawowa	<ol style="list-style-type: none"> 1. G. Golemund: R for Data Science. O'Reilly Media, 2015 2. M. Szeliga: Data Science i uczenie maszynowe. Wydawnictwo Naukowe PWN, 2017 3. H. Wickham, G. Golemund: Język R. Kompletny zestaw narzędzi dla analityków danych. Helion, 2017 	
Literatura uzupełniająca	<ol style="list-style-type: none"> 1. 5 Best Machine Learning APIs for Data Science, KDnuggets 2. https://www.kdnuggets.com/2015/11/machine-learning-apis-data-science.html 3. Bootstrapping Machine Learning, The first guide to Prediction APIs, Luis Dorard, 2017 4. Google Vision API: Image Analysis as a Service : https://cloudacademy.com/blog/google-vision-api-image-analysis/ 5. Computer Vision API - Image Processing Microsoft Azure : https://azure.microsoft.com/en-us/services/cognitive-services/computer-vision/ 	
Jednostka realizująca	Wydział Informatyki Politechniki Białostockiej	Data opracowania programu
Program opracował	<i>dr inż. Paweł Tadejko, Dominik Batorski</i>	14.03.2019

Wydział Informatyki									
Kierunek studiów	Data Science						Poziom i forma studiów	studia podyplomowe	
Specjalność / ścieżka dyplomowania	Analityk danych						Profil kształcenia		
Nazwa przedmiotu	Zaawansowane narzędzia analityczne						Kod przedmiotu	SPDS23	
							Rodzaj przedmiotu	obowiązkowy	
Formy zajęć i liczba godzin	W	C	L	P	Ps	T	S	Semestr	2
					20			Punkty ECTS	3
Przedmioty wprowadzające									
Cele przedmiotu	Nabycie umiejętności i wiedzy na temat wybranych możliwości zaawansowanych narzędzi ekosystemu Apache Spark								
Treści programowe	Narzędzia analityczne – Apache Zeppelin (Spark) – na danych Google. Proces przetwarzania danych w Zeppelinie. Podstawowe komponenty środowiska Zeppelin UI. Współpraca Zeppelina z Apache Spark.								
Metody dydaktyczne	1. prezentacja multimedialna; 2. analiza wybranych przypadków; 3. samodzielne opracowanie wskazanych zagadnień; 4. ćwiczenia indywidualne, ćwiczenia grupowe; 5. przygotowanie przez słuchacza samodzielnej koncepcji projektowej; 8. dyskusje dydaktyczne.								
Forma zaliczenia	Projekt końcowy lub projekty cząstkowe na zajęciach, praca domowa								
Symbol efektu uczenia się	Zakładane efekty uczenia się						Odniesienie do kierunkowych efektów uczenia się		
EU1	Zna i stosuje podstawowe możliwości analizy danych za pomocą wybranych narzędzi						DS_W02, DS_W05, DS_U02		
EU2	Zna, stosuje i wykorzystuje możliwości analizy i wizualizacji danych Apache Zeppelin						DS_U02, DS_U05, DS_U05		
EU3	Potrafi wykorzystać w analizie danych aplikacje front-end na system Apache Spark						DS_U02, DS_U05, DS_S02		
EU4	Potrafi wizualizować dane za pomocą zapytań Zeppelin i SQL						DS_U02, DS_U05, DS_S03		
Symbol efektu uczenia się	Sposoby weryfikacji efektów uczenia się						Forma zajęć, na której zachodzi weryfikacja		

EU1	Projekt końcowy lub projekty cząstkowe na zajęciach	Ps	
EU2	Projekt końcowy lub projekty cząstkowe na zajęciach	Ps	
EU3	Praca domowa, ocena projekty cząstkowe	Ps	
EU4	Praca domowa, projekty cząstkowe	Ps	
Bilans nakładu pracy studenta (w godzinach)		Liczba godz.	
Wyliczenie	Udział w pracowni specjalistycznej	20	
	Opracowanie sprawozdań z pracowni	20	
	Realizacja zadań domowych	40	
	RAZEM:	80	
Wskaźniki ilościowe		GODZINY	ECTS
Nakład pracy studenta związany z zajęciami wymagającymi bezpośredniego udziału nauczyciela		40	1,5
Nakład pracy studenta związany z zajęciami o charakterze praktycznym		80	3
Literatura podstawowa	1. Introduction to Data Science with Apache Spark and Zeppelin : https://hortonworks.com/blog/introduction-to-data-science-with-apache-spark/ 2. Apache Zeppelin · Mastering Apache Spark 2 : https://jaceklaskowski.gitbooks.io/mastering-apache-spark/content/interactive-notebooks/apache-zeppelin.html 3. Spark. Zaawansowana analiza danych, Sandy Ryza, Uri Laserson, Sean Owen, Josh Wills, Helion, 2015		
Literatura uzupełniająca	1. Big Data visualization with Apache Spark and Zeppelin : https://www.slideshare.net/prajods/big-data-visualization-with-apache-spark-and-zeppelin 2. Open-Source Big Data Visualization: Learning Apache Zeppelin on Spark - DZone Big Data : https://dzone.com/articles/best-apache-spark-presentations		
Jednostka realizująca	Wydział Informatyki Politechniki Białostockiej	Data opracowania programu	
Program opracował	<i>mgr inż. Łukasz Prusiel</i> <i>dr inż. Magdalena Topczewska</i>	18.03.2019	

Wydział Informatyki									
Kierunek studiów	Data Science						Poziom i forma studiów	studia podyplomowe	
Specjalność / ścieżka dyplomowania	Analityk danych						Profil kształcenia		
Nazwa przedmiotu	Język SQL dla Data Science						Kod przedmiotu	SPDS24	
							Rodzaj przedmiotu	obowiązkowy	
Formy zajęć i liczba godzin	W	Ć	L	P	Ps	T	S	Semestr	2
	8				16			Punkty ECTS	4
Przedmioty wprowadzające	Nabycie wiedzy o podstawowych narzędziach raportowania oraz języku SQL. Nabycie umiejętności wykorzystania i zastosowania arkusza kalkulacyjnego oraz pisania zapytań SQL								
Cele przedmiotu	Przedmiot obejmuje zagadnienia związane z językiem SQL dla baz operacyjnych (OLTP): zapytania proste, zapytania grupujące, podzapytania, jak również rozszerzenie języka SQL dla baz danych typu OLAP: zapytania z operatorami grupowania wielokrotnego, zaawansowane funkcje analityczne.								
Treści programowe	Zapytania proste. Zapytania grupujące. Podzapytania. SQL dla OLAP: operatory grupowania wielokrotnego. SQL dla OLAP: zaawansowane funkcje analityczne.								
Metody dydaktyczne	1. prezentacja multimedialna; 2. analiza wybranych przypadków; 3. samodzielne opracowanie wskazanych zagadnień; 4. ćwiczenia indywidualne, ćwiczenia grupowe; 5. przygotowanie przez słuchacza samodzielnej koncepcji projektowej; 8. dyskusje dydaktyczne.								
Forma zaliczenia	W: Zaliczenie ustne lub test pisemny								
	Ps: Projekt końcowy lub projekty cząstkowe na zajęciach, praca domowa								
Symbol efektu uczenia się	Zakładane efekty uczenia się						Odniesienie do kierunkowych efektów uczenia się		
EU1	Zna składnię zapytania SQL z podzapytaniami i zapytaniami grupującymi						DS_W06		
EU2	Ma wiedzę jak korzystać z operatorów grupowania wielokrotnego						DS_W06		
EU3	Potrafi formułować zapytania SQL z użyciem podzapytań i zapytań grupujących, funkcji analitycznych						DS_U06, DS_S03		
EU4	Potrafi zastosować poprawnie rozszerzenia języka SQL dla OLAP						DS_U06, DS_S03		
Symbol efektu uczenia się	Sposoby weryfikacji efektów uczenia się						Forma zajęć, na której zachodzi weryfikacja		
EU1	Zaliczenie ustne lub test pisemny						W		
EU2	Zaliczenie ustne lub test pisemny						W		

EU3	Projekt końcowy lub projekty cząstkowe na zajęciach, praca domowa	Ps	
EU4	Projekt końcowy lub projekty cząstkowe na zajęciach, praca domowa	Ps	
Bilans nakładu pracy studenta (w godzinach)		Liczba godz.	
Wyliczenie	Udział w wykładach	8	
	Udział w pracowni specjalistycznej	16	
	Przygotowanie do pracowni specjalistycznej	20	
	Opracowanie sprawozdań z pracowni	20	
	Udział w konsultacjach związanych z wykładem/pracownią specjalistyczną	10	
	Realizacja zadań domowych	30	
	RAZEM:		104
Wskaźniki ilościowe		GODZINY	ECTS
Nakład pracy studenta związany z zajęciami wymagającymi bezpośredniego udziału nauczyciela		24	0,9
Nakład pracy studenta związany z zajęciami o charakterze praktycznym		104	4
Literatura podstawowa	<ol style="list-style-type: none"> 1. T. Connolly, C. Begg: Database Systems: A Practical Approach to Design, Implementation, and Management. 5th edition, Addison-Wesley, 2009 2. J. Casteel: Oracle 11g: SQL. Course Technology, 2010 3. S. Mishra, A. Beaulieu: Mastering Oracle SQL, O'Reilly Media, 2004 		
Literatura uzupełniająca	<ol style="list-style-type: none"> 1. J. Murach. Murach's Oracle SQL and PL/SQL for Developers, Mike Murach & Associates, 2014 		
Jednostka realizująca	Wydział Informatyki Politechniki Białostockiej	Data opracowania programu	
Program opracował	<i>dr hab. inż. Agnieszka Drużdżel</i>	18.03.2019	

Wydział Informatyki									
Kierunek studiów	Data Science						Poziom i forma studiów	studia podyplomowe	
Specjalność / ścieżka dyplomowania	Analityk danych						Profil kształcenia		
Nazwa przedmiotu	Dedykowane narzędzia analityczne						Kod przedmiotu	SPDS28	
							Rodzaj przedmiotu	obowiązkowy	
Formy zajęć i liczba godzin	W	C	L	P	Ps	T	S	Semestr	1
					8			Punkty ECTS	1
Przedmioty wprowadzające									
Cele przedmiotu	Celem przedmiotu jest nabycie wiedzy i umiejętności zastosowania wybranego współczesnego narzędzia do agregacji i wizualizacji danych, np. Tableau, Qlik, Microsoft PowerPivot, w szczególności łączenia danych z różnych źródeł i metod efektywnej wizualizacji danych								
Treści programowe	Łączenie danych z różnych źródeł oraz ich importu do narzędzia. Metody efektywnej wizualizacji danych. Prezentacja danych na mapach geograficznych. Zastosowanie grupowania, sortowania i tworzenia dynamicznych filtrów. Tworzenia raportów tabelarycznych z formatowaniem warunkowym. Wykorzystania parametrów w kalkulacjach. Budowa interaktywnych dashboardów.								
Metody dydaktyczne	1. prezentacja multimedialna; 2. analiza wybranych przypadków; 3. ćwiczenia indywidualne, ćwiczenia grupowe; 4. dyskusje dydaktyczne.								
Forma zaliczenia	Projekt końcowy lub projekty cząstkowe na zajęciach, praca domowa								
Symbol efektu uczenia się	Zakładane efekty uczenia się						Odniesienie do kierunkowych efektów uczenia się		
EU1	Zna i stosuje możliwości współczesnych dedykowanych narzędzi do agregacji i wizualizacji danych						DS_W06, DS_U06		
EU2	Zna i umie stosować idee działania i kategorie funkcjonalności współczesnych narzędzi analitycznych						DS_W06, DS_U06		
EU3	Potrafi stosować łączenie danych z różnych źródeł, ich importu oraz metod efektywnej wizualizacji						DS_U06, DS_S03		
EU4	Potrafi analizować i tworzy raporty tabelaryczne						DS_U06, DS_S03		
Symbol efektu uczenia się	Sposoby weryfikacji efektów uczenia się						Forma zajęć, na której zachodzi weryfikacja		
EU1	Projekty cząstkowe na zajęciach						Ps		

EU2	Projekty cząstkowe na zajęciach	Ps	
EU3	Praca domowa, projekty cząstkowe	Ps	
EU4	Praca domowa, projekty cząstkowe	Ps	
Bilans nakładu pracy studenta (w godzinach)		Liczba godz.	
Wyliczenie	Udział w pracowni specjalistycznej	8	
	Przygotowanie do pracowni specjalistycznej	10	
	Opracowanie sprawozdań z pracowni i/lub wykonanie zadań domowych	10	
	RAZEM:	28	
Wskaźniki ilościowe		GODZINY	ECTS
Nakład pracy studenta związany z zajęciami wymagającymi bezpośredniego udziału nauczyciela		18	0,6
Nakład pracy studenta związany z zajęciami o charakterze praktycznym		28	1
Literatura podstawowa	<ol style="list-style-type: none"> 1. D. G. Murray: Tableau Your Data. Wiley, 2016 2. Tableau 10 A-Z: Hands-On Tableau Training For Data Science! Udemy, https://www.udemy.com/tableau10/ 3. B. Jones: Communicating Data with Tableau. O'Reilly Media, 2014 		
Literatura uzupełniająca	<ol style="list-style-type: none"> 1. Tableau Training and Tutorials : https://www.tableau.com/learn/training 2. Qlikview - Learning Path From A Starter To A QlikView Expert! : https://www.analyticsvidhya.com/learning-paths-data-science-business-analytics-business-intelligence-big-data/qlikview-learning-path/ 3. Power Pivot for Excel Tutorial: Top Use Cases and Examples Toptal : https://www.toptal.com/finance/excel-experts/powerpivot-for-excel-tutorial-top-use-cases-and-examples 		
Jednostka realizująca	Wydział Informatyki Politechniki Białostockiej	Data opracowania programu	
Program opracował	<i>dr inż. Paweł Tadejko</i>	18.03.2019	