

KARTA KURSU (realizowanego w module specjalności)
Data Science

Nazwa	Przetwarzanie języka naturalnego
Nazwa w j. ang.	Natural Language Processing

Koordynator	Dr hab. J. Kapusta, prof. UKEN	Zespół dydaktyczny
		Dr hab. J. Kapusta, prof. UKEN
Punktacja ECTS*	3	

Opis kursu (cele kształcenia)

Pogłębienie wiedzy i rozwijanie umiejętności w zakresie zaawansowanych metod **przetwarzania języka naturalnego (NLP)**, w szczególności analizy danych tekstowych z wykorzystaniem technik uczenia maszynowego i narzędzi stosowanych w Data Science.

Warunki wstępne

Wiedza	
Umiejętności	
Kursy	

Efekty uczenia się

Wiedza	Efekt uczenia się dla kursu	Odniesienie do efektów kierunkowych
	Po zakończeniu kursu student: W01:	SD_W06

Umiejętności	Efekt uczenia się dla kursu	Odniesienie do efektów kierunkowych
	Po zakończeniu kursu student: U01:	SD_U03

Kompetencje społeczne	Efekt uczenia się dla kursu	Odniesienie do efektów kierunkowych
	Po zakończeniu kursu student: K01:	SD_K02 SD_K03

Studia stacjonarne

Organizacja							
Forma zajęć	Wykład (W)	Ćwiczenia w grupach					
		A	K	L	S	P	E
Liczba godzin	15			30			

Studia niestacjonarne

Organizacja							
Forma zajęć	Wykład (W)	Ćwiczenia w grupach					
		A	K	L	S	P	E
Liczba godzin	10			20			

Opis metod prowadzenia zajęć

Kurs prowadzony jest w formie zajęć teoretycznych i praktycznych, łącząc wykłady, laboratoria z zadaniami indywidualnymi oraz projekty grupowe.

Formy sprawdzania efektów uczenia się

	E – learning	Gry dydaktyczne	Ćwiczenia w szkole	Zajęcia terenowe	Praca laboratoryjna	Projekt indywidualny	Projekt grupowy	Udział w dyskusji	Referat	Praca pisemna (esej)	Egzamin ustny	Egzamin pisemny	Inne
W01					X	X	X	X					
W02					X	X	X	X					
U01					X	X	X	X					
U02					X	X	X	X					
U03					X	X	X	X					
K01						X	X	X					
K02						X	X	X					

Kryteria oceny	•
----------------	---

Uwagi	
-------	--

Treści merytoryczne (wykaz tematów)

1. Wprowadzenie do przetwarzania języka naturalnego – zadania, wyzwania i zastosowania w analizie danych.
2. Zaawansowane metody wstępnego przetwarzania danych tekstowych.
3. Metody reprezentacji tekstu: Bag-of-Words, TF-IDF, wektory osadzeń słów (Word2Vec, GloVe, FastText).
4. Modele statystyczne i metody uczenia maszynowego w analizie tekstu.
5. Modele sekwencyjne w NLP (RNN, LSTM, GRU).
6. Architektura transformerów i współczesne modele językowe.
7. Zadania NLP: klasyfikacja tekstu, analiza sentymentu, rozpoznawanie nazwanych jednostek.
8. Modelowanie tematów i eksploracja danych tekstowych.

9. Fine-tuning i wykorzystanie wstępnie wytrenowanych modeli językowych.
10. Implementacja i ocena modeli NLP w środowiskach programistycznych wykorzystywanych w Data Science.

Wykaz literatury podstawowej

1.

Wykaz literatury uzupełniającej

1.

Bilans godzinowy zgodny z CNPS (Całkowity Nakład Pracy Studenta) **studia stacjonarne**

liczba godzin w kontakcie z prowadzącymi	Wykład	15
	Konwersatorium (ćwiczenia, laboratorium itd.)	30
	Pozostałe godziny kontaktu studenta z prowadzącym	5
liczba godzin pracy studenta bez kontaktu z prowadzącymi	Lektura w ramach przygotowania do zajęć	5
	Przygotowanie krótkiej pracy pisemnej lub referatu po zapoznaniu się z niezbędną literaturą przedmiotu	5
	Przygotowanie projektu lub prezentacji na podany temat (praca w grupie)	5
	Przygotowanie do egzaminu/zaliczenia	10
Ogółem bilans czasu pracy		75
Liczba punktów ECTS w zależności od przyjętego przelicznika		3

Bilans godzinowy zgodny z CNPS (Całkowity Nakład Pracy Studenta) **studia niestacjonarne**

liczba godzin w kontakcie z prowadzącymi	Wykład	10
	Konwersatorium (ćwiczenia, laboratorium itd.)	20
	Pozostałe godziny kontaktu studenta z prowadzącym	10
liczba godzin pracy studenta bez kontaktu z prowadzącymi	Lektura w ramach przygotowania do zajęć	5
	Przygotowanie krótkiej pracy pisemnej lub referatu po zapoznaniu się z niezbędną literaturą przedmiotu	5
	Przygotowanie projektu lub prezentacji na podany temat (praca w grupie)	10
	Przygotowanie do egzaminu/zaliczenia	15
Ogółem bilans czasu pracy		75
Liczba punktów ECTS w zależności od przyjętego przelicznika		3