

Nazwa przedmiotu	Text mining i eksploracja mediów społecznościowych
Kod przedmiotu	<i>do wypełnienia przez dziekanat</i>
Profil studiów	ogólnoakademicki
Stopień studiów	drugi
Kierunek studiów	Ekonomia
Specjalność (jeśli dotyczy)	Analityka Biznesowa
Forma studiów	studia stacjonarne
Rok studiów	II
Semestr	zimowy
Liczba godzin	30
Forma/typ zajęć	konwersatorium
Punkty ECTS	3
Rodzaj przedmiotu (przedmiot obowiązkowy / kierunkowy do wyboru / fakultatywny / seminarium)	obowiązkowy
Sposób realizacji przedmiotu (stacjonarny / zdalny / kurs internetowy / kurs realizowany metodą blended learning)	stacjonarny
Język wykładowy	polski
Wymagania formalne	brak
Założenia wstępne	Zakłada się, że student posiada podstawową wiedzę z zakresu matematyki i statystyki, rozumie elementarne pojęcia algebry oraz potrafi logicznie analizować problemy. Oczekiwana jest również ogólna orientacja w zagadnieniach informatycznych oraz umiejętność pracy z danymi.
Skrócony opis przedmiotu	Celem zajęć jest dokonanie zaawansowanego studium metod przetwarzania języka naturalnego oraz wykrywania wzorców w nieustrukturyzowanych danych tekstowych. Studenci zapoznają się z metodologiami uczenia maszynowego nienadzorowanego i nadzorowanego, włączając analizę częstościową, analizę sentymentu oraz modelowanie tematów. W trakcie zajęć studenci będą pracować na rzeczywistych zbiorach danych, ze szczególnym uwzględnieniem eksploracji treści pochodzących z mediów społecznościowych, przy wykorzystaniu narzędzi i algorytmów stosowanych w nowoczesnej analityce tekstowej. Studenci nabędą umiejętności samodzielnego projektowania i realizacji badań opartych na danych tekstowych, interpretacji uzyskanych wyników oraz doboru odpowiednich technik w praktycznych zastosowaniach akademickich i biznesowych.
Treści kształcenia dla przedmiotu	<p>Przedmiotem zajęć jest zapoznanie studentów z nowoczesnymi metodami przetwarzania języka naturalnego oraz analizy nieustrukturyzowanych danych tekstowych. Omawiany materiał obejmuje zarówno klasyczne podejścia do eksploracji tekstu, jak i zaawansowane metody uczenia maszynowego wykorzystywane w analityce tekstowej.</p> <p>Celem laboratorium jest rozwinięcie kompetencji w zakresie projektowania i realizacji badań opartych na danych tekstowych, ze szczególnym uwzględnieniem treści pochodzących z mediów społecznościowych. Studenci</p>

	<p>pracują na rzeczywistych zbiorach danych, ucząc się doboru odpowiednich metod, narzędzi i algorytmów analitycznych.</p> <p>W kolejnych blokach omawiane są następujące treści:</p> <ul style="list-style-type: none"> • Wprowadzenie do analityki tekstowej i NLP. Charakterystyka danych tekstowych nieustrukturyzowanych. Podstawowe pojęcia przetwarzania języka naturalnego. Specyfika danych z mediów społecznościowych. • Wstępne przetwarzanie i reprezentacja tekstu. Czyszczenie danych tekstowych i normalizacja. Tokenizacja, lematyzacja i usuwanie słów funkcyjnych. Reprezentacja tekstu: bag-of-words, n-gramy, TF i TF-IDF. • Analiza częstościowa i eksploracyjna tekstu. Analiza częstości występowania słów i fraz. Wizualizacja danych tekstowych (chmury słów, wykresy częstości). Eksploracyjna analiza danych tekstowych jako etap wstępny badań. • Analiza sentymentu i emocji. Podejścia słownikowe do analizy sentymentu. Wprowadzenie do nadzorowanych metod klasyfikacji sentymentu. Ograniczenia analizy sentymentu w danych społecznościowych. • Modelowanie tematów i metody nienadzorowane. Klasteryzacja dokumentów tekstowych. Modelowanie tematów (LDA, NMF). Metody nienadzorowane w badaniach akademickich i analizach biznesowych. • Projektowanie badań i zastosowania praktyczne NLP. Dobór metod analizy do problemu badawczego. Łączenie różnych technik analityki tekstowej w jednym projekcie. Aspekty etyczne wykorzystania danych tekstowych z mediów społecznościowych
Przedmiotowe efekty uczenia się	<p>Po ukończeniu przedmiotu, student:</p> <p>W ZAKRESIE WIEDZY:</p> <ul style="list-style-type: none"> • zna i rozumie podstawowe oraz zaawansowane metody przetwarzania języka naturalnego i analizy nieustrukturyzowanych danych tekstowych; • zna i rozumie różnice między podejściami uczenia maszynowego nadzorowanego i nienadzorowanego w analityce tekstowej, włączając analizę częstościową, analizę sentymentu oraz modelowanie tematów; • zna i rozumie możliwości oraz ograniczenia stosowania metod text mining w badaniach akademickich i zastosowaniach biznesowych, w szczególności w analizie mediów społecznościowych; <p>W ZAKRESIE UMIEJĘTNOŚCI:</p> <ul style="list-style-type: none"> • potrafi przygotować dane tekstowe do analizy, stosując odpowiednie techniki wstępnego przetwarzania i reprezentacji tekstu; • potrafi samodzielnie zastosować wybrane metody analizy tekstu oraz interpretować uzyskane wyniki; • potrafi zaprojektować i zrealizować badanie oparte na danych tekstowych, dobierając adekwatne narzędzia i algorytmy do postawionego problemu badawczego; <p>W ZAKRESIE KOMPETENCJI:</p> <ul style="list-style-type: none"> • jest gotów do krytycznej oceny wyników analiz tekstowych oraz świadomego korzystania z narzędzi automatycznej analizy języka;

	<ul style="list-style-type: none"> wykazuje świadomość etyczną i metodologiczną w analizie danych tekstowych, w tym danych pochodzących z mediów społecznościowych.
Powiązanie efektów przedmiotowych z efektami kierunkowymi/specjalnościowymi (oznaczonymi kodami z programu nauczania)	specjalność Analityka Biznesowa: S_W04 S_U02 S_U03 S_K01 S_K03
Nakład pracy studenta	<p>Szacunkowy nakład pracy studenta: 3ECTS x 25h = 75h</p> <p>(K) - godziny kontaktowe (S) - godziny pracy samodzielnej</p> <p>konwersatorium (zajęcia): 30h (K) 0h (S)</p> <p>konsultacje: 5h (K) 0h (S)</p> <p>przygotowanie do ćwiczeń: 0h (K) 10h (S)</p> <p>przygotowanie projektu zaliczeniowego: 0h (K) 20h (S)</p> <p>praca z materiałami dodatkowymi: 0h (K) 10h (S)</p> <p>Razem: 35h (K) + 40h (S) = 75h</p>
Metody i kryteria oceniania	<p>Uzyskanie zaliczenia przedmiotu wymaga:</p> <ol style="list-style-type: none"> obecności na ćwiczeniach (dopuszczalne są dwie nieobecności) i ich zaliczenia; podstawą zaliczenia ćwiczeń jest aktywny udział w pracy zespołowej podczas realizacji projektu zaliczeniowego oraz aktywność w czasie zajęć, polegająca w szczególności na udziale w dyskusji czy wykonywaniu zadań przy użyciu narzędzi informatycznych; wykonania projektu zaliczeniowego w grupach, obejmującego przygotowanie, analizę i interpretację danych tekstowych z wykorzystaniem poznanych metod analityki tekstowej; uzyskania co najmniej 51% punktów możliwych do zdobycia w ramach projektu zaliczeniowego, na którego pozytywną ocenę składają się: poprawność metodologiczna zastosowanych rozwiązań, adekwatność doboru narzędzi i algorytmów, jakość interpretacji wyników oraz prezentacja efektów pracy zespołu; Skala ocen: [0%-51%) – ndst [51%-60%) – dst [60%-70%) – dst + [70%-80%) – db [80%-90%) – db+ [90%-100%) – bdb.
Literatura	<p>Literatura obowiązkowa (wybrane rozdziały):</p> <ul style="list-style-type: none"> Sowmya Vajjala, Bodhisattwa Majumder, Anuj Gupta, Harshit Surana, Przetwarzanie języka naturalnego w praktyce. Przewodnik po budowie rzeczywistych systemów NLP, Helion 2023 Piotr Wróblewski, Machine learning i natural language processing w programowaniu. Podręcznik z ćwiczeniami w Pythonie, Helion 2024 Alex J. Gutman, Jordan Goldmeier, Analityk danych. Przewodnik po data science, statystyce i uczeniu maszynowym, Helion 2023 Jon Krohn, Grant Beyleveld, Aglaé Bassens, Uczenie głębokie i sztuczna inteligencja. Interaktywny przewodnik ilustrowany, Helion 2021 <p>Literatura uzupełniająca (wybrane rozdziały):</p>

	<ul style="list-style-type: none">• Julia Silge, David Robinson, Text Mining with R, O'Reilly Media 2025, [https://www.tidytextmining.com/]• Rafael A. Irizarry, Introduction to Data Science, Data Wrangling and Visualization with R, CRC Press 2024, [https://rafalab.dfci.harvard.edu/dsbook-part-1/]• Daniel Jurafsky, James H. Martin, Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models, 3rd edition 2025, [https://web.stanford.edu/~jurafsky/slp3/]• Steven Bird, Ewan Klein, Edward Loper, Natural Language Processing with Python, O'Reilly Media 2019, [https://www.nltk.org/book/]
Metody dydaktyczne	<p>W ramach zajęć wykorzystywane są następujące metody dydaktyczne:</p> <ul style="list-style-type: none">• mini-wykład interaktywny• ćwiczenia rachunkowe i analityczne• analiza danych w programach komputerowych• webquest
Oprogramowanie	MS Office, zintegrowane środowisko programistyczne (RStudio / Positron / Jupyter Notebook / PyCharm / Google Colab)