Uniwersytet Gdański

Karta przedmiotu

Nazwa i kod przedmiotu	Text and social network mining, PG_00177507		
Kierunek studiów	Informatyka i ekonometria (O)		
Data rozpoczęcia studiów	październik 2026 r.	Rok akademicki realizacji przedmiotu	2027/2028
Poziom kształcenia	II stopnia	Grupa zajęć	Grupa zajęć obowiązkowych z zakresu kierunku studiów Grupa zajęć fakultatywnych Grupa zajęć powiązanych z prowadzonymi badaniami naukowymi w dziedzinie nauki związanej z kierunkiem - profil ogólnoakademicki
Forma studiów	stacjonarne	Sposób realizacji	na uczelni
Rok studiów	2	Język wykładowy	polski
Semestr studiów	3	Liczba punktów ECTS	5.0
Profil kształcenia	ogólnoakademicki	Forma zaliczenia	egzamin
Jednostka prowadząca	Rektor -> Wydział Zarządzania -> Katedra Statystyki		
Imię i nazwisko wykładowcy (wykładowców)	Odpowiedzialny za przedmiot	dr hab. Krzysztof Najman	
	Prowadzący zajęcia z przedmiotu		

Formy zajęć

Forma zajęć	Wykład	Ćwiczenia	Laboratorium	Projekt	Seminarium	RAZEM
Liczba godzin zajęć	30.0	0.0	30.0	0.0	0.0	60

W tym liczba godzin zajęć na odległość: 0.0

Aktywność studenta i liczba godzin pracy

Aktywność studenta	Udział w zajęciach dydaktycznych, objętych planem studiów	Udział w konsultacjach	Praca własna studenta	RAZEM
Liczba godzin pracy studenta	60	4.0	61.0	125

Cel przedmiotu

Zapoznanie studentów z teoretycznymi i praktycznymi aspektami eksploracji danych tekstowych.

Zapoznanie studentów z procesem analizy text mining, przeszukiwaniem dokumentów niestrukturalnych i praktycznym zastosowaniem text mining.

Zapoznanie studentów z teoretycznymi i praktycznymi aspektami budowy i analizy sieci społecznościowych jako struktur opisujących relacje między jednostkami (osobami, organizacjami, obiektami).

Efekty uczenia się przedmiotu	Efekt kierunkowy	Efekt z przedmiotu	Sposób weryfikacji i oceny efektu
	[IiEMU2_U06] Potrafi wykorzystywać i integrować uporządkowaną i szczegółową wiedzę z zakresu nauk o zarządzaniu i jakości oraz ekonomii i finansów na potrzeby rozstrzygania dylematów i opracowywania innowacyjnych rozwiązań złożonych lub nietypowych problemów, pojawiających się w pracy zawodowej	Student potrafi budować i stosować w praktyce wybrane modele analizy Text Mining i sieci społecznościowych.	[SU2] prezentacja/projekt/referat/ raport
	[IiEMU2_U03] Potrafi pozyskiwać i weryfikować dane z właściwie dobranych źródeł, gromadzić je, przetwarzać i wizualizować za pomocą nowoczesnych narzędzi ekonometrycznych, informatycznych lub statystycznych	Student potrafi zebrać, zapisać, zweryfikować i ocenić dane niezbędne dla analizy Text Mining i sieci społecznościowych.	[SU2] prezentacja/projekt/referat/ raport
	[IiEMU2_W06] Ma uporządkowaną i podbudowaną teoretycznie wiedzę z zakresu procesów, metod i narzędzi projektowania, tworzenia, rozwoju i zapewnienia odpowiednich warunków użytkowania narzędzi informatycznych, ekonometrycznych lub statystycznych	Student zna zasady budowy, weryfikacji i oceny wybranych metod analizy Text Mining i sieci społecznościowych.	[SW4] test/egzamin - ustny lub pisemny [SW2] prezentacja/projekt/referat/ raport
	[IiEMU2_W08] W pogłębionym stopniu zna i rozumie metody, uwarunkowania, kierunki rozwoju oraz dylematy związane z zastosowaniem zaawansowanych narzędzi ekonometrycznych, informatycznych lub statystycznych, w kontekście dynamicznych zmian otoczenia	Student zna teoretyczne podstawy analizy Text Mining i analizy sieci społecznościowych.	[SW4] test/egzamin - ustny lub pisemny [SW2] prezentacja/projekt/referat/ raport

Treści przedmiotu	**Definicja, istota, cel, zakres i korzysci ze stosowania text miningu** (eksploracja danych tekstowych). **Obszary i przykłady zastosowan tekst miningu**. **Proces analizy text mining**. Okreslenie celu, zakresu i kosztow analizy, okreslenie zrodeł danych nieustrukturyzowanych, przekształcanie dokumentow zrodłowych do postaci pliku tekstowego. **Proces leksykalny** - oznaczanie czesci mowy. Ekstrakcja rdzeni wyrazen (**stemming**) automatyczne odnajdowanie rdzeni lub pni wyrazow, przypisywanie słowom ich bazowej formy. **Parsing** wyłaczanie wyrazow o niskiej wartosci informacyjnej. **Tagowanie** (tagging) wybor opisu morfoskładniowego. **Lematyzacja** - identyfikacja leksemu. Budowa **stop listy**, usuwanie słow (tagow) nieistotnych. Przycinanie (**pruning**) usuwanie niepotrzebnych słow: podejscie oparte na usuwaniu wyrazen wystepujacych najczesciej (most frequent) i najrzadziej (least frequent). Budowa **macierzy czestosci** wystepowania wyrazow dokumencie, filtrowanie tekstu i ekstrakcja faktow (macierz dokumentow-wyrazen, term-document matrix). **Reprezentacja dokumentow tekstowych**, reprezentacja wektorowa (vector space model VSM) oraz reprezentacja grafowa: standardowa, prosta, n-odległosci, prostej n-odległosci, bezwzglednej czestosci, wzglednej czestosci.Wykorzystanie modeli językowych typu BERT do reprezentacji danych tekstowych. **Przeszukiwanie dokumentow niestrukturalnych**: wyszukiwanie pełnotekstowe FTS (Full Text Searching). Macierz term-by-document frequency. Wazenie słow. Metody okreslania liczby wystapien danego wyrazenia w zbiorze dokumentow (kolekcji). **Modelowanie tematyczne** (topic modeling) dokumentów tekstowych. Wprowadzenie do modeli probabilistycznych: Latent Semantic Analysis (LSA), Probabilistic Latent Semantic Analysis (PLSA), oraz Latent Dirichlet Allocation (LDA). Interpretacja tematów jako rozkładów słów oraz dokumentów jako rozkładów tematów. Wyodrębnianie tematów metodami opartymi o modele językowe typu BERT, takimi jak BERTopic. **Określanie emocjonalnego charakteru wypowiedzi** (sentiment analysis). Podejście oparte na analizie słownikowej (lexicon-based) i na analizie statystycznej (statistical-based): uczenie modeli klasyfikujących (np. Naiwnego Klasyfikatora Bayesa) na podstawie oznakowanych zbiorów danych. Ocena wydźwięku na różnego rodzaju reprezentacjach tekstowych. **Definicja, istota i charakterystyka sieci społecznościowych** jako struktur opisujących relacje między jednostkami (osobami, organizacjami, obiektami). Elementy składowe sieci: węzły (aktorzy) i krawędzie (relacje). Przykłady zastosowań sieci społecznych w kontekście społecznym, informacyjnym i biznesowym. Znaczenie sieci w analizie struktur społecznych, przepływu informacji, dynamiki grupowej i wpływu społecznego. **Podstawowe pojęcia z teorii grafów**: graf skierowany i nieskierowany, graf ważony, spójność grafu, ścieżka, cykl. Konwersja danych relacyjnych do formatu sieciowego. Reprezentacja relacji w grafie za pośrednictwem: prostej listy relacji (linked list), ważonej listy relacji (edged list), macierz sąsiedztwa (adjacency matrix), macierz incydencji (incidence matrix). Zasady konstruowania grafów z danych tekstowych. **Techniki prezentacji struktury sieci** przy użyciu narzędzi graficznych. Estetyka i czytelność wizualizacji: zastosowanie różnych układów rozmieszczenia węzłów (layouts), kolorowanie, skalowanie węzłów i krawędzi. Interpretacja wizualizacji jako narzędzia wspomagającego analizę wyników i wyodrębnianie wniosków. **Identyfikacja i interpretacja pozycji aktorów** (węzłów) w strukturze sieci społecznej. Miary centralności: centralność stopnia (degree), pośrednictwa (betweenness), bliskości (closeness) oraz wektorów własnych (eigenvector centrality). Zastosowanie wskaźników centralności do wykrywania wpływowych użytkowników oraz oceny ich znaczenia w kontekście przepływu informacji, opinii i zasobów w sieci. Analiza roli aktorów w sieci. **Badanie właściwości topologicznych sieci**: połączalności (connectivity), gęstości (density), średniej długości ścieżki, średniego stopnia wierzchołków. Wykorzystanie wskaźników do oceny struktury sieci, analizy obszarów nietypowych oraz węzłów o krytycznym znaczeniu dla jej integralności. Ocena odporności sieci na zakłócenia oraz ocena jej efektywności w kontekście rozprzestrzeniania się informacji. **Analiza grup węzłów**. Wyodrębnienie kluczowych struktur, takich jak największe komponenty, kliki i n-kliki. Przegląd metod wykrywania wspólnot, w tym algorytmów Girvana-Newmana, Louvain oraz metody modularności. Zastosowanie technik detekcji wspólnot do segmentacji użytkowników oraz identyfikacji spójnych grup w sieciach społecznych.

Wykorzystanie modeli Markowa do predykcji zasięgu wiadomości, identyfikacji potencjalnych ścieżek rozprzestrzeniania się wiadomości w sieci, wykrywania anomalii oraz predykcji dalszych zachowań użytkowników w sieci. Symulacja rozchodzenia się informacji z wykorzystaniem losowego spaceru po grafie.

Wprowadzenie do sieci Bayesa jako narzędzia probabilistycznego w analizie relacji między aktorami. Reprezentacja zależności warunkowych w sieci społecznej przy użyciu grafów skierowanych acyklicznych (directed acyclic graph, DAG). Modelowanie wpływu, przewidywania zachowań użytkowników, identyfikacji ukrytych zmiennych oraz symulacji rozprzestrzeniania się informacji w warunkach niepewności. Integracja podejścia bayesowskiego z klasyczną analizą grafową.

Wymagania wstępne i dodatkowe

Podstawy algebry liniowej, statystyki i statystycznej analizy wielowymiarowej.

Sposoby i kryteria oceniania osiąganych efektów uczenia się

Sposób oceniania (składowe)	Próg zaliczeniowy	Składowa oceny końcowej
egzamin pisemny	51.0%	40.0%
projekt semestralny	51.0%	60.0%

Zalecana lista lektur

Podstawowa lista lektur

Feldman R., Sanger J., 2006, Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data, Cambridge University Press New York

Berry M.W., Kogan J., 2010, Text Mining, Applications and Theory, A John Wiley and Sons, Ltd.

Solka J. L., 2008, Text Data Mining: Theory and Methods, Statistic Survey

Majkowska, A., Migdał-Najman, K., Najman, K., i Raca, K. 2021. Identification of the words most frequently used by different generations of twitter users. Data analysis and classification: methods and applications, Springer, 2747.

R. A. Hanneman, M. Riddle, Introduction to social network methods. Riverside, CA: University of California, Riverside, 2005.

P.J. Carrington, J. Scott, S. Wasserman (eds.), Models and Methods in Social Network Analysis. Cambridge University Press, 2005.

M. Kamola, P. Arabas, Sieci społeczne i technologiczne. Jak zrozumieć, jak wykorzystać, wyd. PWN, 2018

M. A. Russell, M. Klassen, Data mining. Eksploracja danych w sieciach społecznościowych, wyd. Helion, 2019

Uzupełniająca lista lektur

P.Lula, Statystyczne modelowanie zawartości dokumentów tekstowych, WUEK, 2018.

Adresy eZasobów

Przykładowe zagadnienia/ przykładowe pytania/ realizowane zadania

Praktyki zawodowe w ramach przedmiotu

Nie dotyczy

Dokument wygenerowany elektronicznie. Nie wymaga pieczęci ani podpisu.