

Data mining - opis przedmiotu

Informacje ogólne	
Nazwa przedmiotu	Data mining
Kod przedmiotu	04.2-WE-BizEIP-DataMining-Er
Wydział	Wydział Informatyki, Elektrotechniki i Automatyki
Kierunek	Biznes elektroniczny
Profil	praktyczny
Rodzaj studiów	Program Erasmus pierwszego stopnia
Semestr rozpoczęcia	semestr zimowy 2020/2021

Informacje o przedmiocie	
Semestr	2
Liczba punktów ECTS do zdobycia	3
Typ przedmiotu	obowiązkowy
Język nauczania	angielski
Sylabus opracował	<ul style="list-style-type: none">dr hab. inż. Marek Kowal, prof. UZ

Formy zajęć					
Forma zajęć	Liczba godzin w semestrze (stacjonarne)	Liczba godzin w tygodniu (stacjonarne)	Liczba godzin w semestrze (niestacjonarne)	Liczba godzin w tygodniu (niestacjonarne)	Forma zaliczenia
Wykład	15	1	-	-	Zaliczenie na ocenę
Laboratorium	30	2	-	-	Zaliczenie na ocenę

Cel przedmiotu

Presentation of the software used for data mining. Familiarize students with the methods of data cleaning. Presentation of data classification methods. Presentation of methods of association and sequences rules discovery. Presentation of data clustering methods. Developing practical skills in operating selected data mining systems. Developing skills in the application of data mining methods in e-business (customer segmentation, credit risk scoring, cross-selling strategies, fraud detection).

Wymagania wstępne

Zakres tematyczny

Review and characteristics of the software used for data mining. Introduction to data mining software (SAS). Data structures used in data mining. Types and roles of variables in data mining tasks.

Preparation of data for exploration. Data profiling. Data cleansing. Data sampling. Transformation of variables. Variable Selection.

Data classification. Classification trees, k-nearest neighbors, naive Bayes classifier, neural networks, logistic regression. Measures of classification accuracy. Practical exercises from data classification.

Discovering association and sequence rules. Measures describing the statistical importance and strength of association and sequence rules. Market basket analysis. Computational complexity of association rules discovery. Discussion of the Apriori and Generalized Sequential Pattern algorithm. Practical exercises from association and sequence rules discovery.

Data clustering. Methods of hierarchical clustering. Clustering methods based on iterative optimization. Distance measures used in clustering algorithms. Clusters summary. Methods for estimating the number of clusters. Practical exercises from data clustering.

Metody kształcenia

Lecture - conventional lecture using a video projector.

Laboratory - practical exercises in the computer laboratory.

Efekty uczenia się i metody weryfikacji osiągnięcia efektów uczenia się

Opis efektu	Symbole efektów	Metody weryfikacji	Forma zajęć
Knows IT technologies and tools used in data mining		<ul style="list-style-type: none">test z pytaniami zamkniętymi i otwartymi	<ul style="list-style-type: none">Wykład

Opis efektu	Symboleefektów Metody weryfikacji	Forma zajęć
Student can prepare data for exploration by applying cleaning procedures	<ul style="list-style-type: none"> • obserwacja i ocena aktywności na zajęciach • obserwacje i ocena umiejętności praktycznych studenta • sprawdzian z progami punktowymi 	<ul style="list-style-type: none"> • Laboratorium
Knows the typical applications of data mining methods in e-business	<ul style="list-style-type: none"> • test z pytaniami zamkniętymi i otwartymi 	<ul style="list-style-type: none"> • Wykład
Is able to use known algorithms to discover association and sequence rules	<ul style="list-style-type: none"> • obserwacja i ocena aktywności na zajęciach • obserwacje i ocena umiejętności praktycznych studenta • sprawdzian z progami punktowymi 	<ul style="list-style-type: none"> • Laboratorium
Student knows the methods of data mining	<ul style="list-style-type: none"> • test z pytaniami zamkniętymi i otwartymi 	<ul style="list-style-type: none"> • Wykład
Student can choose the method of data clustering, carry out the clustering process and determine the summaries of clusters	<ul style="list-style-type: none"> • obserwacja i ocena aktywności na zajęciach • obserwacje i ocena umiejętności praktycznych studenta • sprawdzian z progami punktowymi 	<ul style="list-style-type: none"> • Laboratorium
Student can choose the method of data classification, train a classifier and measure the classification accuracy	<ul style="list-style-type: none"> • obserwacja i ocena aktywności na zajęciach • obserwacje i ocena umiejętności praktycznych studenta • sprawdzian z progami punktowymi 	<ul style="list-style-type: none"> • Laboratorium
The student is able to apply data mining methods in innovative solutions for e-business	<ul style="list-style-type: none"> • aktywność w trakcie zajęć • wykonanie sprawozdań laboratoryjnych 	<ul style="list-style-type: none"> • Laboratorium
The student understands the need to constantly supplement and expand his/her knowledge in the field of data mining methods	<ul style="list-style-type: none"> • aktywność w trakcie zajęć • wykonanie sprawozdań laboratoryjnych 	<ul style="list-style-type: none"> • Laboratorium

Warunki zaliczenia

Lecture - the passing criteria is to obtain positive grades from tests carried out at least once in a semester.

Laboratory - the passing criterion is to obtain positive marks for laboratory exercises and tests.

Final mark components = lecture: 50% + teaching laboratory: 50%

Literatura podstawowa

1. Aggarwal C.C.: Data Mining, Springer, 2015.
2. Han J., Kamber, M., Data Mining: Concepts and Techniques, Morgan Kaufmann, 2011.
3. Hastie T., Tibshirani R., Witten D., James G: An Introduction to Statistical Learning, Springer, 2013

Literatura uzupełniająca

1. Hastie T., Tibshirani R., Friedman J.H.: The Elements of Statistical Learning, Springer 2001

Uwagi

Zmodyfikowane przez dr hab. inż. Marcin Mrugalski, prof. UZ (ostatnia modyfikacja: 24-04-2020 14:51)

Wygenerowano automatycznie z systemu SyllabUZ