**Nazwa przedmiotu:**

Data Mining

**Koordynator przedmiotu:**

Marzena Kryszkiewicz

**Status przedmiotu:**

Fakultatywny ograniczonego wyboru

**Poziom kształcenia:**

Studia II stopnia

**Program:**

Informatyka

**Grupa przedmiotów:**

Przedmioty techniczne - zaawansowane

**Kod przedmiotu:**

EDAMI

**Semestr nominalny:**

3 / rok ak. 2018/2019

**Liczba punktów ECTS:**

6

**Liczba godzin pracy studenta związanych z osiągnięciem efektów uczenia się:**

30 hours of lectures
30 hours preparation for tests
15 hours of laboratory exercises
15 hours of preparation for the laboratory exercises
15 hours of project meetings
45 hours of implementation of project assignments

**Liczba punktów ECTS na zajęciach wymagających bezpośredniego udziału nauczycieli akademickich:**

30 hours of lecture
15 hours of laboratory exercises
15 hours of project meetings
which gives approx. 2.5 ECTS

**Język prowadzenia zajęć:**

angielski

**Liczba punktów ECTS, którą student uzyskuje w ramach zajęć o charakterze praktycznym:**

15 hours of laboratory exercises
15 hours of preparation for laboratory exercises
15 hours of project meetings
45 hours of implementation of project assignments
which gives approx. 4 ECTS

**Formy zajęć i ich wymiar w semestrze:**

|  |  |
| --- | --- |
| Wykład:  | 30h |
| Ćwiczenia:  | 0h |
| Laboratorium:  | 15h |
| Projekt:  | 15h |
| Lekcje komputerowe:  | 0h |

**Wymagania wstępne:**

knowledge of data bases is recommended

**Limit liczby studentów:**

30

**Cel przedmiotu:**

The objective of the course is to make students familiar with important topics in the area of data mining. The techniques and algorithms to be presented are of practical value – they are well suited to the discovery of hidden data in real large data sources. The methods to be presented are anticipated to have a great impact on the evolution of database systems towards effective knowledge base systems. As a result of participating in the course, students should become capable of efficiently discovering novel, non-trivial and useful knowledge from large data resources.

**Treści kształcenia:**

Data mining as a multidisciplinary area: Roots and development of data mining area. Current challenges in data mining. Classification of data mining tasks.
Data preprocessing: Data cleaning. Data integration and transformation. Data reduction. Discretization and concept hierarchy generation.
Data mining language: Specifying required properties of knowledge to be discovered by means of a sample data mining language.
Frequent patterns and association rules: Scalable methods of discovering frequent patterns and association rules in transactional and relational databases. Modifications of algorithms capable to deal with hierarchy and negation. Usage of imposed constraints for efficient reduction of a discovery process.
Concise models of frequent patterns: Generators, closed itemsets and k-disjunction-free sets as basic elements of lossless representations of frequent patterns. Discovery of concise representations of frequent patterns. Usage of the models for derivation of all frequent patterns.
Concise models of association rules: Generators, closed itemsets, and pseudo-closed sets as building blocks of lossless representations of association rules. Mechanisms of deriving association rules from their representations.
Functional and approximate dependencies: Scalable methods of discovering functional and approximate dependencies in large databases.
Other patterns and rules: Scalable methods of discovering sequential patterns, episode rules, quantitative rules, frequent patterns, decision tree classifiers, and rough set decision rules.
Clustering: Scalable methods of clustering objects. Usage of multidimensional indexing techniques to support the process of discovering clusters and outliers.
Reasoning under incompleteness: Legitimate approach to reasoning from data with missing values. Mining from partial knowledge.
Data mining applications: Sample applications of data mining in the financial, telecommunication, biomedical and DNA areas. Brief overview of selected data mining systems.

**Metody oceny:**

In order to pass the EDAMI course, students must achieve a pass grade from each of the three course components: the lecture part (assessed on the basis of two tests), the project part (assessed on the basis of an implemented software and carried out tests, a report and presentation of the project) and the laboratory part (recognized as successfully completed if all 5 laboratory tasks are done correctly). A positive final grade is determined on the basis of the average of the grade from the tests and the grade from the project. If the grade from the tests is lower than the grade from the project, the final assessment is determined as the rounding down of that average. Otherwise, it is determined as the rounding up of that average.

**Egzamin:**

nie

**Literatura:**

Han J., Kamber M., Pei, J., Data Mining: Concepts and Techniques, The Morgan Kaufmann Series in Data Management Systems, 3rd edition, Morgan Kaufmann, 2011
Fayyad U.M. , Piatetsky-Shapiro G. , Smyth P., Uthurusamy R. (eds.), Advances in Knowledge Discovery and Data Mining, AAAI, Menlo Park, California, 1996
Kryszkiewicz M., Concise Representations of Frequent Patterns and Association Rules, Prace Naukowe, Elektronika, Oficyna Wydawnicza Politechniki Warszawskiej, z. 142 (2002)
Communications of the ACM, November 1996, Vol. 39. No 11., 1996
Ganter B., Wille R., Formal Concept Analysis, Mathematical Foundations, Springer-Verlag, 1999
and a number of recent data mining publications accessible via Internet. The instructor will recommend the respective publications during the course.

**Witryna www przedmiotu:**

https://usosweb.usos.pw.edu.pl/kontroler.php?\_action=katalog2/przedmioty/pokazPrzedmiot&prz\_kod=103B-CSCSN-MSA-EDAMI

**Uwagi:**

A project task is to design, implement and perform an experimental evaluation of selected data mining algorithms.

The aim of the laboratory is to acquaint students with modern technologies of data mining. During the laboratory classes, students will become familiar with possibilities of carrying out data mining using a selected commercial system, for example, IBM Warehouse Design Studio.

## Charakterystyki przedmiotowe

### Profil ogólnoakademicki - wiedza

**Charakterystyka EDAMI\_W01:**

has knowledge of discovering patterns and dependencies by means of data mining methods

Weryfikacja:

test

**Powiązane charakterystyki kierunkowe:** K\_W06, K\_W08, K\_W09

**Powiązane charakterystyki obszarowe:** I.P7S\_WG, III.P7S\_WG.o

**Charakterystyka EDAMI\_W02:**

has knowledge of methods of representing frequent patterns and reasoning about them

Weryfikacja:

test

**Powiązane charakterystyki kierunkowe:** K\_W06

**Powiązane charakterystyki obszarowe:** I.P7S\_WG, III.P7S\_WG.o

**Charakterystyka EDAMI\_W03:**

 has knowledge of modern data mining technologies

Weryfikacja:

laboratory excercises

**Powiązane charakterystyki kierunkowe:** K\_W11

**Powiązane charakterystyki obszarowe:** I.P7S\_WG

### Profil ogólnoakademicki - umiejętności

**Charakterystyka EDAMI\_U01:**

is capable of planning and implementing a knowledge discovery process as well as of interpreting its results

Weryfikacja:

project

**Powiązane charakterystyki kierunkowe:** K\_U01, K\_U06, K\_U09, K\_U13

**Powiązane charakterystyki obszarowe:** I.P7S\_UK, I.P7S\_UW, III.P7S\_UW.1.o, III.P7S\_UW.3.o

**Charakterystyka EDAMI\_U02:**

is capable of presenting a plan, implementation and results of a knowledge discovery process in an oral and written form

Weryfikacja:

project

**Powiązane charakterystyki kierunkowe:** K\_U03, K\_U06

**Powiązane charakterystyki obszarowe:** I.P7S\_UK, I.P7S\_UW

**Charakterystyka EDAMI\_U03:**

is capable of discovering knowledge using modern data mining technologies

Weryfikacja:

laboratory excercises

**Powiązane charakterystyki kierunkowe:** K\_U13

**Powiązane charakterystyki obszarowe:** I.P7S\_UW, III.P7S\_UW.3.o