1. Presenting DM as a process that should involve a methodology id applied at its best.
2. Introducing the students to the new concept of DM for processes, called Process Mining.
3. Delving into some detail in one of the stages of DM: data exploration
4. Dealing in detail with the problem of data visualization for exploration as a key issue in DM.
5. Introducing the students to the basics of probability theory as applied in Intelligent Data Analysis (IDA)
6. Introducing the students to the probabilistic variant of IDA in the form of Statistical Machine Learning, both for supervised and unsupervised learning models.
7. Dealing in detail with different unsupervised models for data visualization, including case studies.
8. Approaching the multi-faceted concept of data mining (DM) from different perspectives.

Capacidades previas

Students are expected to have at least some basic background in the area of artificial intelligence and, more specifically, with the areas of Machine Learning and Computational Intelligence. Some basic knowledge of probability theory and statistics would be beneficial. Other than this, the course is open to students and researchers of all types of background.

Metodologías docentes

This course will build on different teaching methodology (TM) aspects, including:
TM1: Expositive seminars
TM2: Expositive-participative seminars
TM3: Orientation for individual assignments (essays)
TM4: Individual tutorization

Objetivos de aprendizaje de la asignatura

1. Presenting DM as a process that should involve a methodology id applied at its best.
2. Introducing the students to the new concept of DM for processes, called Process Mining.
3. Delving into some detail in one of the stages of DM: data exploration
4. Dealing in detail with the problem of data visualization for exploration as a key issue in DM.
5. Introducing the students to the basics of probability theory as applied in Intelligent Data Analysis (IDA)
6. Introducing the students to the probabilistic variant of IDA in the form of Statistical Machine Learning, both for supervised and unsupervised learning models.
7. Dealing in detail with different unsupervised models for data visualization, including case studies.
8. Approaching the multi-faceted concept of data mining (DM) from different perspectives.
## Introduction to the concept of data mining (DM).

**Competencias de la titulación a las que contribuye el contenido:**

**Descripción:**
DM is a multi-faceted concept that requires discussion and clarification. We will do this at the beginning of the course.

---

## DM as a methodology.

**Competencias de la titulación a las que contribuye el contenido:**

**Descripción:**
We argue that DM should not be focused on the concept of data analysis/modeling, but, instead, should be treated as a methodology with diverse inter-related stages.

---

## DM for processes: Process Mining.

**Competencias de la titulación a las que contribuye el contenido:**

**Descripción:**
A new development in DM methodologies is that which deals with one specifically suited for processes. It is called Process Mining and will be described and discussed in this course.

---

## Data exploration in DM.

**Competencias de la titulación a las que contribuye el contenido:**

**Descripción:**
One of the main stages of well-structures DM methodologies is Data exploration. It will be discussed as a preamble to data visualization.

---

## Basics of probability theory in Intelligent Data Analysis (IDA)

**Competencias de la titulación a las que contribuye el contenido:**

**Descripción:**
For a long time in the last half-century, multivariate statistics and artificial intelligence (mostly in the field of machine learning) have developed in parallel without fully meeting. Statistical machine learning has bridged that field over the last two decades. We introduce it by first providing some basic principles of probability theory (Bayesian inference).

---

## Data visualization for exploration.

**Competencias de la titulación a las que contribuye el contenido:**
One of the aspects of the problem of data exploration is data visualization. It has a research 'life' of its own as it involves not only computer-based mathematical models, but also natural perception and processing.

**Statistical Machine Learning for IDA: supervised models.**

**Descripción:**
Once the basics of Bayesian inference are set, we will delve into the field of Statistical Machine Learning for IDA, starting with supervised learning models, with an emphasis on feed-forward artificial neural networks.

**Statistical Machine Learning for IDA: unsupervised models.**

**Descripción:**
Once the basics of Bayesian inference and of Statistical Machine Learning for IDA in supervised models are set, we will continue with unsupervised models, focusing on self-organizing maps and related models.

**Unsupervised models for data visualization, with case studies.**

**Descripción:**
In the final item of the contents of the course, we will bring statistical machine learning and data visualization together by discussing some probabilistic unsupervised learning models for data visualization, including some case studies as an example.
## Introduction to Data Mining and its Methodologies

**Descripción:**
Introduction to Data Mining as a general concept and to its methodologies for practical implementation

**Objetivos específicos:**
1, 9

**Dedicación:** 15h
- Grupo grande/Teoría: 6h
- Grupo mediano/Prácticas: 0h
- Grupo pequeño/Laboratorio: 0h
- Actividades dirigidas: 0h
- Aprendizaje autónomo: 9h

## Process Mining

**Descripción:**
Introduction to the novel concept of Process Mining and its application within the DM framework.

**Objetivos específicos:**
2

**Dedicación:** 9h
- Grupo grande/Teoría: 3h
- Grupo mediano/Prácticas: 0h
- Grupo pequeño/Laboratorio: 0h
- Actividades dirigidas: 0h
- Aprendizaje autónomo: 6h

## Data Visualization

**Descripción:**
As part of the DM stage of Data Exploration, we focus in the problem of Data Visualization.

**Objetivos específicos:**
3, 4

**Dedicación:** 15h
- Grupo grande/Teoría: 6h
- Grupo mediano/Prácticas: 0h
- Grupo pequeño/Laboratorio: 0h
- Actividades dirigidas: 0h
- Aprendizaje autónomo: 9h

## Basics of probability theory for intelligent data analysis

**Dedicación:** 15h
- Grupo grande/Teoría: 6h
- Grupo mediano/Prácticas: 0h
- Grupo pequeño/Laboratorio: 0h
- Actividades dirigidas: 0h
- Aprendizaje autónomo: 9h
Descripción:
Introduction to probability theory for intelligent data analysis, with a focus on Bayesian statistics

Objetivos específicos:
5

Statistical Machine Learning methods

Descripción:
The meeting of statistics and machine learning: Statistical Machine Learning methods, from the point of view of both supervised and supervised learning

Objetivos específicos:

SML in data visualization, with case studies

Descripción:
We merge the topics of SML and data visualization, illustrating its use with some real case studies

Objetivos específicos:
7

Sistema de calificación

The course will be evaluated through a final essay that will take one of these three modalities:
1. State of the art on a specific IDA-DM topic
2. Evaluation of an IDA-DM software tool with original experiments
3. Pure research essay, with original experimental content
270717 - IDADM - Análisis y Minería de Datos Inteligente

Bibliografía

Básica:


Complementaria:

