

FROM DATA TO KNOWLEDGE

An introduction to big data, open data and datamining technologies

UNIVERSITY OF ALICANTE

ALICANTE, 9-10TH MAY 2023

About the workshop

This workshop provides an overview of the knowledge discovery in databases pipeline, that is, the process of discovering useful knowledge from large collections of data.

The first part of the workshop focuses on the concepts of big data and open data. The adoption of open data initiatives by public administrations is increasing and there is a broad consensus in the public and private sector about the need for data management (within big data scenarios) to generate positive social and economic impact. However, there is a lack of understanding of the concept of open data itself within these big data ecosystems, as well as the opportunities and benefits of reusing open data when used in conjunction with the latest technologies (e.g., IoT and artificial intelligence). The lectures introduce the concept of open data and the opportunities and benefits that its reuse has for society, as well as other interesting concepts such as data sharing between private and public entities, open data in smart cities and smart tourist destinations, data-driven entrepreneurship and its promotion among university students, as well as the opportunity for universities that the proper management of open research data represents.

The second part of the workshop addresses the exploitation of data to extract knowledge by means of data mining techniques. The lectures introduce the concepts of data mining, exploratory data analysis (including different visualisation techniques) and machine learning. In addition to the analysis of structured data by means of data mining techniques, the workshop also covers different approaches to analyse and extract knowledge from textual information (i.e., text mining). Theoretical explanations are complemented with practical examples so that students can see these techniques in action.

Learning outcomes

After completion of this workshop the student should be able to:

- Understand the concept of big data scenarios and their relationship with the latest technologies (e.g. IoT and Artificial Intelligence).
- Understand technological and legal issues of open data (formats and licenses).
- Identify opportunities and benefits of reusing open data, as well as data sharing between private and public entities.
- Understand the fundamental concepts of data mining, including its definition, purpose, and applications in various fields.

- Describe the various stages of the data mining process, including data collection, preparation, modelling, and evaluation.
- Apply exploratory data analysis techniques, including central tendency measures and visualisation, to gain insights into the data and prepare it for modelling.
- Apply machine learning algorithms such as support vector machines, decision trees, and neural networks, to build predictive models and evaluate their performance.
- Understand the principles and techniques of text mining.
- Apply text mining techniques to analyse text data, such as customer reviews, and news articles.

Duration

35 hours. 21 hours of individual work online. 14 hours of face-to-face lessons.

Literature Review: students will be provided with relevant papers to read before attending to the workshop.

Final report: students should write a report of a particular section of the workshop that will be assigned.

Workshop teachers

Jose Norberto Mazón Lopez, University of Alicante

David Tomás Díaz, University of Alicante

Workshop Schedule

Day 1

Time	Content
09:00 – 10:00	Introduction to Big Data and Open Data
10:00 – 11:00	Licenses and Open Data
11:00 – 12:30	Open Data formats
12:30 – 13:00	FAIR principles for data
13:00 – 14:00	Lunch break
14:00 – 15:00	Reusing Open Data from Open Data portals
15:00 – 16:00	European Strategy for Data
16:00 – 17:00	Case studies on Open Data

Day 2

Time	Content
09:00 – 10:00	Introduction to Data Mining
10:00 – 11:30	Exploratory Data Analysis
11:30 – 13:00	Machine Learning techniques
13:00 – 14:00	Lunch break
14:00 – 16:00	Introduction to Text Mining
16:00 – 17:00	Text Mining applications