

The background of the entire image is a dense, abstract network of colored lines. These lines are primarily in shades of green, blue, red, and purple, creating a complex web-like pattern that suggests data flow or connectivity. The lines are thicker in some areas, forming a more structured grid, while in others, they form a chaotic, organic mass.

Universitat de Barcelona

Data Science & Big Data

A part-time course to train the new generation of data scientists.



Oriol Pujol



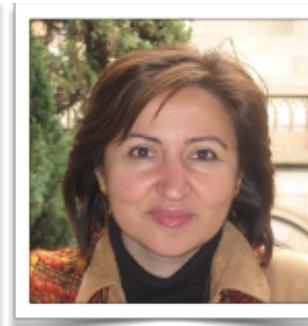
Jordi Vitrià



Albert Diaz



Josep Perelló



Petia Radeva



Laura Igual



Lluis Garrido



Eloi Puertas



Santi Seguí



Montse Guillen



Jordi Nin



Sergio Escalera



Introduction

Universitat de Barcelona's Data Science and Big Data course offers students a program that covers the concepts and tools you will need throughout the entire data science pipeline: asking the right questions; wrangling and cleaning data; generating hypothesis; making inferences; visualizing data; assessing solutions; and building data products.

Schedule

Oct 2, 2018 - July 2, 2019, every Tuesday and Thursday
18h-20h

Location

Aula T1, Edifici Històric de la Universitat de Barcelona,
Gran Via de les Corts Catalanes 585, 08007, Barcelona.



Requirements:

The program is specially designed for students with a background in computer science, mathematics, and applied statistics, but other scientific and engineering backgrounds can be considered.

We will require you to follow lessons and complete class exercises using personal laptops. You will not be able to complete all your assignments in class if you rely solely on desktop equipment at home.

<https://github.com/DataScienceUB/Postgrau>

Before the first class you need to:

- Install Anaconda Python 3.6 version: Anaconda Distribution is a free, easy-to-install package manager, environment manager and Python distribution with a collection of over 720 open source packages with free community support.
- Have a (free) account at GitHub: GitHub is a web-based Git or version control repository and Internet hosting service. It offers all of the distributed version control and source code management (SCM) functionality of Git as well as adding its own features.

Raw Data

2. Processing: How I do clean and separate my data?

- Identification: filter data.
- Outliers.
- Imputation: missing value processing.
- Reduction: dimensionality reduction.
- Normalization: duplicates, ranges, format, coordinates, units, etc.
- Feature extraction.

3. Enrichment: How do I add more information to my data?

- Feature engineering.
- Search for additional data sources.

6. Analyze: How do I model my data?

- Variable selection (How do I determine important variables?)
- Probabilistic modeling (How are my variables related?)

8. Evaluate: Are the outcomes generic and robust?

- Statistical Testing.
- Model performance.

Data Science Path

1. Acquire: How do I get my data?

- Web Scraping.
- Data Base queries.
- Access to bulk data stores.

4. Aggregation: How do I collect and summarize my data?

- Basic Statistics: mean, std, box plots, scatter plots, counts, etc.
- Distribution fitting.
- Feature aggregation.

5. Discover: What are the key relationships in my data?

- Clustering (How do I segment the data to find natural groupings?)
- Visualization (Are there unexpected relationships?)

7. Predict: What are the likely future outcomes?

- Regression (How do I predict the future?)
- Classification (How do I predict a category?)
- Recommendation (How do I predict relevant conditions?)

Insights

Activities

twitter: [@datascienceub](#) | [@databeersbcn](#)

DataScience@alumni

MasterClasses

Evaluation

Capstone Project: An important part of the course is the capstone project!