Universitat de Barcelona

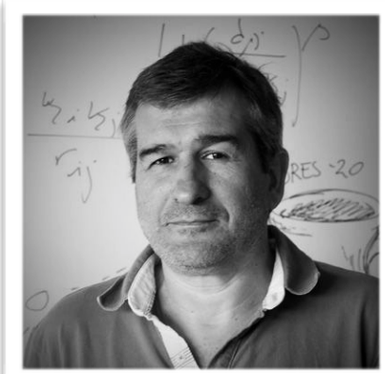
Data Science & Big Data

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

www.ub.edu/datascience

Oriol Pujol

Jordi Vitrià



Albert Diaz

Josep Perelló

Petia Radeva

Laura Igual

Lluis Garrido

Eloi Puertas

Santi Seguí

Montse Guillen

Mireia Ribera

UNIVERSITAT DE BARCELONA

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 1, 2019 - July 2, 2020, 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.

UNIVERSITAT DE 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 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 3.7 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

1. Acquire: How do I get my data?
- Web Scraping.
- Data Base queries.
- Access to bulk data stores.

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.

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?)

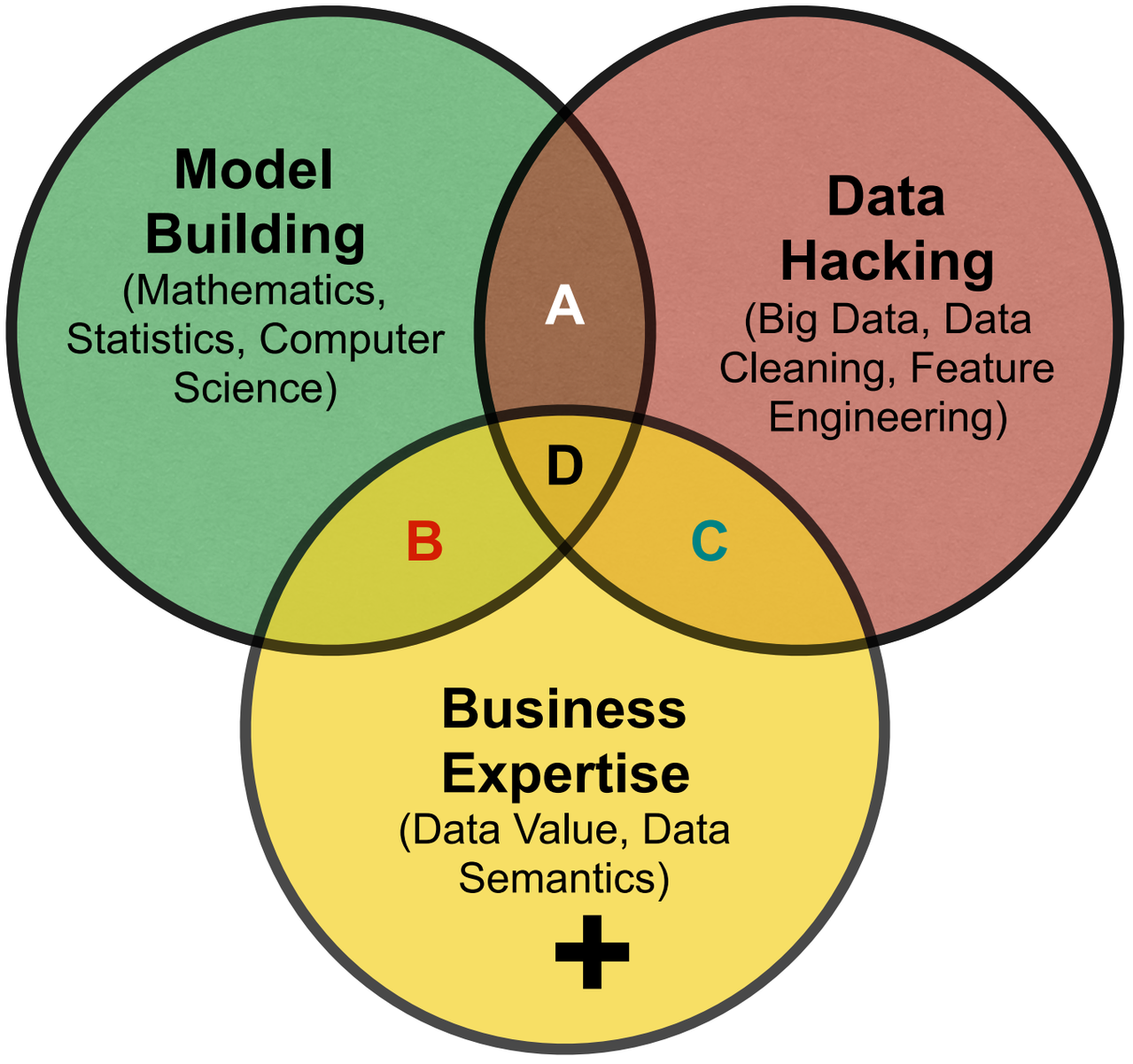
6. Analyze: How do I model my data?
- Variable selection (How do I determine important variables?)
- Probabilistic modeling (How are my variables related?)

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?)

8. Evaluate: Are the outcomes generic and robust?
- Statistical Testing.
- Model performance.

Data Science Path

Insights


Model Building
(Mathematics, Statistics, Computer Science)
Data Hacking
(Big Data, Data Cleaning, Feature Engineering)
Business Expertise
(Data Value, Data Semantics)
+
A
B
C
D

DataLabs

accenture

Masterclasses

letgo Kernel analytics TÜVRheinland® Precisely Right. BBVA DATA & ANALYTICS socialpoint King SATELLOGIC ABB

Activities

twitter: @datascienceub

DataScience@alumni

Job Offers

Evaluation

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