

---

# Capstone Project

## Postgraduate Course on Data Science & Big Data



UNIVERSITAT<sub>DE</sub>  
BARCELONA

---

- 
- **Pablo Martínez**  
pmartima22@gmail.com
  - **Guillem Pascual**  
guillem.pascual@ub.edu
-

---

# The problem

---

[illegible]

---

# of politicians and parties in Twitter



**Xavier García Albiol** ✓ @Albiol\_XG · 18 nov.

Vergüenza e indignidad. No se puede humillar a más catalanes que nos sentimos españoles, en menos tiempo.



**Miquel Iceta Llorens** @miqueliceta · 4h

Cada día, un haiku. Hoy os ofrezco un poema de Natsume Soseki [poesiajaponesa.wordpress.com/2018/11/19/nat...](https://poesiajaponesa.wordpress.com/2018/11/19/nat...) via @miqueliceta



**Quim Torra i Pla** ✓ @QuimTorraIPla · 4h

Una altra evidència que a l'estat espanyol la justícia és un mercadeig entre el PSOE i el PP i que l'estat de dret i les llibertats no estan garantides per a cap ciutadà. No hi ha independència judicial, ni imparcialitat ni integritat.



**Inés Arrimadas** ✓ @InesArrimadas · 23h

Esto pasa cuando el Gobierno tiene como socios a quienes quieren liquidar España: se gastan el dinero de todos en propaganda e ilegalidades. Con tal de estar más tiempo en Moncloa, Sánchez prefiere mirar hacia otro lado que defender a los españoles.



---

# Team-based project

- 4 people per team
- 4 practical sessions
- All year long







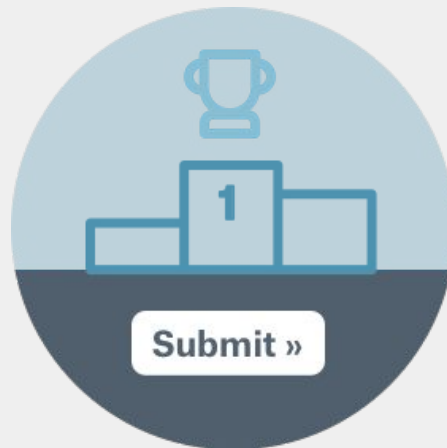
**Design the best learning  
model and beat your  
classmates!**

---

# Kaggle in Class

## Two datasets

- Public
- Private



Real Time classification

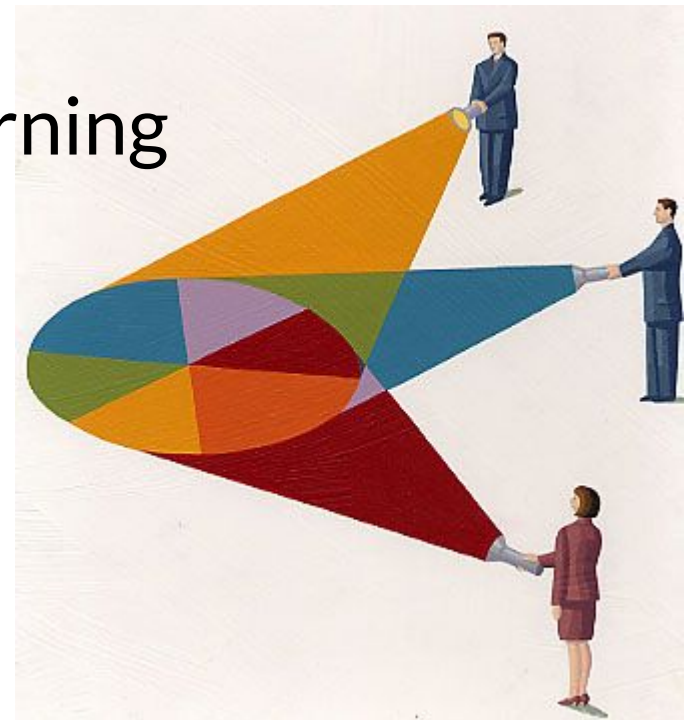
---



---

# Features a wide range of fields

- Feature Engineering
- Natural Language Processing
- Labeling
- Supervised Learning



---

# Data

---



6	Parties
2400	Tweets
12	Politicians

---

# Notebook time!

---

---

# Project outline

---

---

## 0. Get in touch with the data

... Explore the data provided in this session.

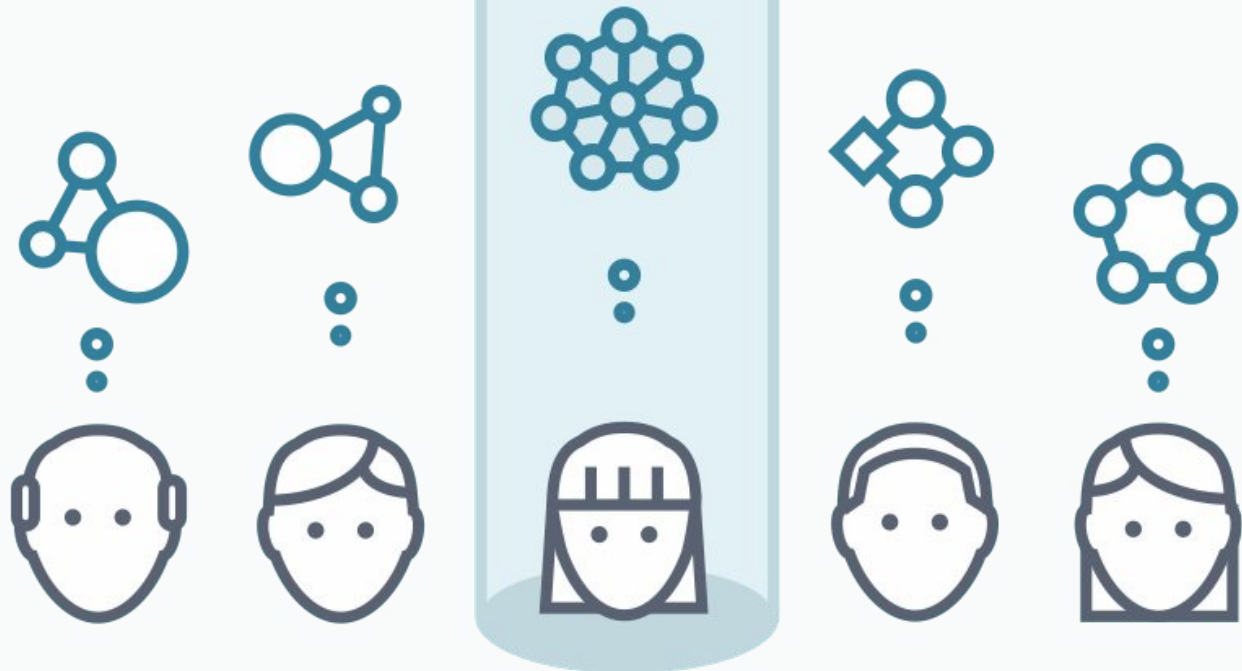




---

## 1. Join Kaggle

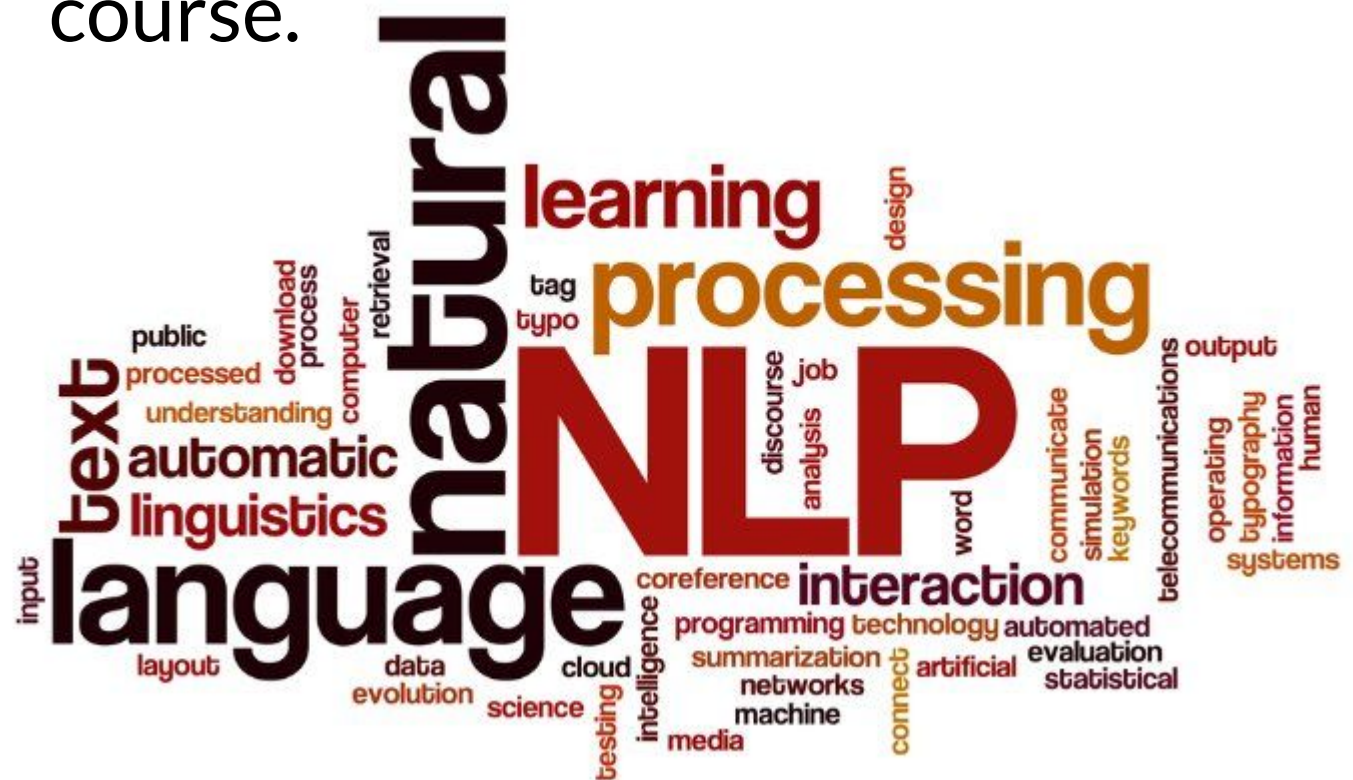
... Start competing with the rest of the groups to obtain the best score.



---

## 2. Feature engineering

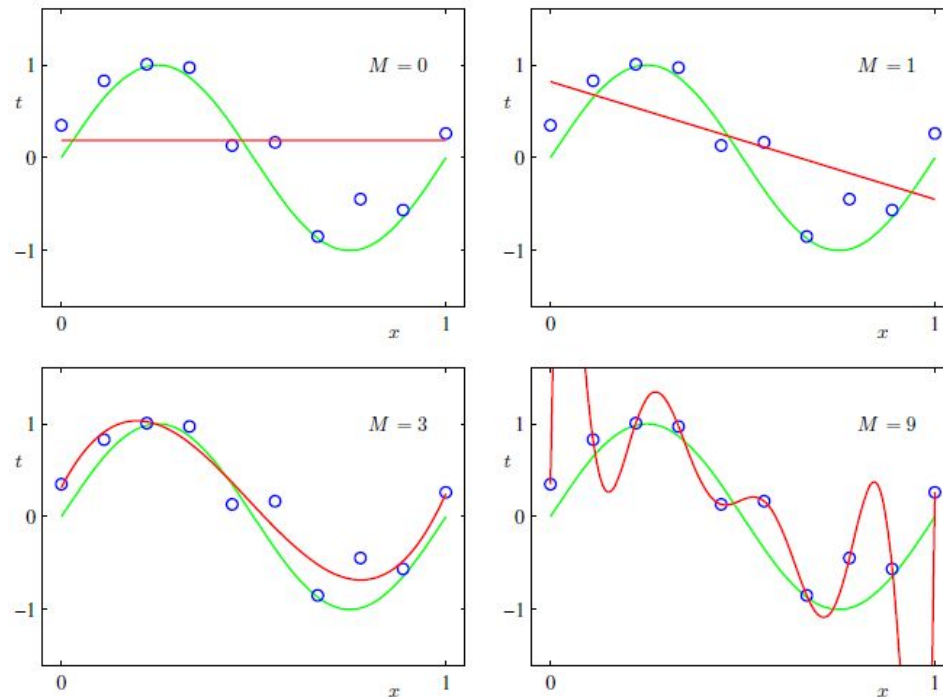
... Improve the model using NLP techniques provided along the course.



---

### 3. Model selection

... Push the limits finding the optimal model and parameters.



---

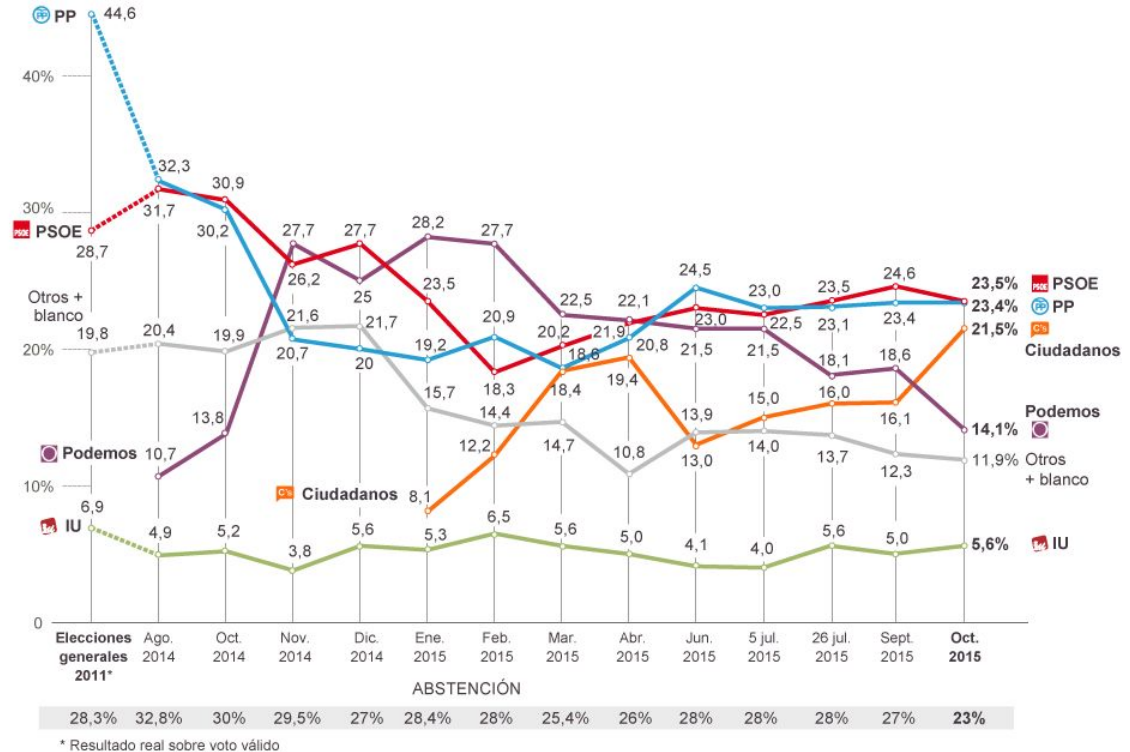
## 4. Get your best score!

... Deliver your model and beat the accuracy obtained by other groups.



## 5. Create a Data Story

... Explain your data story with graphics and models.



---

**Next steps?**

---



---

# Capstone Project Session

## January 31st

- Apply the concepts learnt in class to explore/solve the proposed project **only** by doing data engineering, without modifying the model
  - Mandatorily submit a model to Kaggle
- Professors will act as Expert Consultants during all sessions. So come prepared and ask as many questions as you want!

---

# NOW?

It's time to meet each other  
and  
create a team

---