

# Text Mining II

## Máster Big Data Analytics

### (2015 - 2016)

Dr. Paolo Rosso

Casi-casi-Dr. Francisco Rangel ;-)



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA

# Planificación

Sábado 28	Sábado 4	Viernes 10	Sábado 11
Presentación del módulo (5min)	Construcción de un baseline (1h)	Introducción al Author Profiling (2h)	Prácticas con HispaTweets (3h)
Text Mining en las empresas (1'5h) Caso real: Autoritas	Prácticas con Hispatweets (3h)	Prácticas con HispaTweets (2'5h)	
Exploración del dataset HispaTweets (3h) Ejercicio práctico	EmoGraphs (0'5h)		Presentaciones PechaKucha (1'5h)
30 min descanso			

# Requisitos

- Conocimientos de programación.
  - En el marco del máster se recomienda el uso de Python y/o R.
- Conocimientos de Latex (o en su defecto Word).
- Cuenta en Github.

# Material

- Guión para la exploración del dataset.
- Guión para la generación del baseline.
- Presentaciones y material teórico y auxiliar. Referencias.
- Vídeos de optimización.
- Esqueleto de artículo.

# Entregables

En la cuenta de Github del alumno:

- Artículo realizado individualmente contando la participación del alumno en la tarea.
  - La longitud máxima del artículo será de 4 hojas adaptadas al formato proporcionado.
- Presentación PechaKucha (4 min) de los resultados alcanzados y los aspectos remarcables de su aproximación, especialmente desde el punto de vista del Big Data.
- Código fuente de los scripts utilizados para la experimentación más un guión descriptivo para su reproducibilidad.

# Criterios de evaluación

- Interacción en clase (10%)
- Presentación PechaKucha (40%)
  - Resultados alcanzados, desde la doble perspectiva precisión-tiempo.
- Trabajo entregable (50%)
  - Artículo
  - Código fuente + guión de ejecución