

# Data Science Foundations

Master in Big Data Solutions 2017-2018



Liana Napalkova  
liana.napalkova@bts.tech

Ludovico Boratto  
ludovico.boratto@bts.tech

Francisco Gutierrez  
francisco.gutierrez@bts.tech

# Session 5

## **Introduction to Text Mining (Part 2)**

### *Assignments*

# Assignments

1. Revise the slides to make sure that the different concepts are clear (e.g., could you tell in your own words what's the difference between "Part-of-speech tagging" and "Name Entity Recognition")
2. Create 10 txt files with 10 lyrics from your favorite artist (or of 10 favorite songs). Open each file, removing the new line commands (to have a unique string with the whole lyrics):

```
with open('lyrics.txt', 'r') as myfile:  
    data=myfile.read().replace('\n', '')
```

- a. Provide 5 examples of salient words
- b. Compute the TF-IDF score of each lyric