



MADRID

**inter.noise 2019**

June 16 - 19

NOISE CONTROL FOR A BETTER ENVIRONMENT

# Evaluating noise perception through online social networks: A text mining approach to design a noise-event alarm system based on social media content

Luis Gascó<sup>1,2</sup>, César Asensio<sup>1</sup>, Guillermo de Arcas<sup>1</sup>, Chloé Clavel<sup>3</sup>

1



POLITÉCNICA

1



2



3



# Crowdsourcing

*People act like human sensors voluntarily*



# Social computing



*People express their feelings directly on Social Media*



Our method



Is it a **noise complaint**?

Yes

Which is the **noise source**?

Music

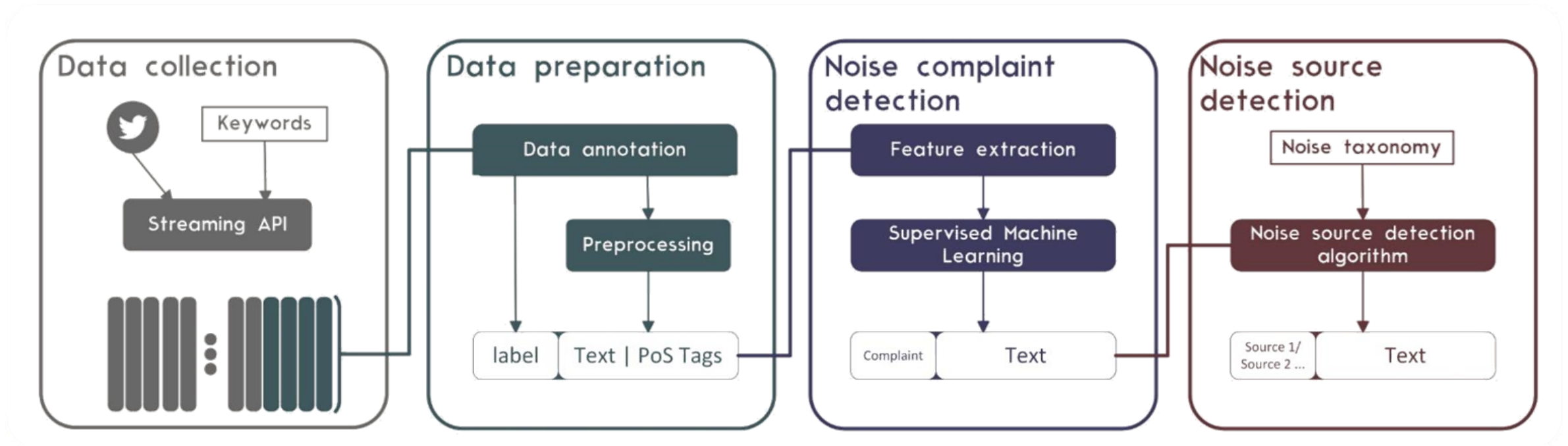
Nature

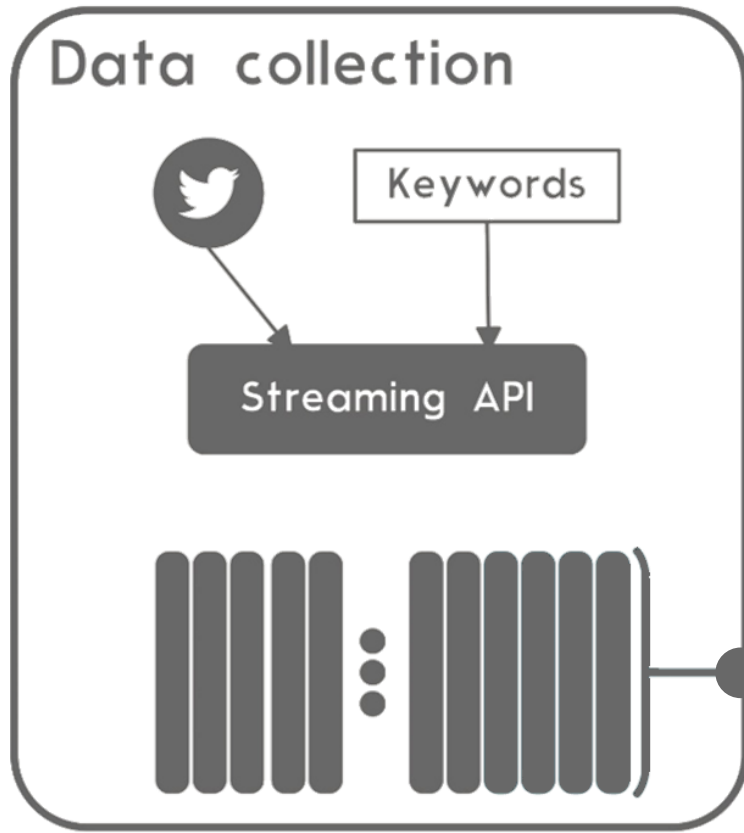
Indoor

Transport

Human

Mechanical





**1<sup>st</sup> June and 1<sup>st</sup> September 2017**

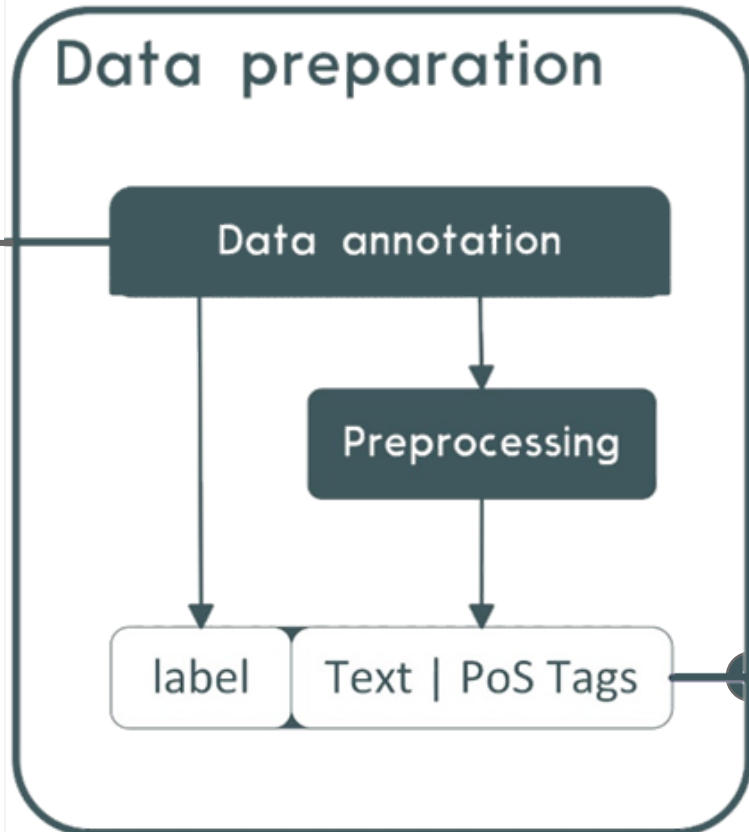
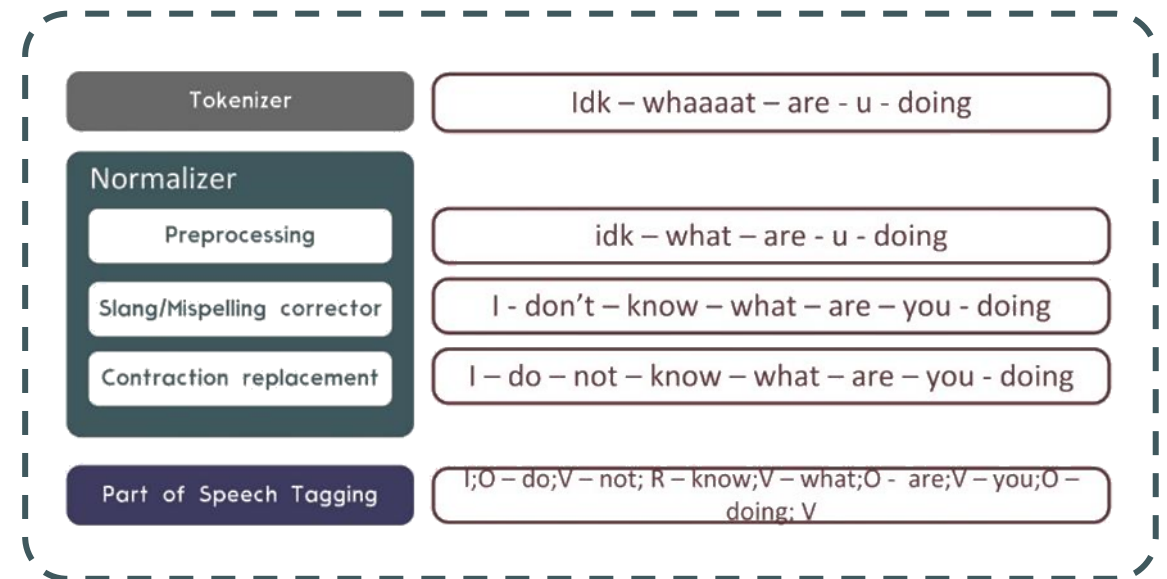
**+5,6M tweets**

**+840k tweets**

## Manual annotation of tweets

- 2 categories: Noise complaint or not
- ~10k annotated tweets
  - Complaints: 580
  - Other: 9427

## Preprocessing



## Noise complaint detection

Feature extraction

Supervised Machine Learning

Complaint

Text

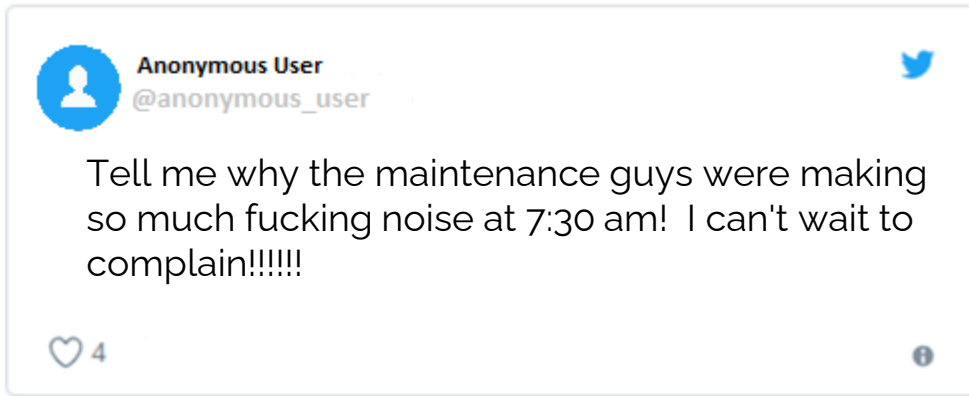
### Feature extraction

- N-grams
- PoS Features
- Sentiment features
- Embeddings

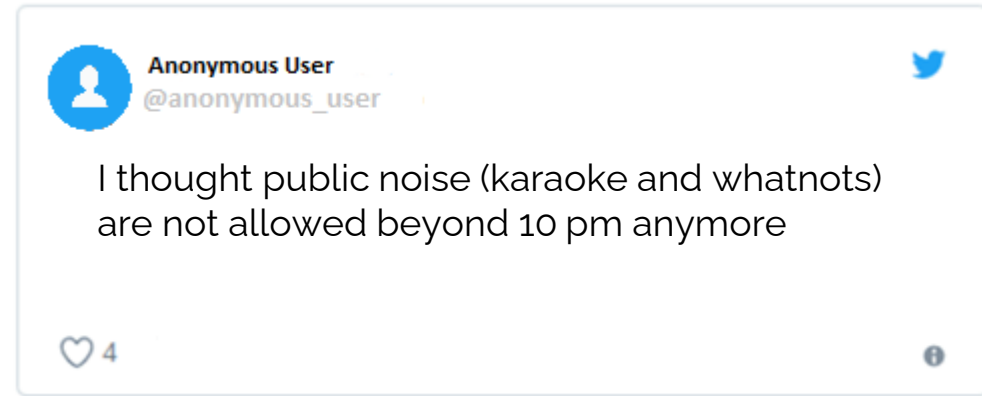
### Maximum Entropy Classifier

- TPR = 0,85
- FPR = 0,16



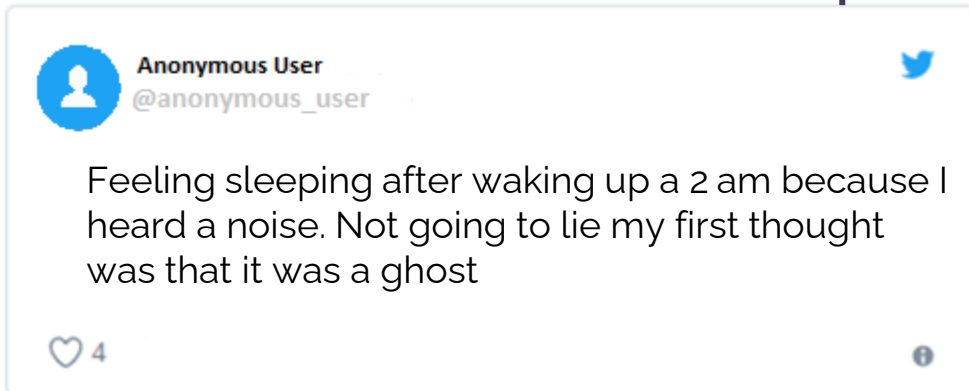


True class: 1  
Predicted class: 1

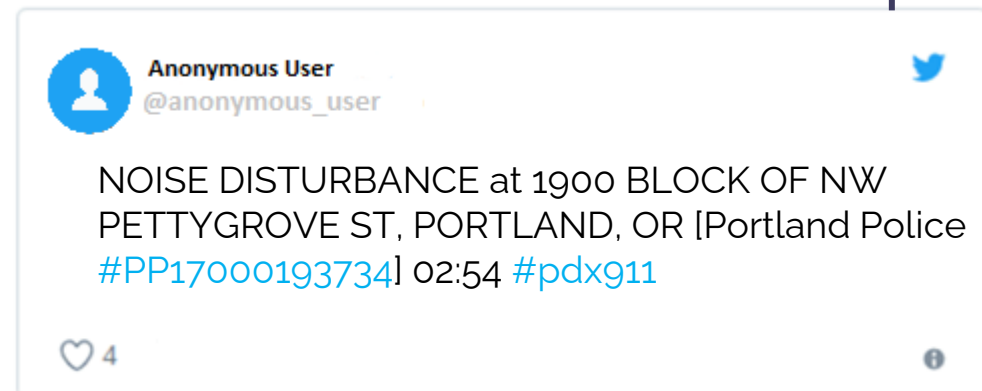


True class: 1  
Predicted class: 0

True class: 0  
Predicted class: 1



True class: 0  
Predicted class: 0



## Noise source detection

Noise taxonomy

Noise source detection algorithm

Source 1/  
Source 2 ...

Text

## Chatty maps taxonomy + WordNet

- Transport
- Nature
- Human
- Music
- Indoor
- Mechanical

## Presence of words in the lexicon

**86% labels correctly predicted**

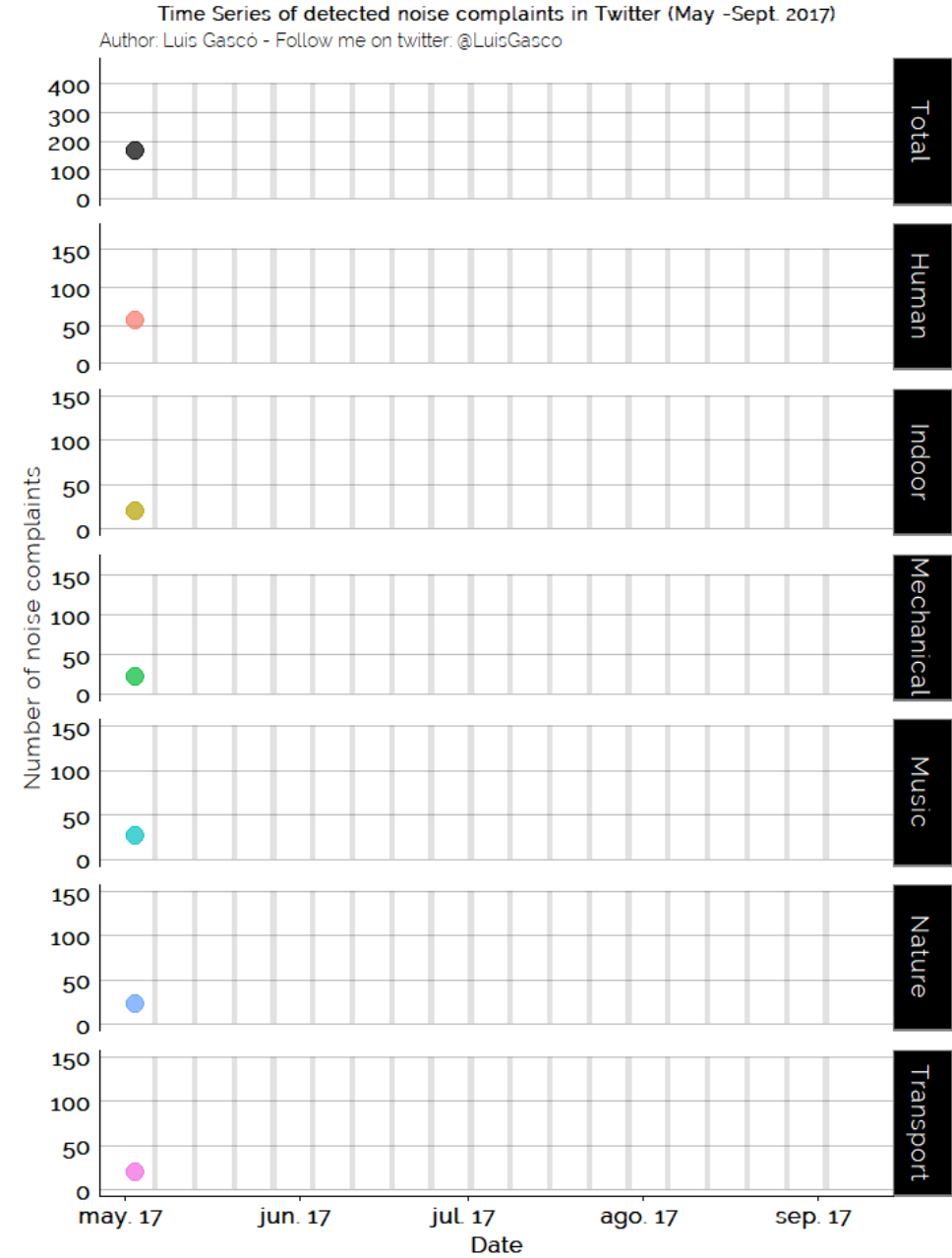
# **Case study**

*Noise event alarm system*

# Noise complaint time series



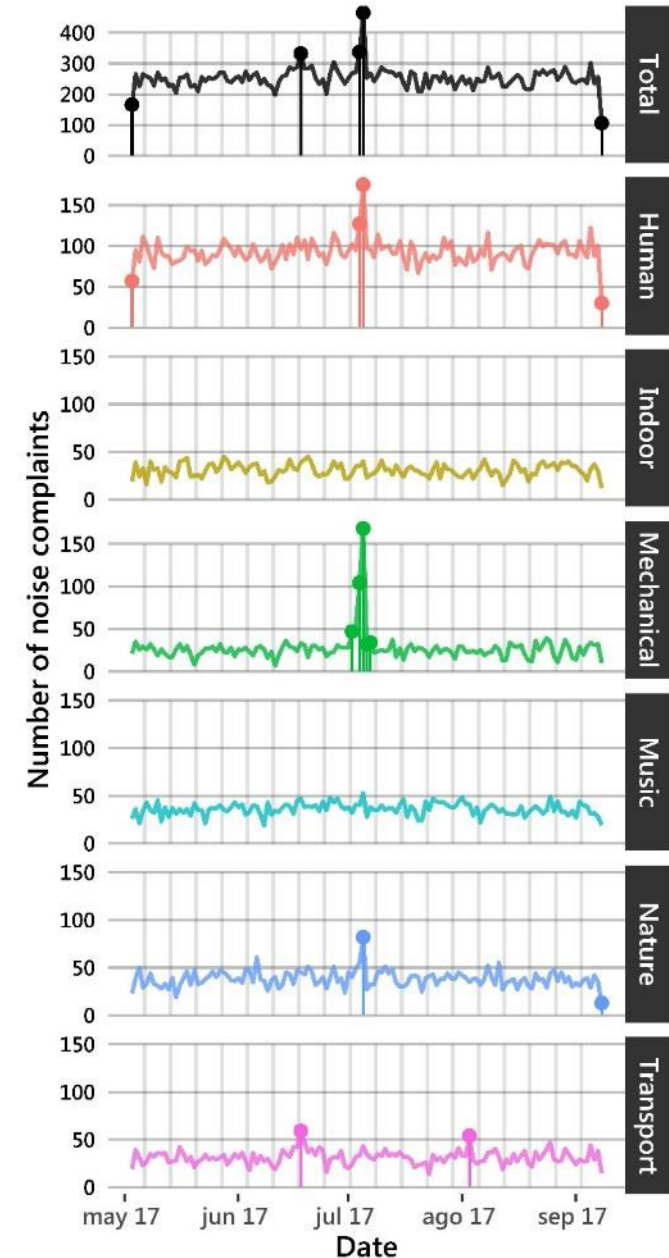
- **32k complaints**
  
- **19k with at least one noise source**



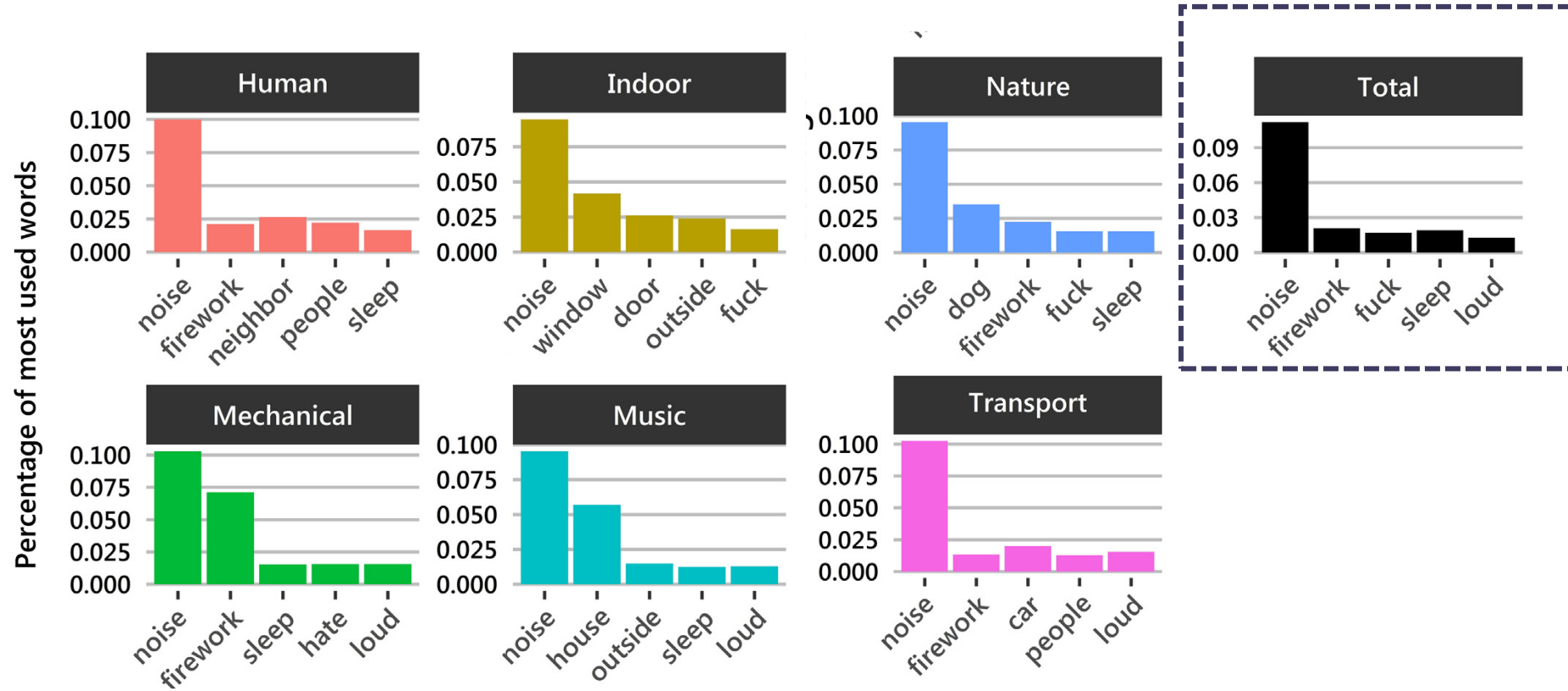
# Anomalous days detection



- **AutoRegressive Integrated Moving Average (ARIMA)**
- **Anomalous days ~ noise problem with those noise sources**
- **How could know the origin of those complaints?**

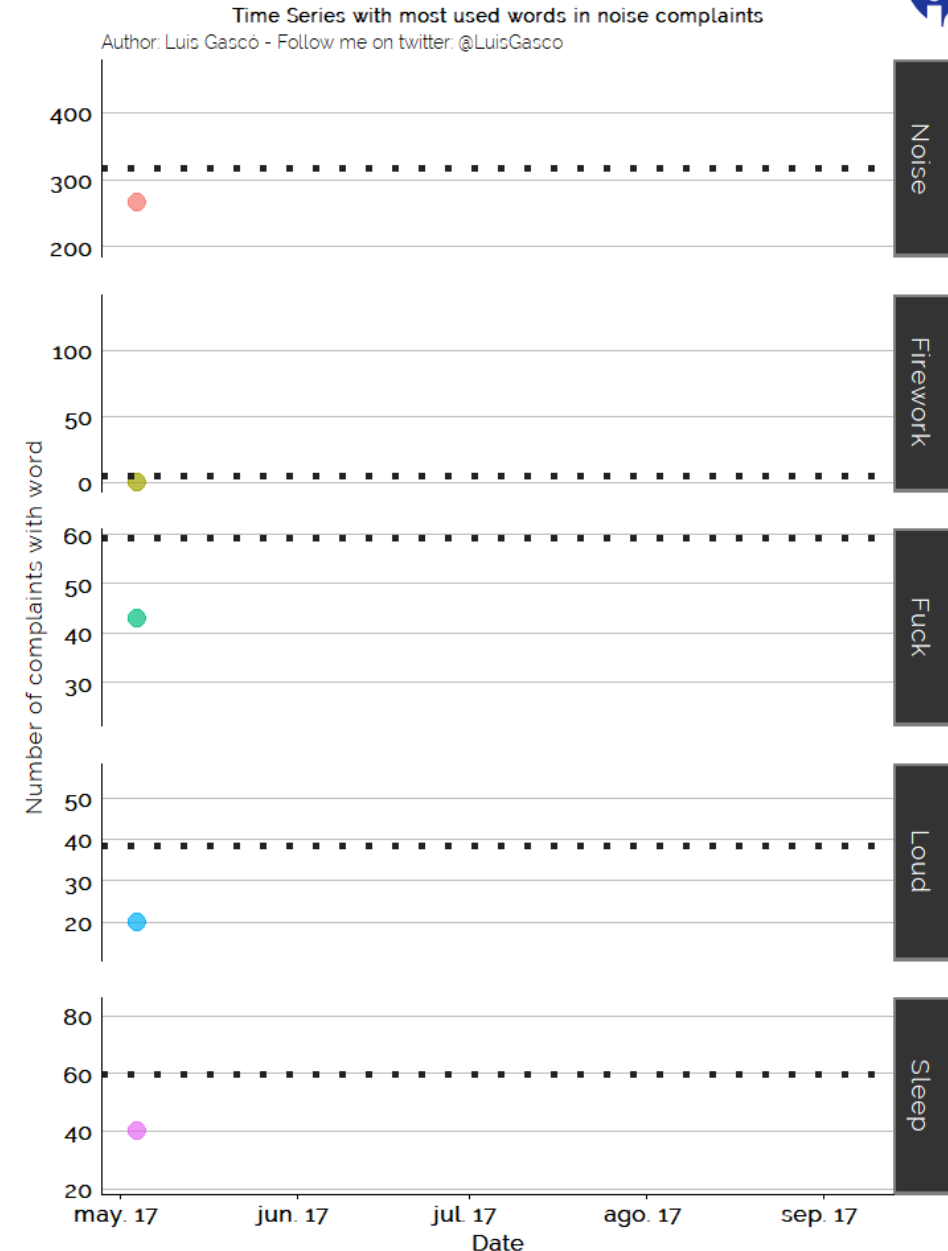


# Most used words during anomalous days for each sound category



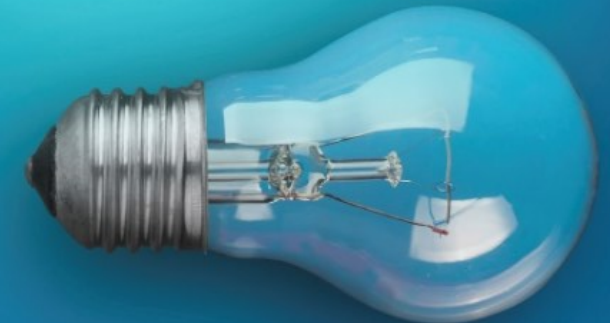
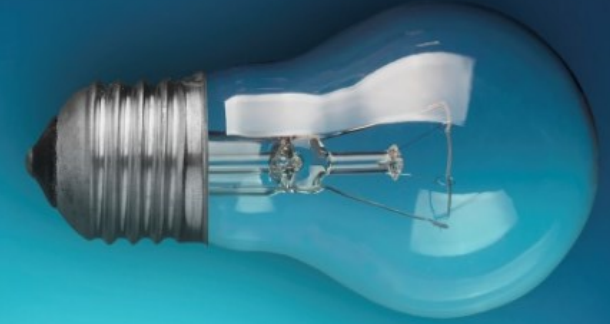
# Noisy activities detector

- **Statistical Process Control theory**
- **Limit cover 99,7% of normal days**
- **If one day that limit is exceeded, we may conclude that a noise event related to fireworks has been detected.**
- **Detecting big festivities based on appearance of some words on the complaints**



# Conclusions

- Development and testing of a noise complaint detector with sound source identification based on data from online social networks.
- Acoustic annoyance alarm system for the detection of noisy events, which can be implemented by city managers to measure the effectiveness of actions against noise in this type of events.





WHAT'S  
NEXT ?



**Incorporate to the methodology the detection of annoyance levels**

**Adapt the methodology to analyze other data sources with longer texts**

**Development of labelled twitter dataset to be used as a basis for future models**

**Development of tools to facilitate research in this field, e.g. collaborative annotating tools**



# Evaluating noise perception through online social networks: A text mining approach to design a noise-event alarm system based on social media content



CAMPUS  
DE EXCELENCIA  
INTERNACIONAL



INSTRUMENTATION &  
APPLIED ACOUSTICS  
RESEARCH GROUP



Digital  
DOCTORAL SCHOOL

## Luis Gasco Sánchez



[luis.gasco@i2a2.upm.es](mailto:luis.gasco@i2a2.upm.es)



<https://www.luisgasco.es>



<https://www.linkedin.com/in/luisgascosanchez/>



[https://www.researchgate.net/profile/Luis\\_Gasco\\_Sanchez](https://www.researchgate.net/profile/Luis_Gasco_Sanchez)



[@luisgasco](https://twitter.com/luisgasco)