

WOMEN IN DATA SCIENCE



A tag is worth a thousand words



Arancibia Sofia Casandra Data Scientist, McAfee









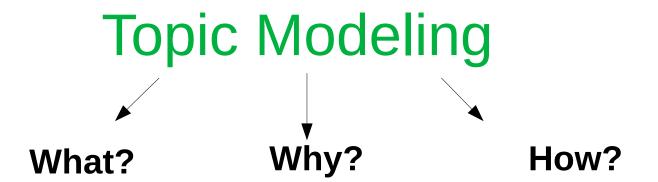


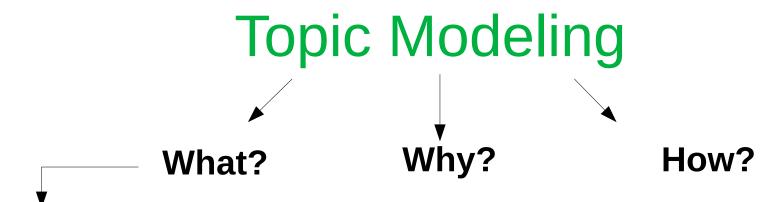




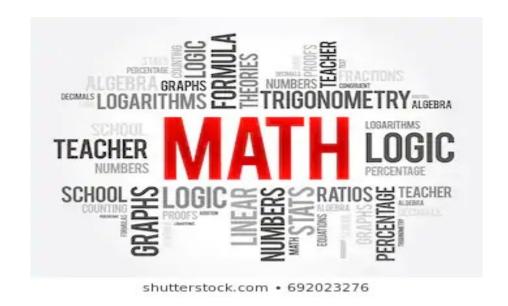








- The subject of a document
 - sports,
 - politics,
 - travelling, ...
- Word clusters



Word Frequency

> 1 latent topic

Topic Modeling What? Why?

- The subject of a document
 - sports,
 - politics,
 - travelling, ...
- Word clusters

How?

- LSA

- pLSA

- LDA

— What?

Why?

- How?
- LSA
- pLSA
- LDA

- The subject of a document
 - sports,
 - politics,
 - travelling, ...
- Word clusters



----- What?

- The subject of a document
 - sports,
 - politics,
 - travelling, ...
- Word clusters

Why?

Document content understanding and classification

How?

- LSA

- pLSA

- LDA





Jugar

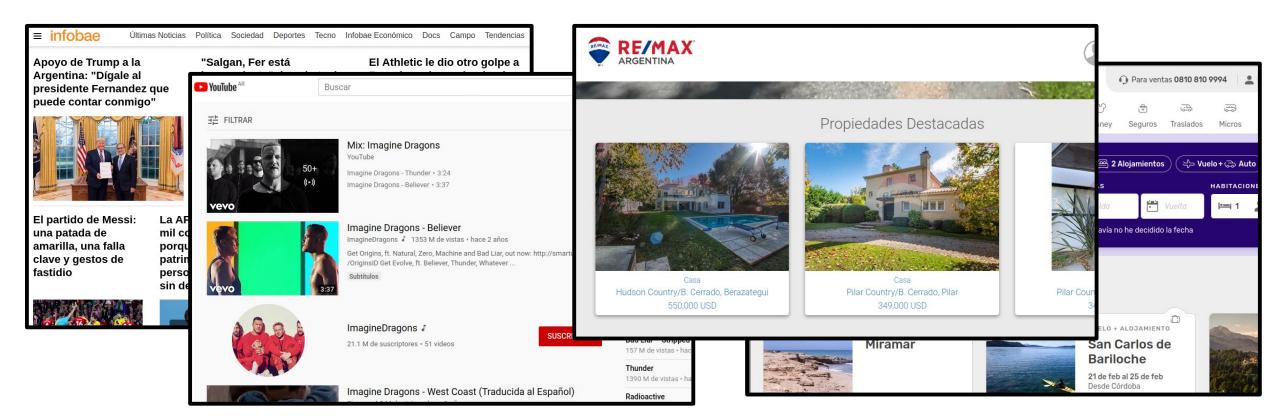
Jugador

Ocio

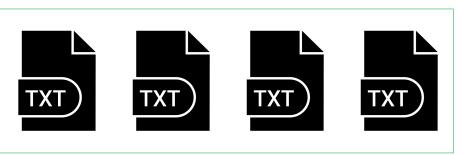
Online

Dinero





raw data





Data Gathering



raw data

Data Preparation

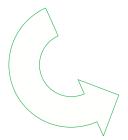




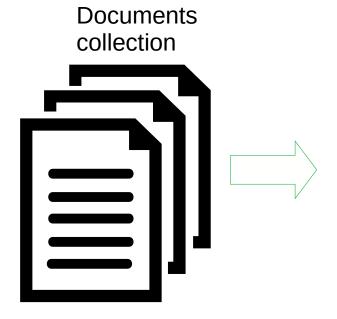




Data preprocessing

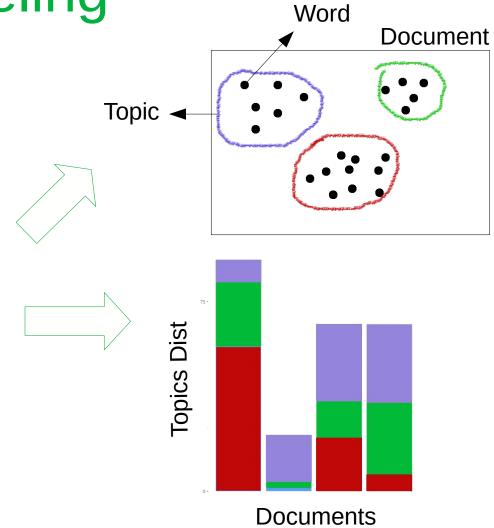


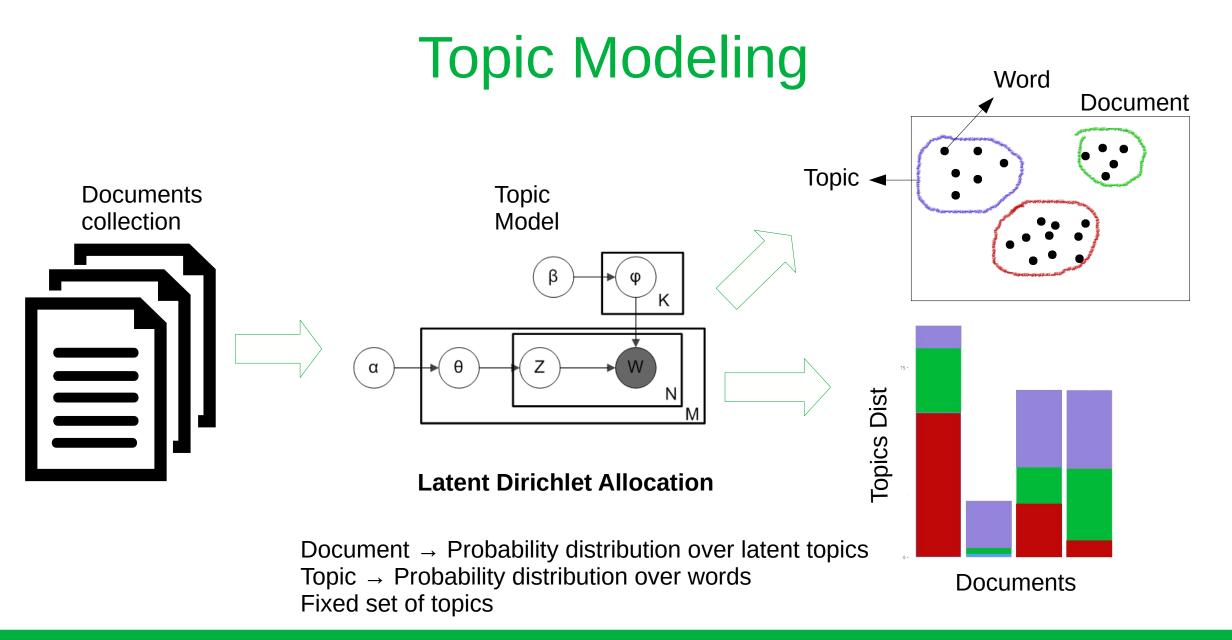


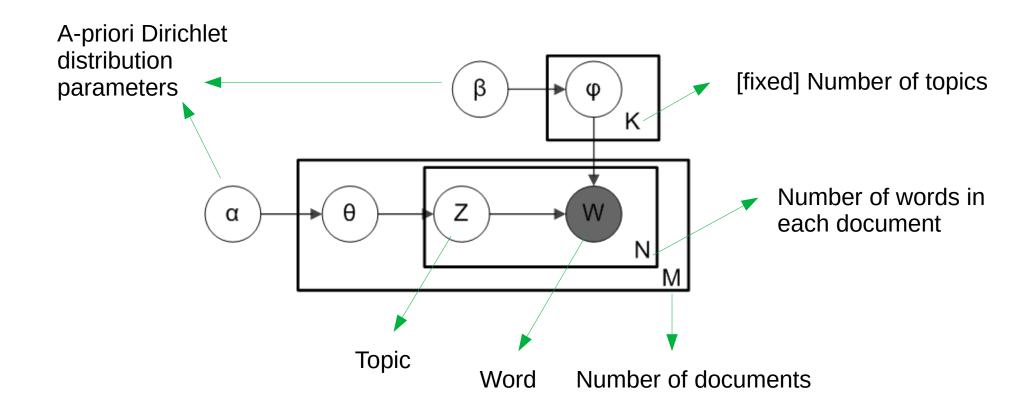


Topic Model









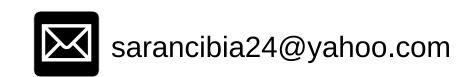
Demo













Gracias!





