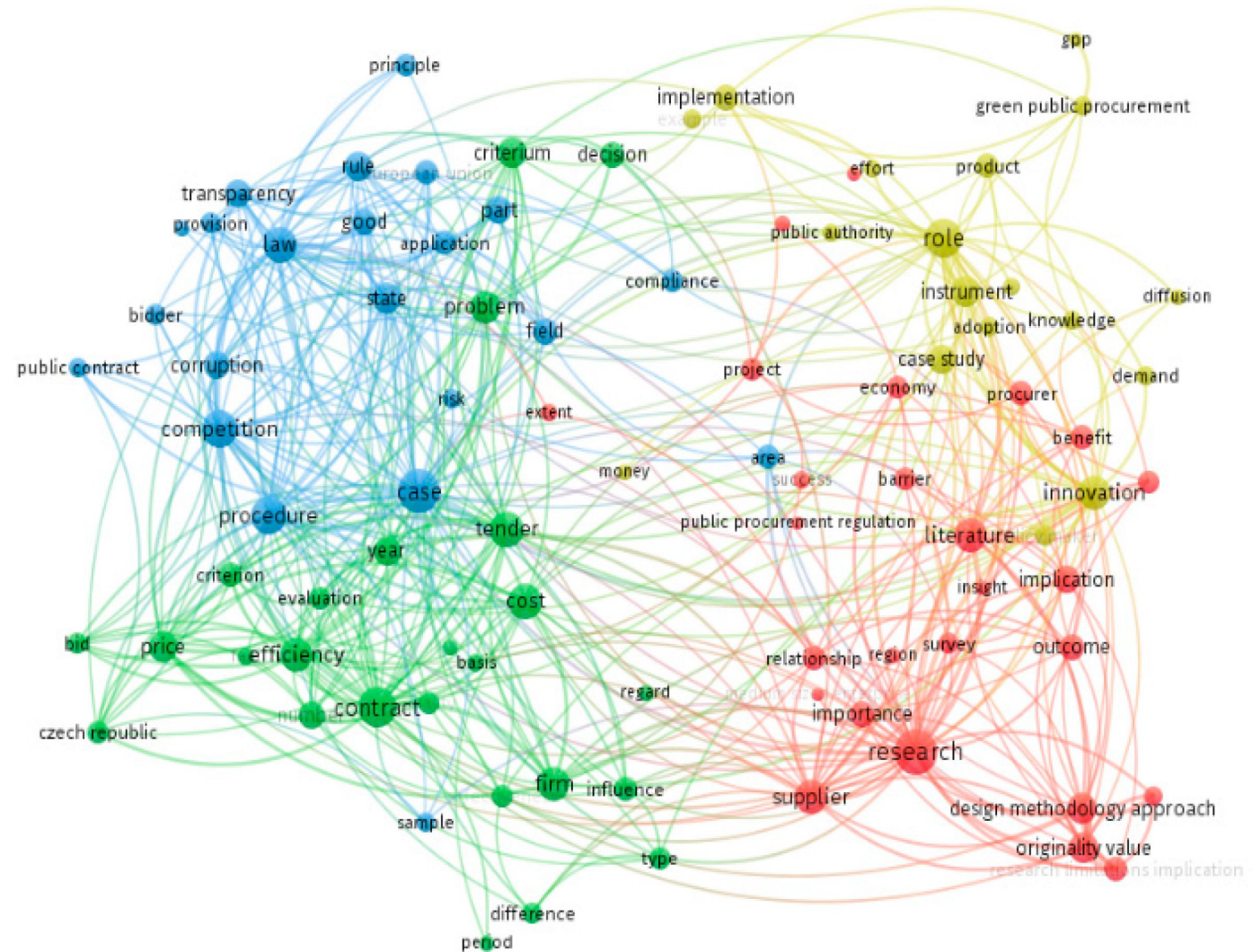# A neuronal network method for predicting fraud and corruption risk in public procurements

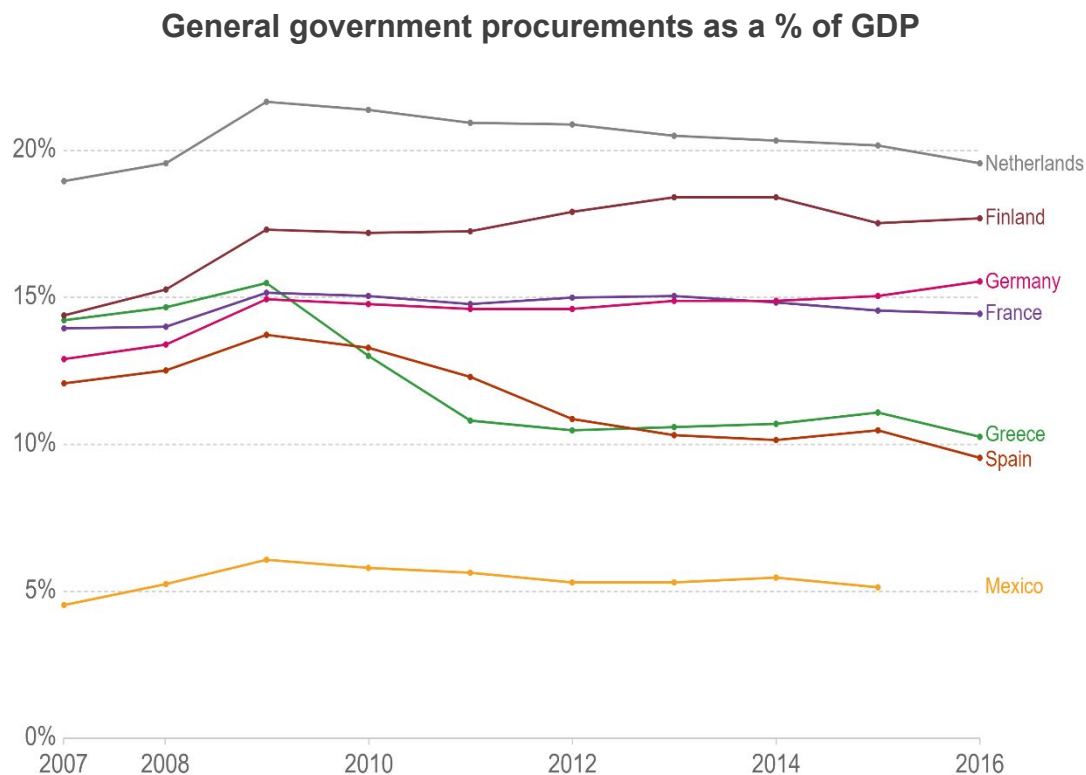**David Blanco Alcántara**
Félix. J López-Iturriaga
Iván Pastor Sanz
Elena Temiño Santamaría

Public procurements: some figures

Public procurement is the process of **purchasing goods, services or works** by the public sector **from the private sector.**

Public procurement represents **a substantial share of world trade**:

General government procurements as a % of GDP

Source: Government at a Glance – OECD (2017)   OurWorldInData.org/public-spending/ • CC BY

It is also considered as **a potential source of corruption**:

Four anti-corruption priorities for Southern Africa during COVID-19

Along with countries in Latin America and the Middle East, Southern African countries face serious corruption risks in addressing vulnerabilities across national health care systems.

As in other regions, public procurement, which plays an essential and often life-saving role in the response to the pandemic, also presents significant opportunities for corruption.
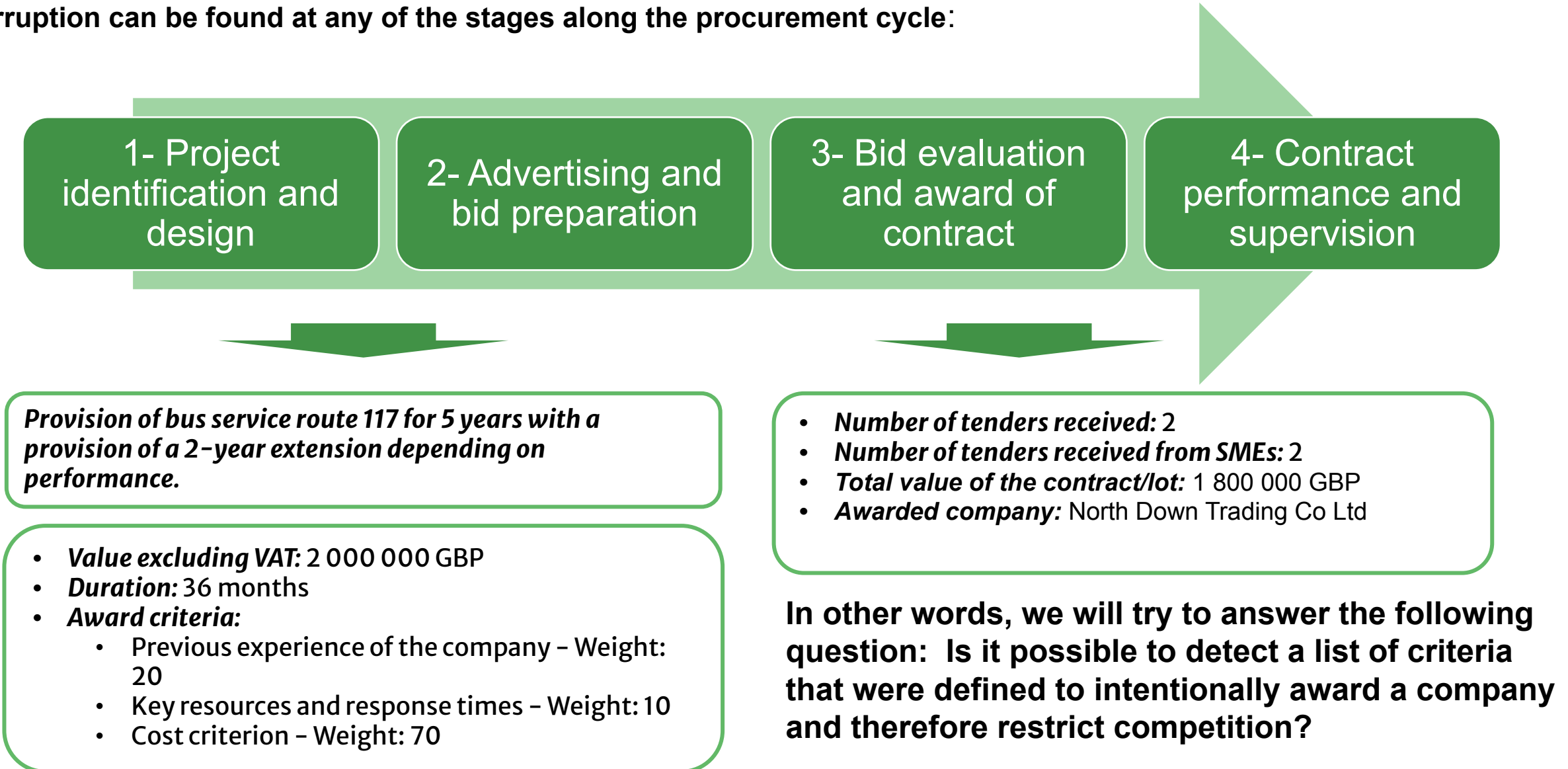
COLUMNISTS

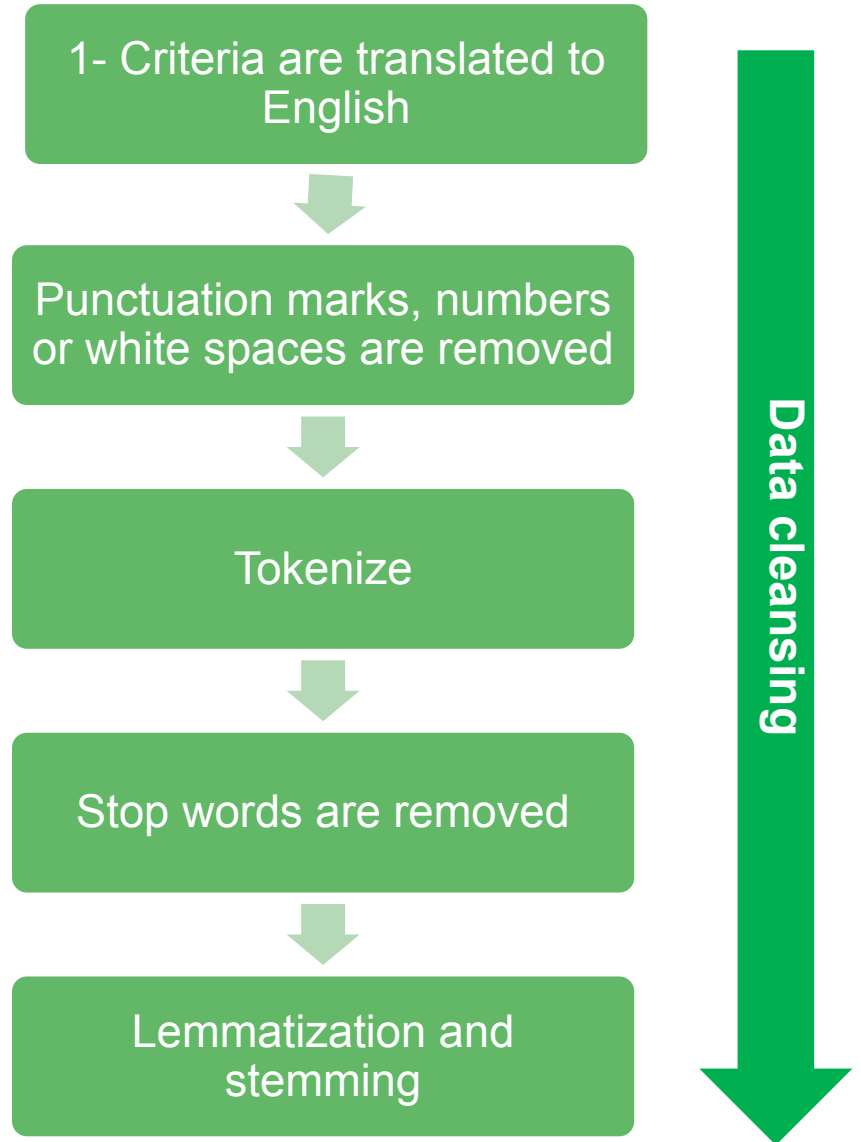## Refine training to tame procurement corruption in Kenya

TUESDAY JANUARY 12 2021

The public procurement cycle

Corruption can be found at any of the stages along the procurement cycle:

1- Project identification and design	2- Advertising and bid preparation	3- Bid evaluation and award of contract	4- Contract performance and supervision

Provision of bus service route 117 for 5 years with a provision of a 2-year extension depending on performance.

- *Value excluding VAT:* 2 000 000 GBP
- *Duration:* 36 months
- *Award criteria:*
  - Previous experience of the company – Weight: 20
  - Key resources and response times – Weight: 10
  - Cost criterion – Weight: 70

- *Number of tenders received:* 2
- *Number of tenders received from SMEs:* 2
- *Total value of the contract/lot:* 1 800 000 GBP
- *Awarded company:* North Down Trading Co Ltd

In other words, we will try to answer the following question: Is it possible to detect a list of criteria that were defined to intentionally award a company and therefore restrict competition?

The dataset

- **Data** used in this paper **comes from the Tenders Electronic Daily (TED)** portal from **January 2016 to December 2018.**

- We use information of **72.466 tenders from 33 different countries**

- We consider the tenders award criteria to identify subjective criteria where the level of competition is significantly low (bid received lower than 3).

- Award criteria should be treated to the analysis, there are more than 1.000 different criteria.

1- Criteria are translated to English
Punctuation marks, numbers or white spaces are removed
Tokenize
Stop words are removed
Lemmatization and stemming

Data cleansing

Latent Dirichlet allocation (LDA)

Each tender set of different criteria consists of a mix of topics
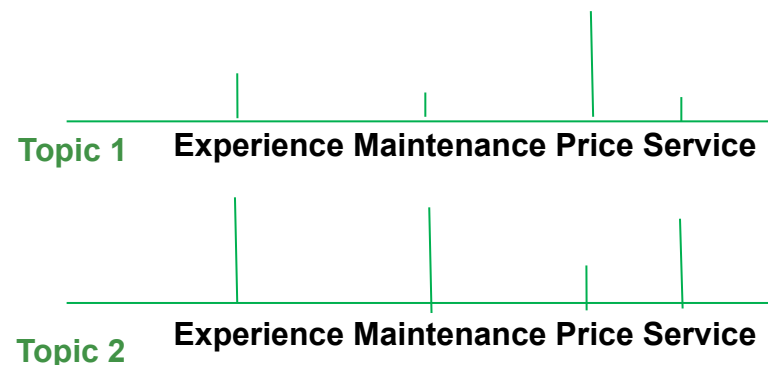

Price

Service Maintenance Experience

Experience Price

100% Topic 1   **100% Topic 2**   **70% Topic 1**
                                          30% Topic 2

Every topic consists of a mix of words

Topic 1   Experience Maintenance Price Service

Topic 2   Experience Maintenance Price Service

How does LDA work?

1. Choose the number of topics you think there are in the corpus

2. Randomly assign each word/criterion in each document to one of the topics

3. Go through every word and its topic assignment in each document. Look at:
   - How often the topic occurs in the document
   - How often the word occurs in the topic overall

4. Based on previous info, assign the word to a new topic.

Goal: LDA **should learn the topic mix** in each tender and **the word mix** in each topic

Latent Dirichlet allocation (LDA)

What is the result of LDA:

1. An optimal number of Topics or groups of award criteria:

Four different Topics

- Topic 1
- Topic 2
- Topic 3
- Topic 4

40,62%
23,04%
23,40%
13,98%
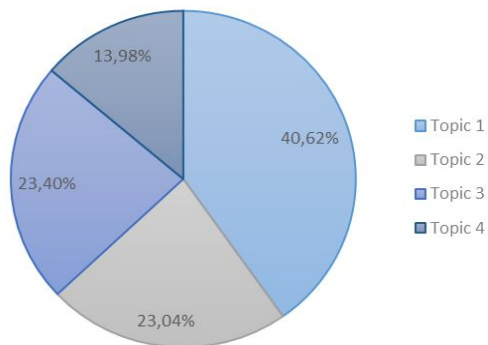
2. A set of distributions of topics for each tender:

Tender ID	Award criteria	Topic1	Topic2	Topic3	Topic4
1	Price, Quality	80,00%	12,50%	5,00%	2,50%
2	Competence and experience of the offered resources	2,00%	5,00%	73,23%	19,77%
...	...	...	...	...	...
72000	Solutions. Price	35,00%	55,00%	2,00%	8,00%

3. A set of distributions of words for each topic

**A topic is defined a vector of words and weights.**

The higher the weight the most importance this word to classify a tender under this topic.

Topic 1 (40.62% of tenders)		Topic 2 (23.04% of tenders)		Topic 3 (22.34% of tenders)		Topic 4 (13.98% of tenders)	
Relevant word	Importance	Relevant word	Importance	Relevant word	Importance	Relevant word	Importance
Technical	32.40%	service	32.52%	Price	63.60%	experience	18.10%
Quality	21.80%	criterion	13.11%	Time	8.12%	contract	12.22%
Offer	15.76%	cost	11.65%	Delivery	6.36%	work	11.31%
Price	9.87%	improvement	7.77%	Period	5.55%	project	10.41%
Proposal	4.12%	maintenance	6.80%	Warranty	5.28%	point	9.95%
Propose	3.68%	system	6.31%	Deadline	3.79%	professional	8.14%
economic	3.24%	specification	6.31%	Guarantee	2.57%	implementation	7.69%
environmental	3.24%	equipment	5.34%	Term	1.89%	month	7.69%
Benefit	3.09%	support	5.34%	Response	1.49%	additional	7.24%
performance	2.80%	management	4.85%	Order	1.35%	number	7.24%

Latent Dirichlet allocation (LDA)

The problem of the most dominant topic:

Topic Number	Keywords	Num. Offers < 3
Topic 1	technical; quality; offer; price; proposal; propose; economic; environmental; benefit; performance	37.33%
Topic 2	service; criterion; cost; improvement; maintenance; system; specification; equipment; support; management	44.07%
Topic 3	price; time; delivery; period; warranty; deadline; guarantee; term; response; order	51.10%
Topic 4	experience; contract; work; project; point; professional; implementation; month; additional; number	42.62%
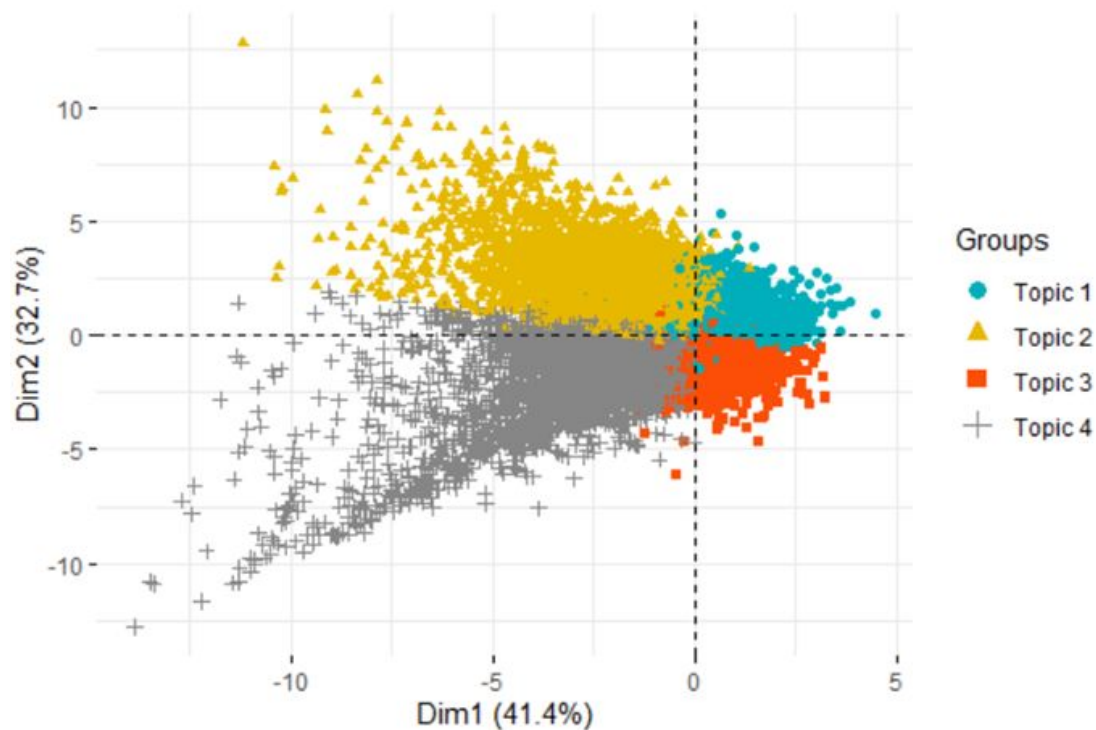
To classify tenders under **the most dominant topic** performs well but **the topic distribution is not exploited**

Self-organizing maps on topic distribution

The problem of the most dominant topic:

Topic 1 distribution   Topic 2 distribution   Topic 3 distribution   Topic 4 distribution
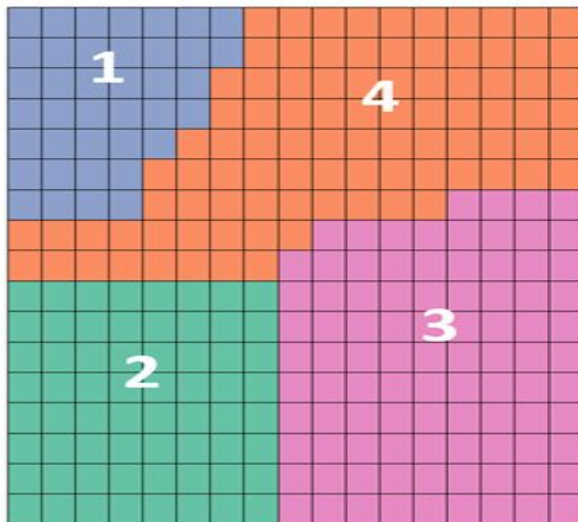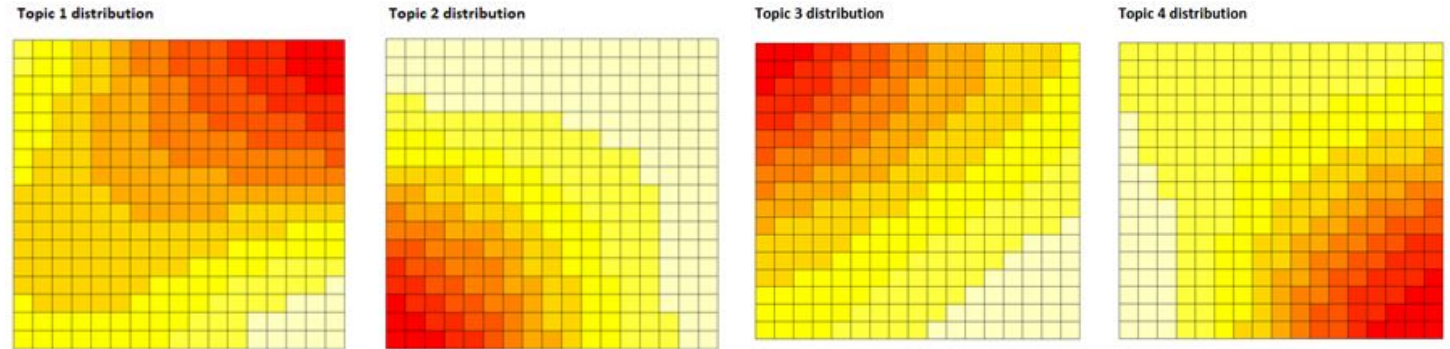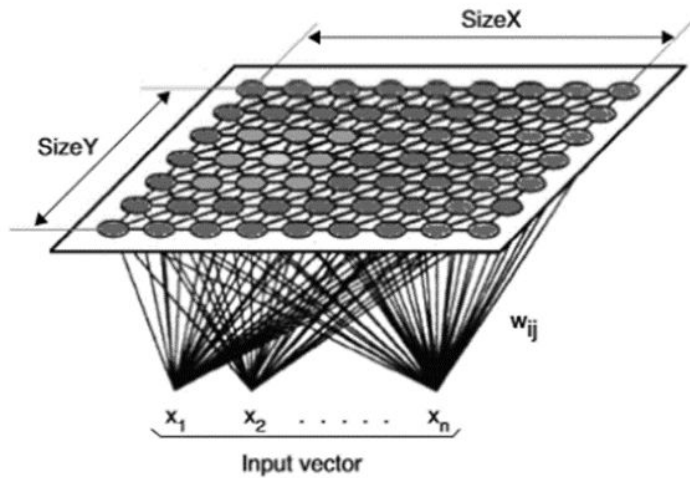
Group	Top 10 criteria	Train sample (2016-2017)		Test sample (2018)	
		Tenders	Red-flag	Tenders	Red-flag
1	price; time; delivery; period; warranty; deadline; guarantee; term; response; order	8,337	54.01%	3,151	56.49%
2	service; criterion; cost; improvement; maintenance; system; specification; equipment; support; management	9,973	45.25%	4,162	46.28%
3	experience; contract; work; project; point; professional; implementation; month; additional; number	11,426	40.52%	4,999	44.05%
4	technical; quality; offer; price; proposal; propose; economic; environmental; benefit; performance	21,645	36.60%	8,773	39.53%

Next steps: Interactive dashboard of Public procurements

A web site has been developed to obtain a visual representation of European public procurements

http://crystaltender.com/
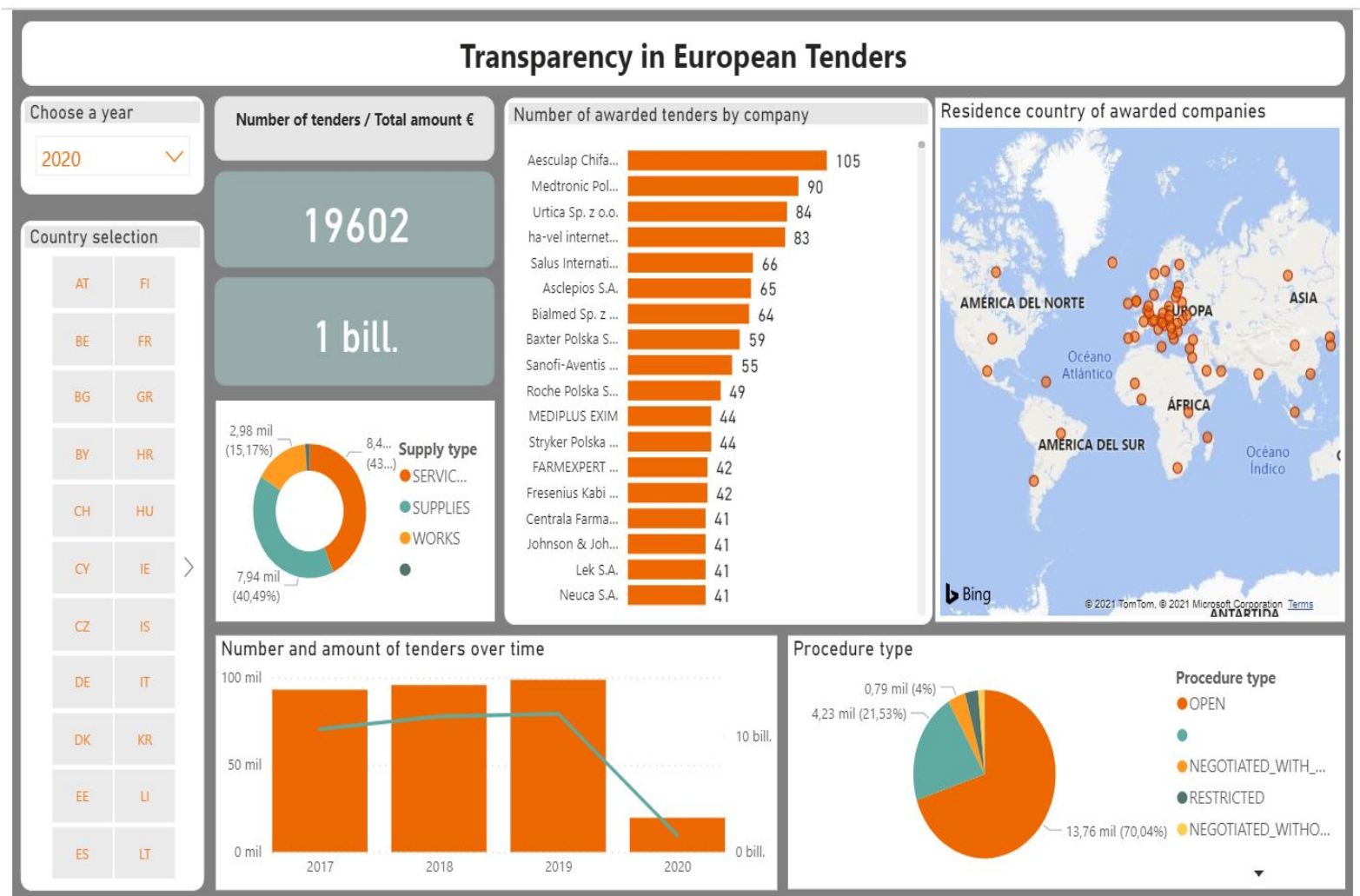
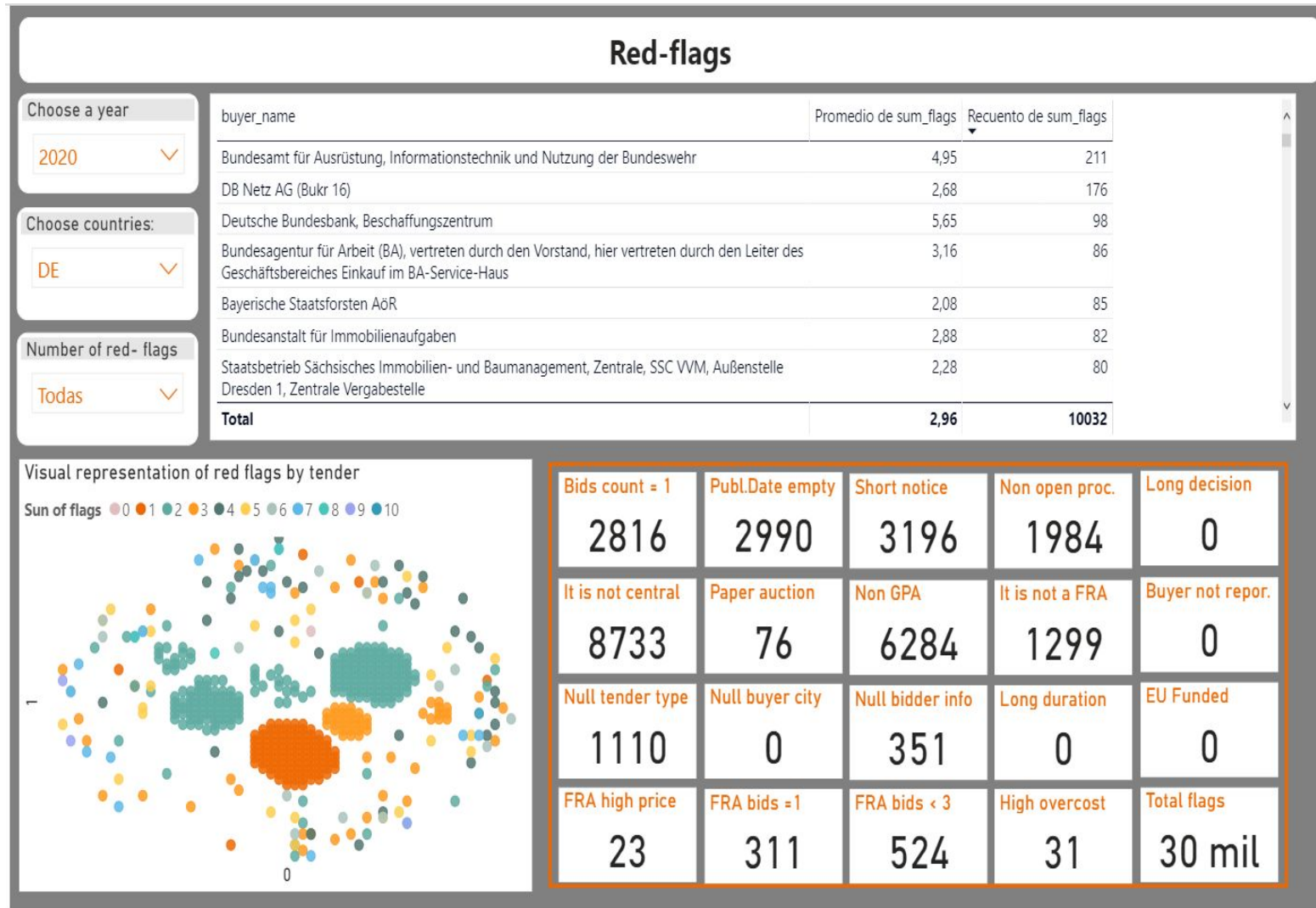**Dashboard of public procurement**

It contains information of public procurements in more than 30 different countries

Information available

- Last five years of information
- Total amount of public procurements
- Detailed list of awarded companies by country
- Details about the procedure and supply types
- Evolution of tenders over time

Interactive dashboard of European Public procurements

It also contains a section where countries and public authorities behavior is assessed:

Red-flags

buyer_name	Promedio de sum_flags	Recuento de sum_flags
Bundesamt für Ausrüstung, Informationstechnik und Nutzung der Bundeswehr	4,95	211
DB Netz AG (Bukr 16)	2,68	176
Deutsche Bundesbank, Beschaffungszentrum	5,65	98
Bundesagentur für Arbeit (BA), vertreten durch den Vorstand, hier vertreten durch den Leiter des Geschäftsbereiches Einkauf im BA-Service-Haus	3,16	86
Bayerische Staatsforsten AöR	2,08	85
Bundesanstalt für Immobilienaufgaben	2,88	82
Staatsbetrieb Sächsisches Immobilien- und Baumanagement, Zentrale, SSC VVM, Außenstelle Dresden 1, Zentrale Vergabestelle	2,28	80
Total	**2,96**	**10032**

Choose a year: 2020
Choose countries: DE
Number of red- flags: Todas

Visual representation of red flags by tender
Sun of flags ● 0 ● 1 ● 2 ● 3 ● 4 ● 5 ● 6 ● 7 ● 8 ● 9 ● 10

Bids count = 1	Publ.Date empty	Short notice	Non open proc.	Long decision
2816	2990	3196	1984	0
It is not central	**Paper auction**	**Non GPA**	**It is not a FRA**	**Buyer not repor.**
8733	76	6284	1299	0
Null tender type	**Null buyer city**	**Null bidder info**	**Long duration**	**EU Funded**
1110	0	351	0	0
FRA high price	**FRA bids =1**	**FRA bids < 3**	**High overcost**	**Total flags**
23	311	524	31	30 mil

Red flags identification

🚩 Most common red-flags are identified by country and buyer name

One example: Germany

- The buyer with the highest number of red flags in 2020 is the Federal Office of Equipment, Information Technology and utilization.
- A total of 211 contracts awarded
- Average number of red flags by tender: 4,95.
- In 190 contracts only one bidder was received

Interactive dashboard of European Public procurements

http://crystaltender.com/

Thank you!