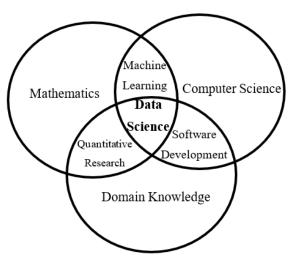


# DATA SCIENCE PROJECT DEVELOPMENT

### Context

 Data Science includes techniques developed in some traditional fields like artificial intelligence, statistics or machine learning.



Aparicio et al. (2019).

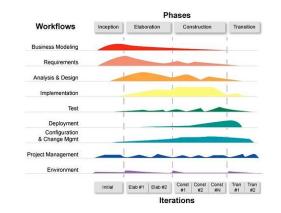








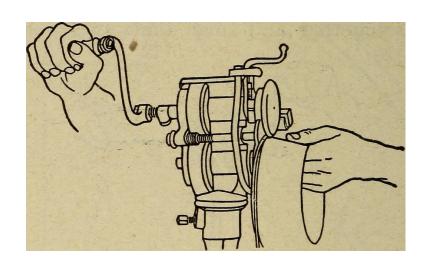






### Context

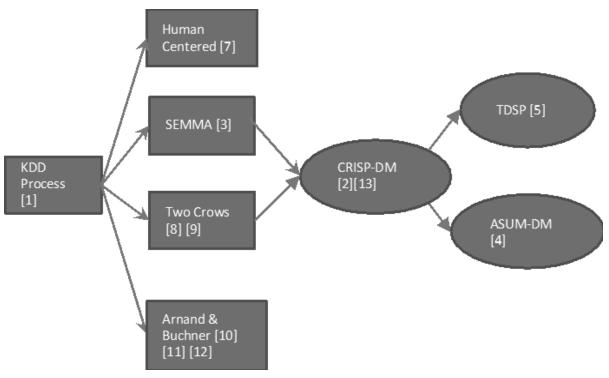
 methodology that may contribute to the improvement of the knowledge creation outputs.



Carlos J. Costa (ISEG) 2023 (4

### Related Work

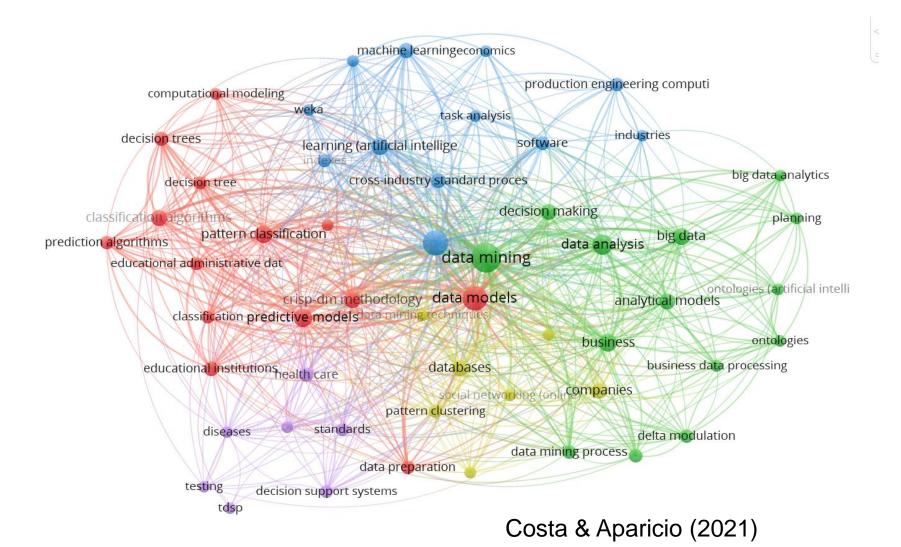
### Process



Costa & Aparicio (2020)

Carlos J. Costa (ISEG) 2023 (5)

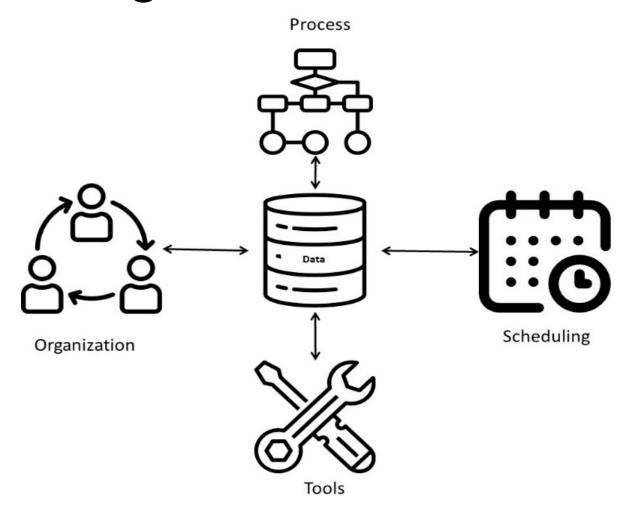
### Related Work



### Related Work

- Summarizing, the approaches related to data mining, machine learning and data science may be interrelated.
- CRISP-DM is one of the most used and the one that inspired many other approaches.
- Nevertheless, other features may be added to this approach:
  - Organization
  - Scheduling
  - Tools

# Proposing a Model



Carlos J. Costa (ISEG) 2023 (8)

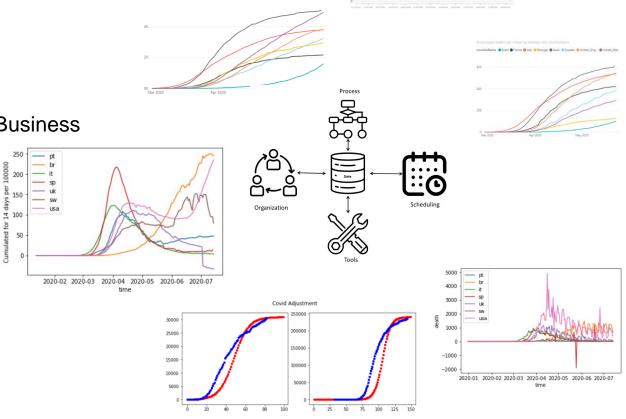
### POST-DS

		ВА	DE	DS	WD	Risk	w1	w2	w3	w4	w5	w6	w7	w8	w9	w10	w11	w12	w13	w14	Tools and Resource
1	Business Understanding																				
1.1.	Define Business Objectives																				
1.2.	Identify ethical values and privacy	A/R				L															meeting
1.3.	Assess Situation	A/R				L															meeting
1.4.	Define Data Science Goals	A/R				L															meeting
1.5.	Produce Project Plan	A/R	R	R		L															WBS, GANTT
2	Data Understanding																				
2.1.	Collect Initial Data		A/R			Н															open data, scraping,
2.2.	Describe Data		A/R			L															use Jupyter/python/Pandas
2.3.	Explore Data		A/R			М															use Jupyter/python/Pandas
2.4.	Verify Data Quality			A/R		Н															use Jupyter/python/Pandas
3	Data Preparation			A/R																	
3.1.	Select Data			A/R		М															Meeting
3.2.	Clean Data			A/R		М															use Jupyter/python/Pandas
3.3.	Construct Data			A/R		М															use Jupyter/python/Pandas
3.4.	Integrate Data			A/R		Н															use Jupyter/python/Pandas
3.4.	Format Data			A/R		Н															use Jupyter/python/Pandas
4	Modeling																				
4.1.	Select Modeling Techniques	_		A/R		Н															MIT flowchart
4.2.	Generate Test Design	L		A/R		Н															use Jupyter/python/Pandas
4.3.	Build Model	L		A/R		М															use Jupyter/python/Pandas
4.4.	Assess Model	I		A/R		Н															use Jupyter/python/Pandas
5	Evaluation																				
5.1.	Evaluate Results, icnluding ethical	A/R		R		Н															use Jupyter/python/Pandas
5.2.	Review Process	A/R				L															meeting
5.3.	Determine Next Steps	A/R				L															meeting
6	Deployment																				
	Plan Deployment	Α		R	R	Н															PowerBI or Flash
6.2.	Plan Monitoring and Maintenance	Α				М															meeting
6.3.	Produce Final Report	A/R	R	R	R	М															PowerBI or Flash
6.4.	Review Project	A/R		R		М															meeting

Carlos J. Costa (ISEG) 2023 (9)

# Using the model

- Covid
- Financial market
- Real estate
- Politics
- Software Development Business
- Academic context
- Professional work



Carlos J. Costa (ISEG) 2023 (10)

# Using the model

#### Sentiment Analysis of Portuguese Political Parties Communication

Carlos J. Costa ADVANCE/CSG, ISEG (Lisbon School of Economics & Management), Universidade de Lisboa, Portugal cjcosta@iseg.ulisboa.pt Manuela Aparicio NOVA Information Management School (NOVA IMS), Universidade Nova de Lisboa, Portugal manuela.aparicio@novaims.unl.pt Joao Tiago Aparicio INESC-ID and Instituto Superior Tecnico/Universidade de Lisboa, Dep Transportes LNEC, Portugal joao.aparicio@tecnico.ulisboa.pt

#### ABSTRACT

Political communication in social media has gained increasing importance in the last years. In this study, we analyze the political parties' communication on Twitter and understand the sentiment of their communication. First by identifying their communication performance regarding the daily number of tweets, favorite tweets. number of retweets per day and per political party. We present a sentiment analysis by the political party using tweets data. In this study, we propose an explanatory model with the main drivers of retweets. To conduct this study, our approach used data analysis and machine learning techniques methods. Results indicate the main determinants that influence future retweets of political posts globally. Here we present a comparison of the communication content between tweets posts and the political parties' programs available on their institutional websites. We identify the similarities between tweets and formal programs per party and among all parties. This study contributes to analyze the coherence and effectiveness of the political parties' communication.

#### CCS CONCEPTS

General and reference; - Cross-computing tools and techniques; - Empirical studies;; - Information systems; - Information retrieval; - Retrieval tasks and goals; - Sentiment analysis; - Computing methodologies; - Artificial intelligence; - Natural language processing;; - Machine learning;

#### KEYWORDS

Twitter, political parties, sentiment analysis, document similarity, machine learning

#### ACM Reference Format:

Carlos J. Costa, Manucka Aparicio, and Joao Tiago Aparicio. 2021. Sentiment Analysis of Portuguese Political Parties Communication. In The 39th ACM International Conference on Design of Communication (SIGDOC '21), October 12–14, 2021, Virtual Event, USA. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3472714.34736.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profil or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be homored. Abstracting with orecett is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions(abean.org).

SIGDOC 21, October 12–14, 2021, Virtual Event, USA © 2021 Association for Computing Machinery. ACM ISBN 978-1-4503-8628-9/21/10...\$15.00 https://doi.org/10.1145/3/27214.3473624

#### 1 INTRODUCTION

Political parties are increasingly using social media to communicate their values and ideas. An example of such behavior was when Barack Obama's staff successfully used Twitter in the 2008 presidential elections [13]. This practice persisted in Donald Trump's administration in 2016 [14]. However, using social media is not a panacea per se. It is also essential to analyze what is the effectiveness of what is being communicated to the public. The cohesion between parties' political agenda versus their social media communication needs to be addressed [27]. Are parties communicating according to their goals, or is it all part of a hype machine? [16] The usage of social media in a political context has been studied by several authors [6, 7, 13, 14]. In this context, the evolution in natural language processing and sentiment analysis is significant. Nevertheless, there is a research gap in the Portuguese language and its application in a political context as well [15]. The purpose of the work performed in this paper is to analyze political parties' communication, expressed explicitly by the official Twitter accounts of such parties. To reach this main goal, we state four research objectives (RO) as follows: RO1: Identify the performance of each political party on Twitter; RO2: Identify the global sentiment per political party in Twitter communication; RO3: Identify the drivers of retweet behavior in political parties: RO4: Understand the similarities between social media communication and political program communication.

We used data analysis and machine learning techniques described in section 3 of this paper for each of these research objectives. This study contributes to a better understanding of how political parties communication in Twitter can be analyzed in terms of the sentiment of posts, and globally we contribute to analyze the coherence and effectiveness of the political parties communication. We also used also conducted our approach following a data science perspective [22, 23].

#### 2 LITERATURE REVIEW

Social media analysis has been a study in the context of design of communication [22, 24, 25]. In this context, many approaches were used [22, 24]. Social media usage has been studied to identify the performance of each political party in Twitter [17]. Twitter is also a relevant data source to identify the global sentiment per political party in Twitter communication [18]. Sentiment analysis refers to using several approaches, like natural language processing, text analysis, computational linguistics, and biometrics, to systematically identify, extract, quantify, and study affective states and subjective information. Emotions can be reactions to internal stimul (such as thoughts or memories) or events in our environment.

J. T. Aparicio, J. Salema de Sequeira and C. J. Costa, "Emotion analysis of Portuguese Political Parties Communication over the covid-19 Pandemic," 2021 16th Iberian Conference on Information Systems and Technologies (CISTI), 2021, pp. 1-6, doi: 10.23919/CISTIS.2073.2021.9476557.

#### Emotion analysis of Portuguese Political Parties Communication over the covid-19 Pandemic

Joao Tiago Aparicio Instituto Superior Técnico, University of Lisbon Lisbon, Portugal joao.aparicio@tecnico.ulisboa.pt

João Salema de Sequeira Instituto de Estudos Políticos Universidade Católica Portuguesa joaosalemasequeira@gmail.com Carlos J. Costa
Instituto Universitário de Lisboa
(ISCTE-IUL), ISTAR-IUL,
Lisboa, Portugal
ISEG, Universidade de Lisboa,
Lisboa, Portugal
carlos.costa@acm.org

Abstract - In this paper, we explore the use of emotions in the Portuguese political parties' (with a seat in the Portuguese Parliament) communication as expressed by their official Twitter accounts, as of March 2020. The chosen period of our investigation is particularly interesting because political parties had a chance to communicate their views during a pandemic situation and over a period of one year. These views include possible solutions to face the crisis and their comments on the development of the whole situation. Using a standard lexicon we classified the amount of particular emotions in different tweets. Using this method we plotted the average positivity and negativity along time per party. We also analyzed the impact of each emotion to classify positivity using the present corpus. Finally, we considered some important words regarding the pandemic and their average positivity score. The analysis allows us to identify different approaches to participation in social media according to different strategies, more than political ideology.

Keywords - political communication; Portuguese political parties; Portuguese parliament; Portuguese; lexicon; sentiment analysis; emotions; visualization; social media; twitter; covid-19.

#### INTRODUCTION

Now-a-days, different political actors are increasingly using social media platforms to communicate their worldviews. American Presidents have used Twitter heavily to communicate their position in relation to specific ideas and to specific policies [5]. Hence it is essential to analyse what is being communicated and even more important how this communication is being done in order to best assess their impact. Political communication can help us explain the ups and downs of the electoral polls and the electoral success of a certain political party or individual in the following election.

The publication of The Gutenberg Galaxy: The Making of Typographic Man [8] considers the effects of social media in different human dimensions. However, a new empirical approach is needed, one that considers the effects of social networks or to put it simply a Zuckerberg Galaxy approach

which demonstrates how Facebook, Twitter, and other social media are used and to what extent they have a more decisive influence on some of the voters, in comparison to the traditional media. In this context, the evolution in Natural Language Processing (NLP) and sentiment analysis is significant, however the political communication in Portugal has not yet been a subject of this kind of study, since the available models and lexicons are not yet adapted to European Portuguese. In this sense, we aim to answer the following question: What are the prevalent emotions in the Portuguese political parties' tweets during over the first year of the covid-19 pandemic;

The purpose of the work performed in this paper is to analyse the communication of the official Twitter accounts of the Portuguese political parties. The time frame ranges through 3200 last tweets, going as back as March 2020, when the first case of covid-19 was registered in the country. This period is specially interesting because political parties had a chance to communicate to the electorate their ideas in face of a social and economic crisis. It is important to take into consideration that the different parties tweeted with a different frequency, however the reality they were facing was one and the same.

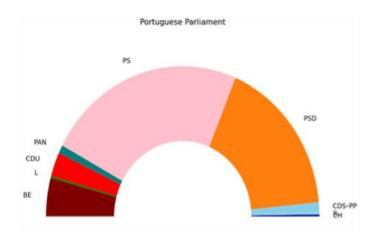
#### II. LITERATURE REVIEW

Sentiment analysis refers to using several approaches, such as: natural language processing, text analysis, computational linguistics, and biometrics, to systematically identify, extract, quantify, and study affective states and subjective information.

Emotions can be reactions to internal stimuli (such as thoughts or memories) or events in our environment. To analyze emotions, Mohammad and Turney [1] proposed a lexicon. This lexicon uses six emotions [2], [3]: joy, sadness, anger, fear, disgust, and surprise, along with how positive and negative the words are. These are a subset of the eight emotions proposed in Plutchik [4] which are still relevant today [10]. Recently the study of the impact of texts on such emotions has been done, namely in the USA political context [5]. This was done with a focus on awareness and topical emergence. However, there was no analysis over the emotion on the content of the message shared by the political parties, instead it was

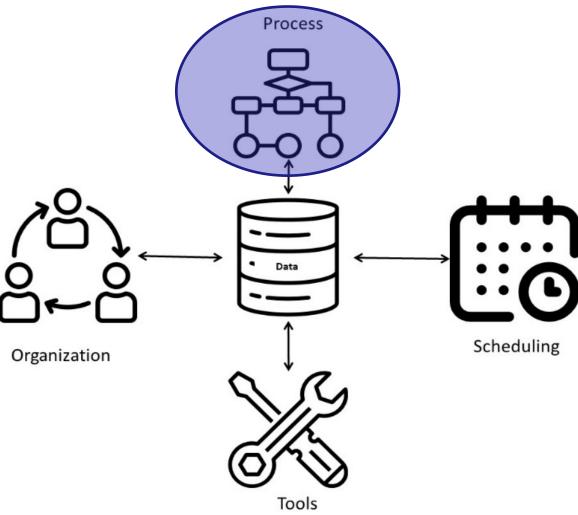
# Using the model

 Emotion analysis of Portuguese Political Parties Communication

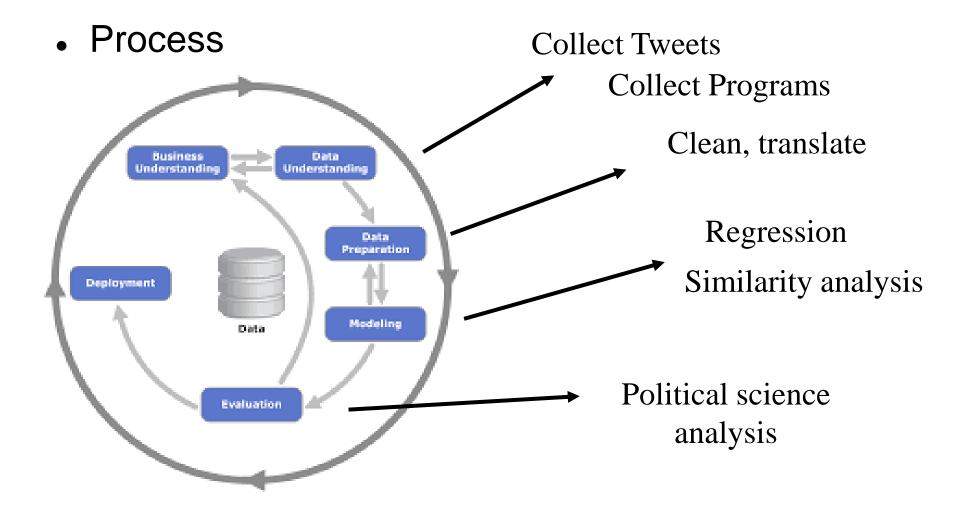




**Carlos J. Costa (ISEG)** 2023 (12)



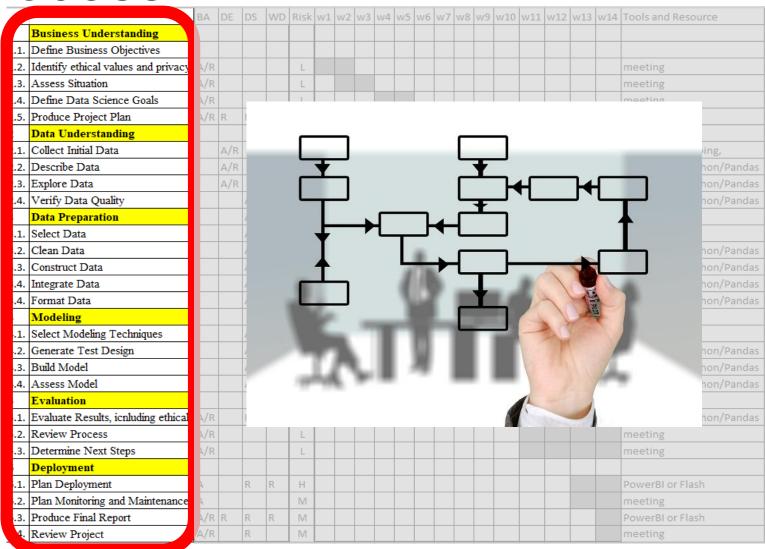
**Carlos J. Costa (ISEG)** 2023 (13)

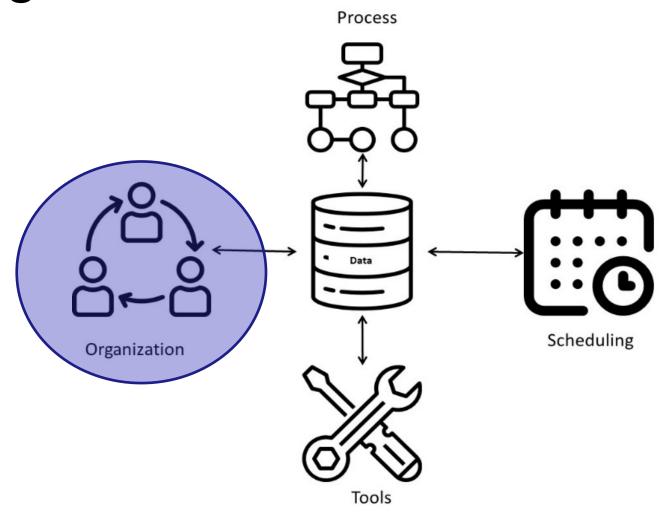


**Carlos J. Costa (ISEG)** 2023 (14)

		BA	D	Е	DS	WD	Risk	w1	w2	w3	w4	w5	w6	w7	w8	w9	w10	w11	w12	w13	w14	Tools and Resource
	Business Understanding																					
.1.	Define Business Objectives																					
.2.	Identify ethical values and privacy	۱,/	R				L															meeting
.3.	Assess Situation	۱,/	'R				L															meeting
.4.	Define Data Science Goals	۱,/	R				L															meeting
5.	Produce Project Plan	۱,/	R R		R		L															WBS, GANTT
	Data Understanding																					
.1.	Collect Initial Data		А	/R			Н															open data, scraping,
.2.	Describe Data		А	/R			L															use Jupyter/python/Pandas
.3.	Explore Data	L	А	/R			M															use Jupyter/python/Pandas
.4.	Verify Data Quality	L		1	A/R		Н															use Jupyter/python/Pandas
	Data Preparation	L			A/R																	
.1.	Select Data	L			A/R		M															Meeting
.2.	Clean Data	L			A/R		M															use Jupyter/python/Pandas
.3.	Construct Data	L		1	A/R		M															use Jupyter/python/Pandas
.4.	Integrate Data	L			A/R		Н															use Jupyter/python/Pandas
.4.	Format Data	L		1	A/R		Н															use Jupyter/python/Pandas
Ŀ	Modeling	L																				
_	Select Modeling Techniques	L		1	A/R		Н															MIT flowchart
_	Generate Test Design	L		ı	A/R		Н															use Jupyter/python/Pandas
_	Build Model	L		ı	A/R		M															use Jupyter/python/Pandas
.4.	Assess Model	L			A/R		Н															use Jupyter/python/Pandas
	Evaluation	L																				
_	Evaluate Results, icnluding ethical	۱,/	R		R		Н															use Jupyter/python/Pandas
	Review Process	-	R				L															meeting
.3.	Determine Next Steps	۱,/	R				L															meeting
	Deployment	L																				
_	Plan Deployment	A			R	R	Н															PowerBI or Flash
_	Plan Monitoring and Maintenance	A					M															meeting
_	Produce Final Report	_	R R	-	R	R	M															PowerBI or Flash
4.	Review Project	A/	R		R		M															meeting

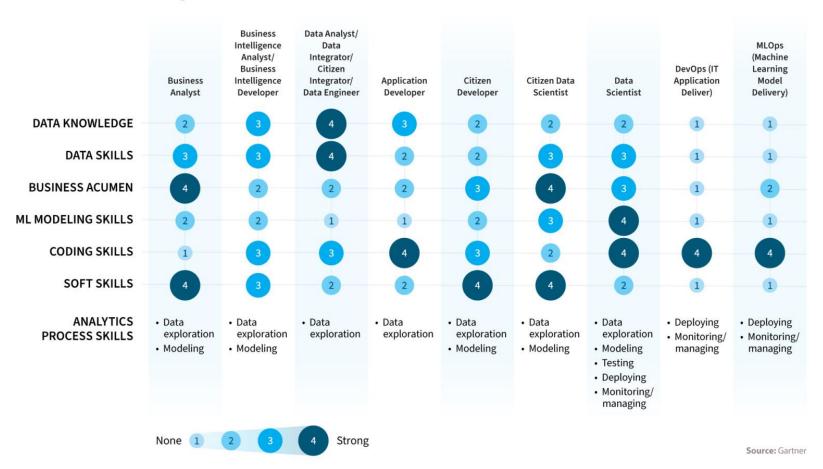
**Carlos J. Costa (ISEG)** 2023 (15)





**Carlos J. Costa (ISEG)** 2023 (17)

#### **Continuum of Analytics Roles and Skills**



Carlos J. Costa (ISEG) 2023 (18)

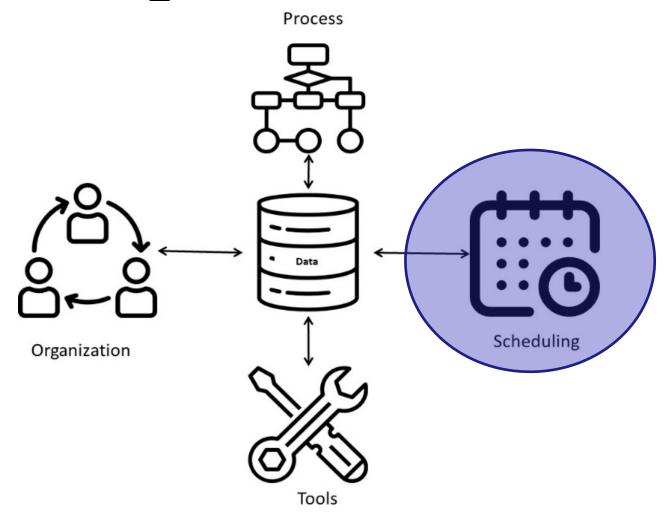
- Process
- Organization
- Scheduling
- Tools



**Carlos J. Costa (ISEG)** 2023 (19)

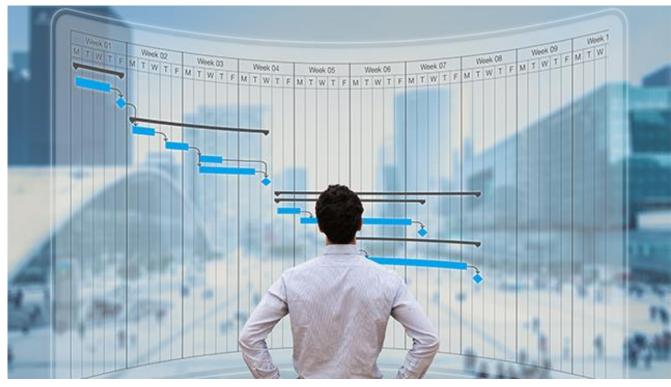
_																					i
		A	DE	DS	WL	Ris	( W	1 w2	w3	w4	w5	w6	w7	w8	w9	w10	w11	w12	w13	w14	Tools and Resource
1	Business Understanding			_		L	$\perp$														
	Define Business Objectives						L														
1.2.	Identify ethical values and privacy	/R				L															meeting
1.3.	Assess Situation	/R				L	L														meeting
1.4.	Define Data Science Goals	/R	_			L															meeting
1.5.	Produce Project Plan	/R	R	R		L															WBS, GANTT
2	Data Understanding																				
2.1.	Collect Initial Data		A/R			Н															open data, scraping,
2.2.	Describe Data		A/R			L															use Jupyter/python/Pandas
2.3.	Explore Data		A/R			M															use Jupyter/python/Pandas
2.4.	Verify Data Quality			A/R		Н															use Jupyter/python/Pandas
3	Data Preparation			A/R																	
3.1.	Select Data			A/R		M															Meeting
3.2.	Clean Data			A/R		M															use Jupyter/python/Pandas
3.3.	Construct Data			A/R		M															use Jupyter/python/Pandas
3.4.	Integrate Data			A/R		Н															use Jupyter/python/Pandas
3.4.	Format Data			A/R		Н															use Jupyter/python/Pandas
4	Modeling																				
4.1.	Select Modeling Techniques			A/R		Н															MIT flowchart
4.2.	Generate Test Design			A/R		Н															use Jupyter/python/Pandas
4.3.	Build Model			A/R		M															use Jupyter/python/Pandas
4.4.	Assess Model			A/R		Н															use Jupyter/python/Pandas
5	Evaluation																				
5.1.	Evaluate Results, icnluding ethical	/R		R		Н															use Jupyter/python/Pandas
5.2.	Review Process	/R				L															meeting
5.3.	Determine Next Steps	/R				L															meeting
6	Deployment																				
6.1.	Plan Deployment			R	R	Н															PowerBI or Flash
6.2.	Plan Monitoring and Maintenance					M															meeting
6.3.	Produce Final Report	/R		R	R	M															PowerBI or Flash
6.4.	Review Project	VR		R		М															meeting

**Carlos J. Costa (ISEG)** 2023 (20)



**Carlos J. Costa (ISEG)** 2023 (21)

- Process
- Organization
- Scheduling
- Tools



Carlos J. Costa (ISEG) 2023 (22)

Process

Organization

Scheduling

• Tools

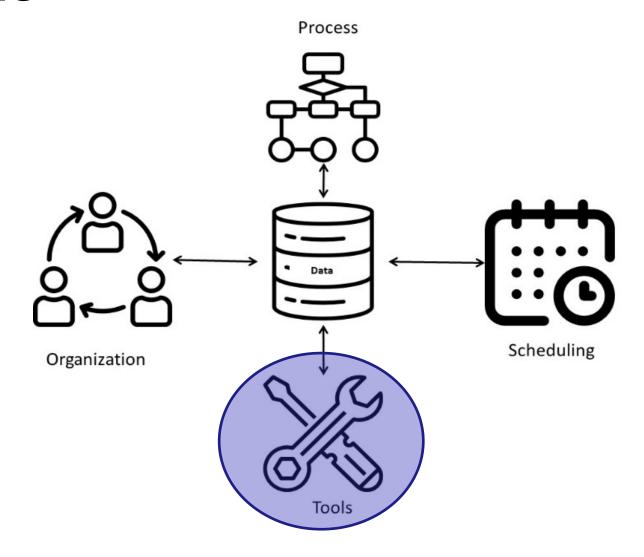


**Carlos J. Costa (ISEG)** 2023 (23)

		ВА	DE	DS	WD	Risk	k y	 W3	W4	W5	Wб	w/	w8	W9	WIO	WII	W12	W15	14	Tools and Resource
1 B	Business Understanding																			
1.1. D	Define Business Objectives																			
1.2. Id	dentify ethical values and privacy	A/R																		neeting
1.3. A	Assess Situation	A/R																		neeting
1.4. D	Define Data Science Goals	A/R																		neeting
1.5. P	Produce Project Plan	A/R	R	R																/BS, GANTT
2 <b>D</b>	Oata Understanding																			
2.1. C	Collect Initial Data		A/R																	pen data, scraping,
2.2. D	Describe Data		A/R																	se Jupyter/python/Pandas
2.3. E	Explore Data		A/R			r														se Jupyter/python/Pandas
2.4. V	Verify Data Quality			A/R																se Jupyter/python/Pandas
3 <b>D</b>	Data Preparation			A/R																
3.1. S	elect Data			A/R		r														leeting
3.2. C	llean Data			A/R		ı														se Jupyter/python/Pandas
3.3. C	Construct Data			A/R		ı														se Jupyter/python/Pandas
3.4. Ir	ntegrate Data			A/R																se Jupyter/python/Pandas
3.4. F	ormat Data			A/R		H														se Jupyter/python/Pandas
4 1	Modeling																			
4.1. S	elect Modeling Techniques	1		A/R		H														IIT flowchart
4.2. G	Generate Test Design	I		A/R		1														se Jupyter/python/Pandas
4.3. B	Build Model	L		A/R		ľ														se Jupyter/python/Pandas
4.4. A	Assess Model	I		A/R		H														se Jupyter/python/Pandas
5 <b>E</b>	Evaluation																			
5.1. E	Evaluate Results, icnluding ethical	A/R		R		H														se Jupyter/python/Pandas
5.2. R	Review Process	A/R																		neeting
5.3. D	Determine Next Steps	A/R																		neeting
6 <b>D</b>	Deployment																			
6.1. P	lan Deployment	А		R	R	H														owerBI or Flash
6.2. P	lan Monitoring and Maintenance	А				ľ														neeting
6.3. P	roduce Final Report	A/R	R	R	R	M														owerBI or Flash
6.4. R	Review Project	A/R		R		M	N													meeting

**Carlos J. Costa (ISEG)** 2023 (24)

### **Tools**



**Carlos J. Costa (ISEG)** 2023 (25)

### **Tools**

		ВА	DE	DS	WD	Risk	w1	w2	w3	w4	w5	w6	w7	w8	w9	w10	w11	w12	w13	w14	ools and Resource
1	Business Understanding																				
1.1.	Define Business Objectives																				
1.2.	Identify ethical values and privacy	A/R				L															meeting
1.3.	Assess Situation	A/R				L															meeting
1.4.	Define Data Science Goals	A/R				L															meeting
1.5.	Produce Project Plan	A/R	R	R		L															WBS, GANTT
2	Data Understanding																				
2.1.	Collect Initial Data		A/R			Н															open data, scraping,
2.2.	Describe Data		A/R			L															use Jupyter/python/Pandas
2.3.	Explore Data		A/R			M															use Jupyter/python/Pandas
2.4.	Verify Data Quality			A/R		Н															use Jupyter/python/Pandas
3	Data Preparation			A/R																	
3.1.	Select Data			A/R		M															Meeting
3.2.	Clean Data			A/R		M															use Jupyter/python/Pandas
3.3.	Construct Data			A/R		M															use Jupyter/python/Pandas
3.4.	Integrate Data			A/R		Н															use Jupyter/python/Pandas
3.4.	Format Data			A/R		Н															use Jupyter/python/Pandas
4	Modeling																				
4.1.	Select Modeling Techniques	1		A/R		Н															MIT flowchart
4.2.	Generate Test Design	I		A/R		Н															use Jupyter/python/Pandas
4.3.	Build Model	I		A/R		M															use Jupyter/python/Pandas
4.4.	Assess Model	I		A/R		Н															use Jupyter/python/Pandas
5	Evaluation																				
5.1.	Evaluate Results, icnluding ethical	A/R		R		Н															use Jupyter/python/Pandas
5.2.	Review Process	A/R				L															meeting
5.3.	Determine Next Steps	A/R				L															meeting
6	Deployment																				
	Plan Deployment	Α		R	R	Н															PowerBI or Flash
6.2.	Plan Monitoring and Maintenance	А				M															meeting
6.3.	Produce Final Report	A/R	R	R	R	M															PowerBI or Flash
6.4.	Review Project	A/R		R		M															reeting

**Carlos J. Costa (ISEG)** 2023 (26)

### Tools

- Process
- Organization

Home > Conferences > DOC > Proceedings > SIGDOC '21 > Sentiment Analysis of Portuguese Political Parties

Authors: <a>Carlos Costa</a>, <a>Manuela Aparicio</a>,

SIGDOC '21: The 30th ACM International Conference on

Design of Communication • October 2021 • Pages

Ioao Aparicio Authors Info & Claims

RESEARCH-ARTICLE FREE ACCESS Sentiment Analysis of Portuguese Political Parties

Communication

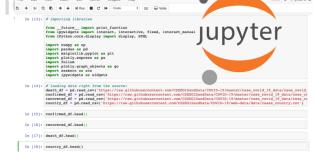






• Schedulin Field New York

Tools







Test Data

### Conclusions

- Adequate Approach
- Many roles and people with different backgrounds
- Improve organization contribution
- Improve scheduling
- Allows results vs. expectations adjustment
- Main limitation: Bureaucracy

### References

- Aparicio, J.T, Salema de Sequeira, J & Costa, J. (2021) Emotion analysis of Portuguese Political Parties Communication over the covid-19 Pandemic in 2021 16th Iberian Conference on Information Systems and Technologies (CISTI), pp. 1-6, doi: 10.23919/CISTI52073.2021.9476557.
- Aparicio, S., Aparicio, J. T., & Costa, C. J. (2019). Data Science and AI: trends analysis. In 2019
  14th Iberian Conference on Information Systems and Technologies (CISTI) (pp. 1-6). IEEE.
  DOI:10.23919/CISTI.2019.8760820
- Costa, C. J. & Aparicio, J.T. (2020). POST-DS: A Methodology to Boost Data Science. In 2020 15th Iberian Conference on Information Systems and Technologies (CISTI) (pp. 1-6). IEEE. Doi:10.23919/CISTI49556.2020.9140932
- Costa, C, Aparicio, M & Aparicio, J.T. (2021). Sentiment Analysis of Portuguese Political Parties Communication. In The 39th ACM International Conference on Design of Communication (SIGDOC '21). Association for Computing Machinery, New York, NY, USA, 63–69.
   DOI:10.1145/3472714.34736241.
- Costa CJ, Aparicio JT.(2021) A Methodology to Boost Data Science in the Context of COVID-19.
   Advances in Parallel & Distributed Processing, and Applications. Published online 2021:65-75.
   doi:10.1007/978-3-030-69984-0

Carlos J. Costa (ISEG) 2023 (29)



Carlos J. Costa (ISEG) 2023 (30)