



Lisbon School
of Economics
& Management
Universidade de Lisboa

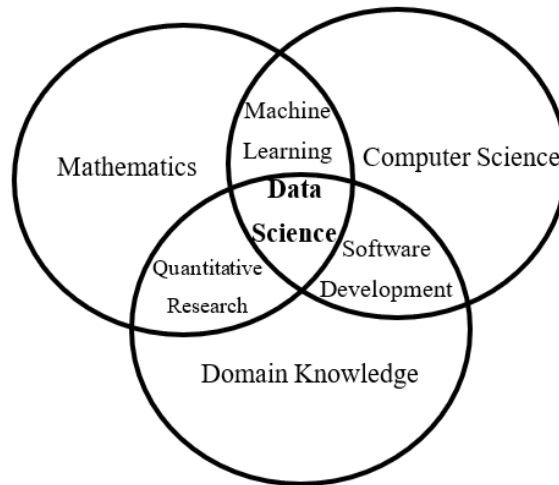


DATA SCIENCE PROJECT DEVELOPMENT

Carlos J. Costa, ISEG (2025)

Context

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



Aparicio et al.(2019).

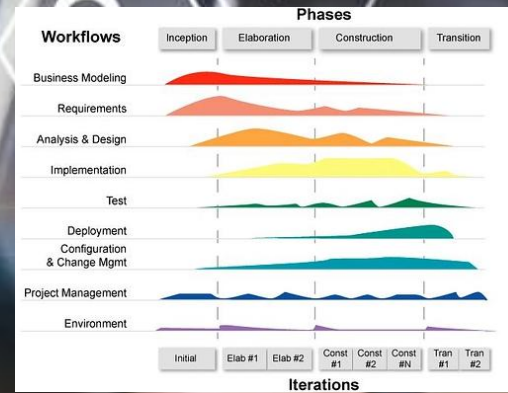
PROJECT MANAGEMENT



IPMA[®]



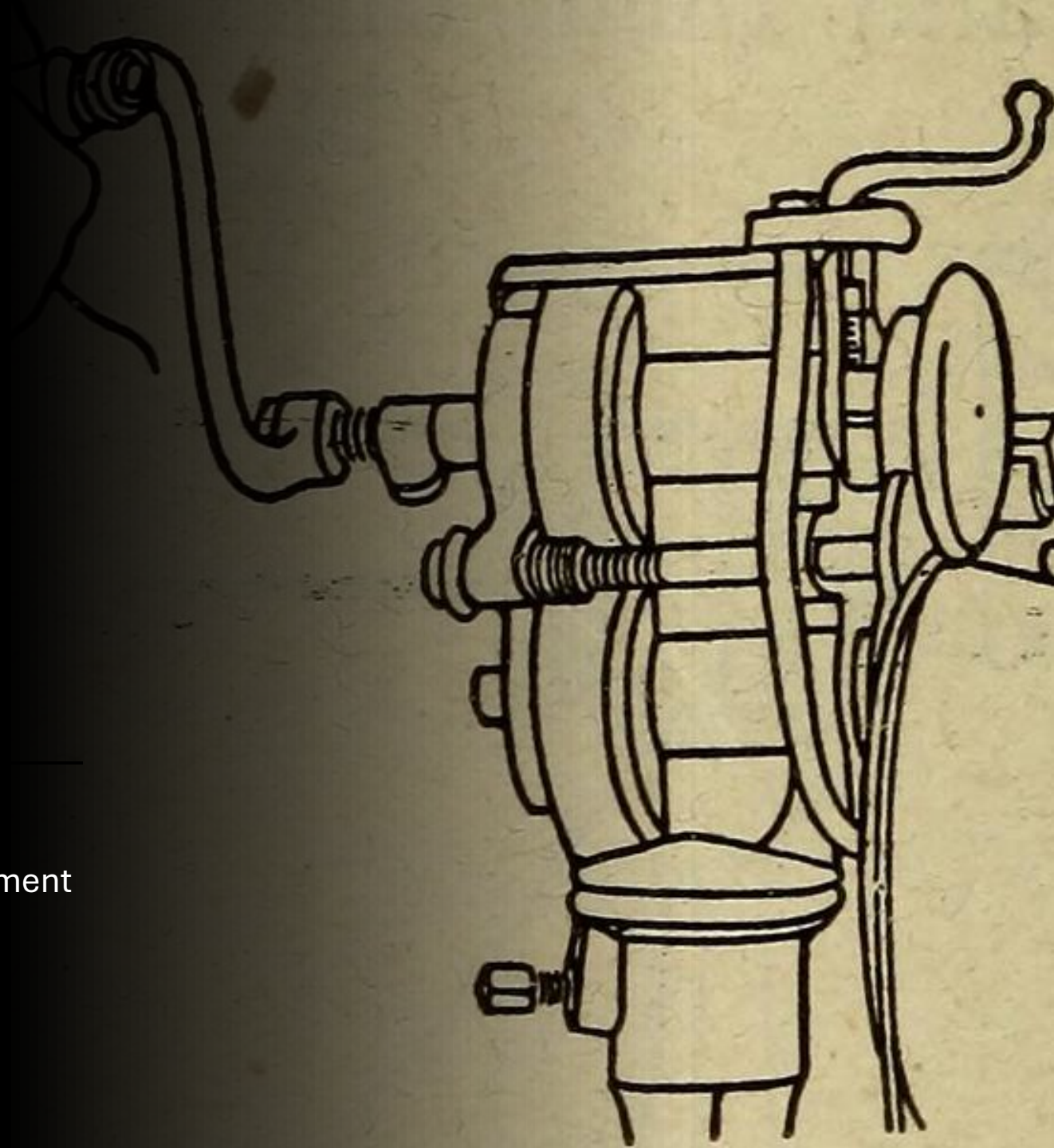
WATER FALL





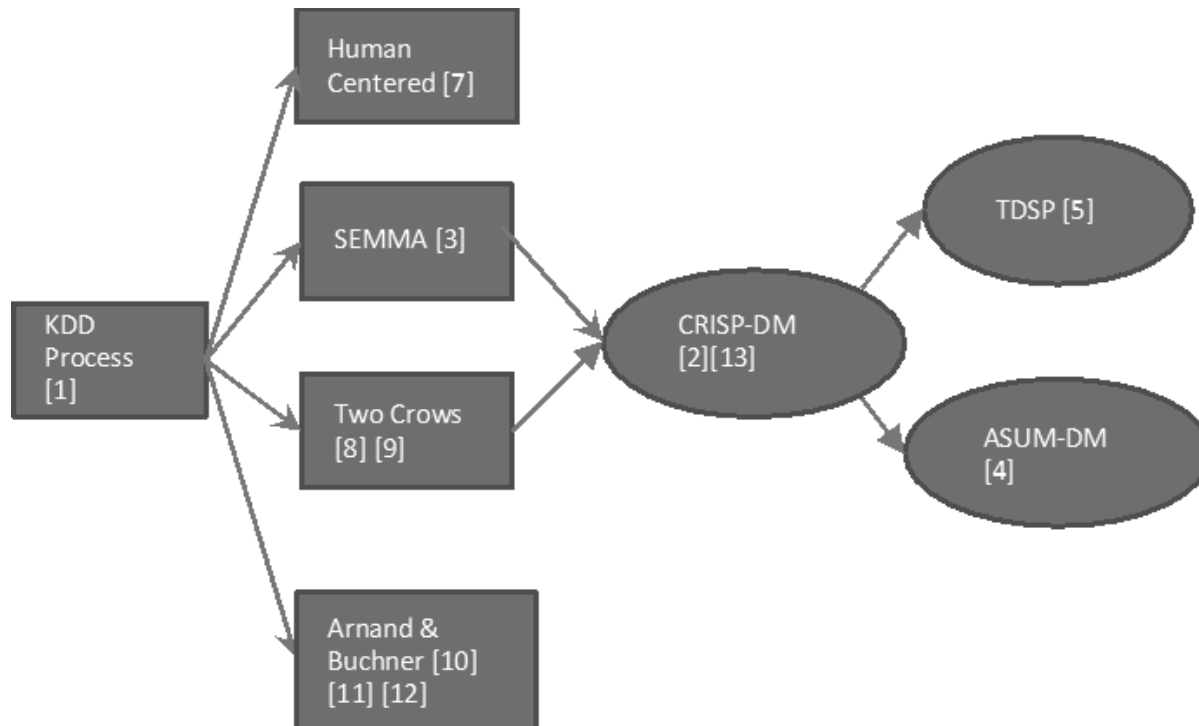
Context

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

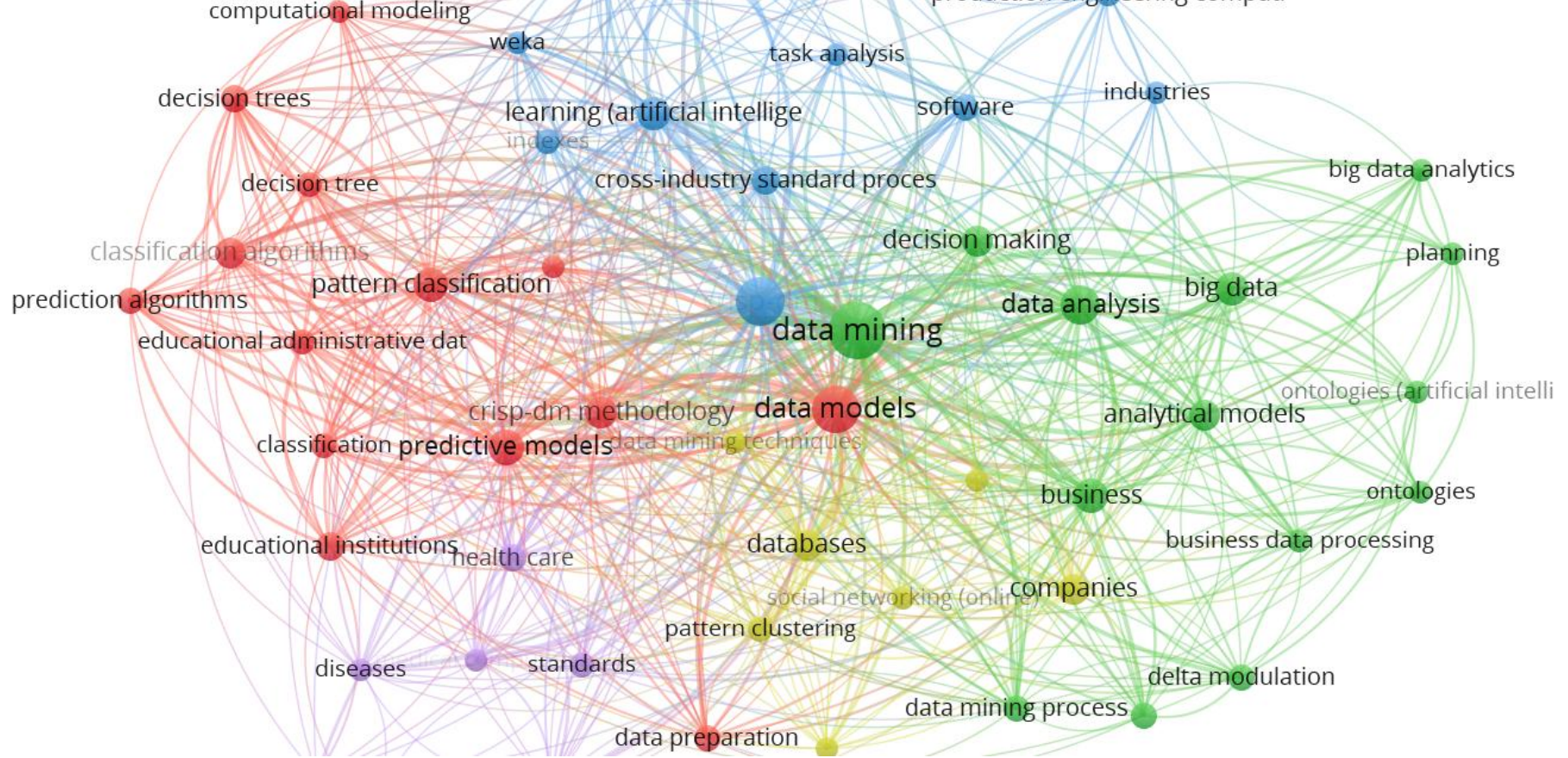


Related Work

- Process



Costa & Aparicio (2020)



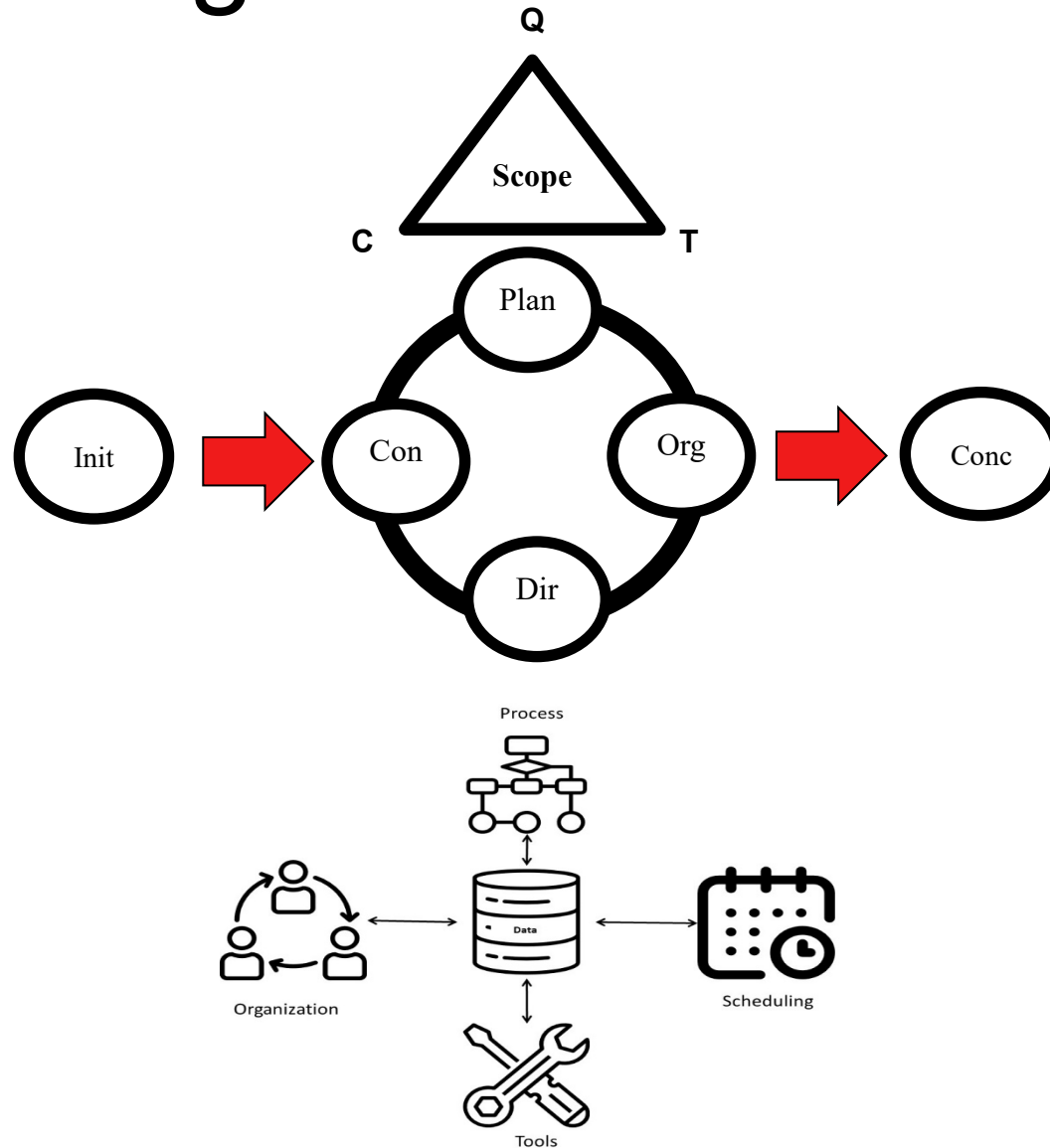
Related Work

Costa & Aparicio (2021)

Related Work

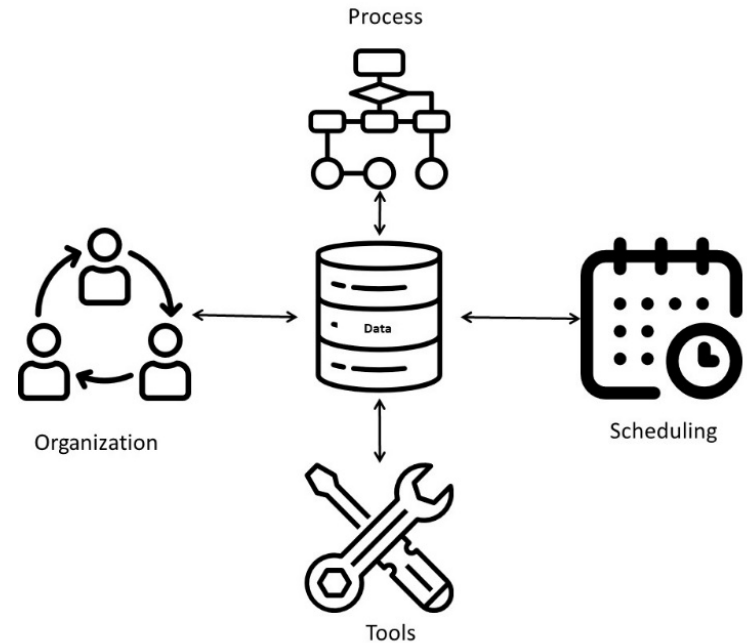
- Summarizing, the approaches related to data mining, machine learning and data science may be interrelated.
- CRISP-DM is one of the most used and the one that inspired many other approaches.
- Nevertheless, other features may be added to this approach:
 - Organization
 - Scheduling
 - Tools

Proposing a Model

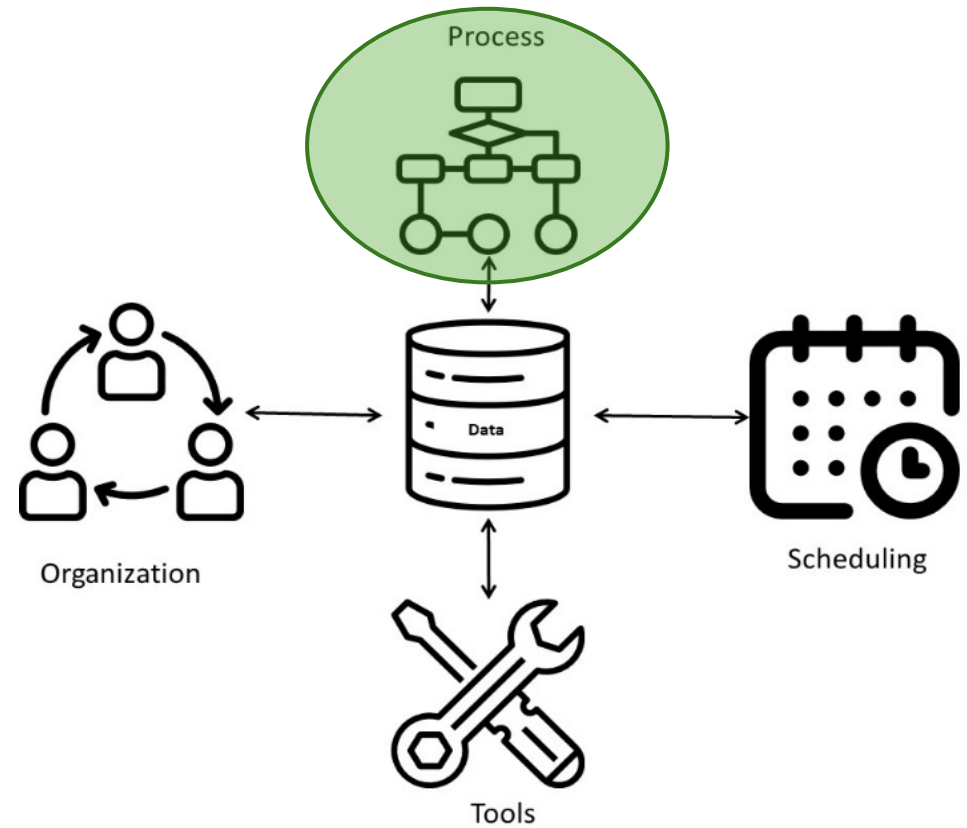


Proposing a Model

- POST-DS



Process



Process

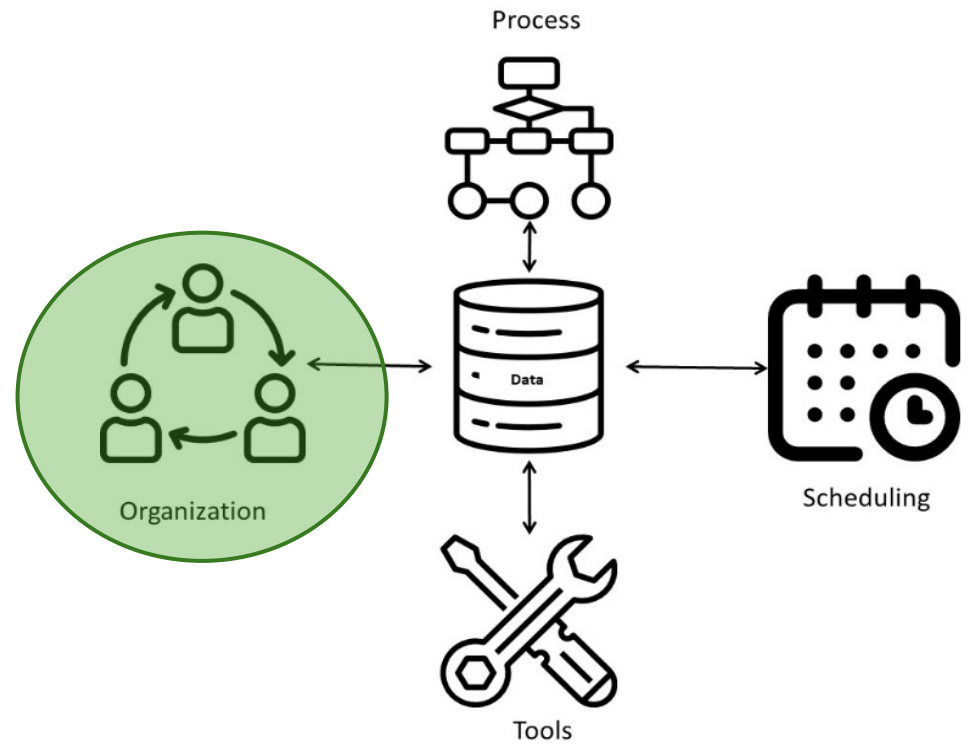
- Process
 - Business Understanding
 - Data Understanding
 - Data Preparation
 - Modelling
 - Evaluation
 - Deployment



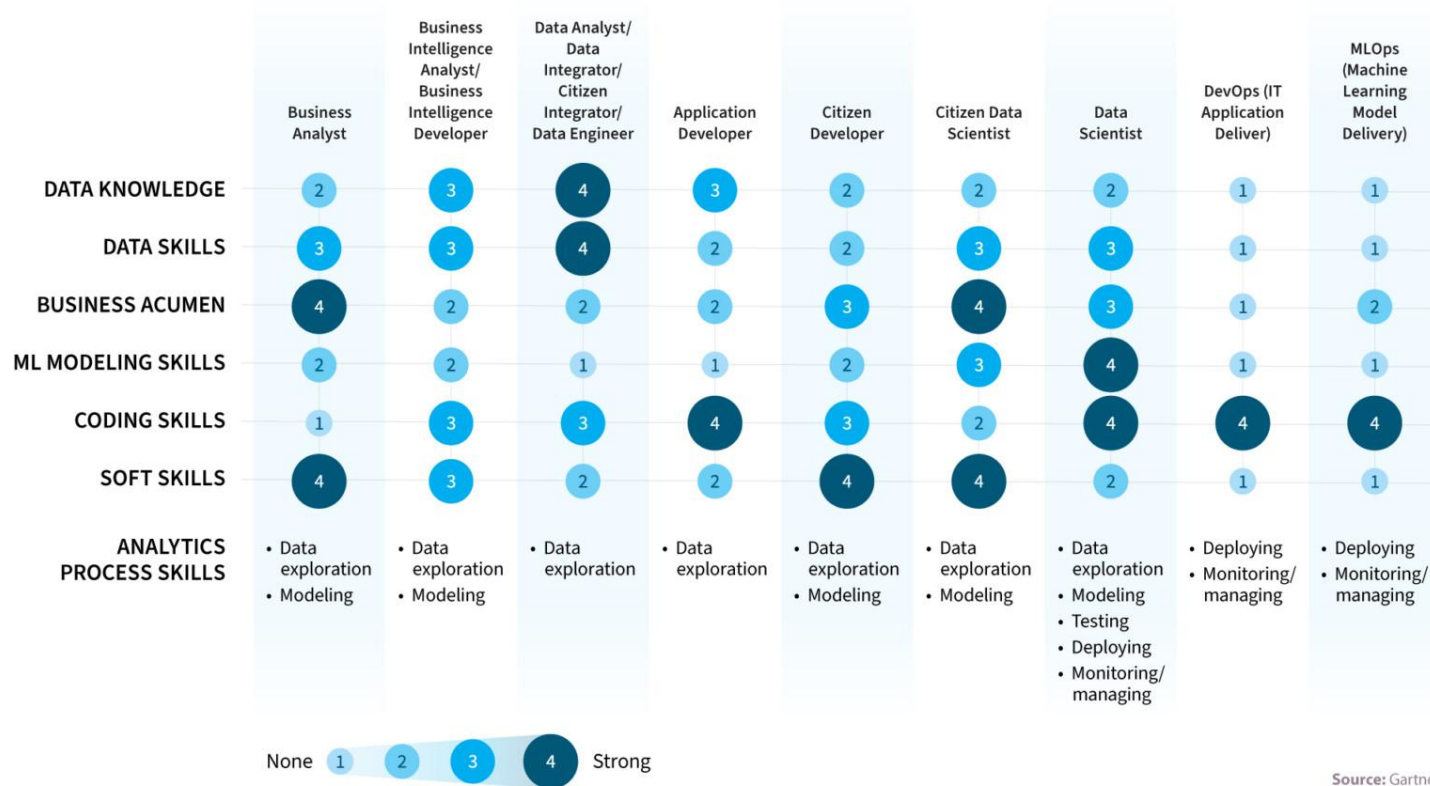
Process

[illegible]

Organization



Continuum of Analytics Roles and Skills



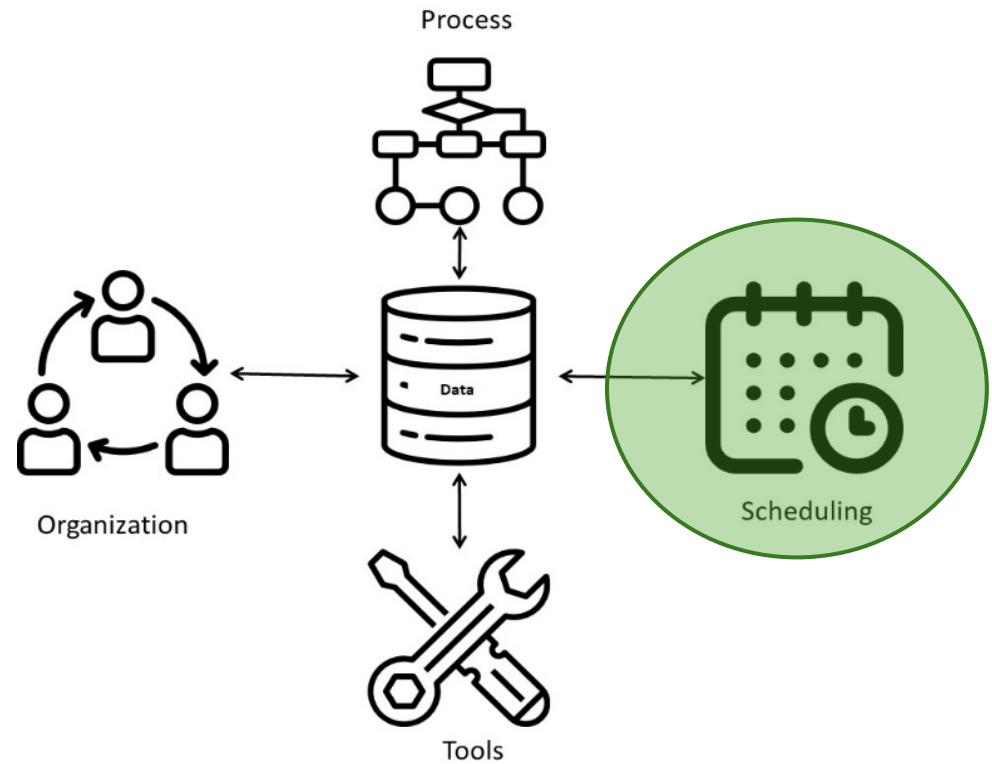
Source: Gartner

Organization

Organization

[illegible]

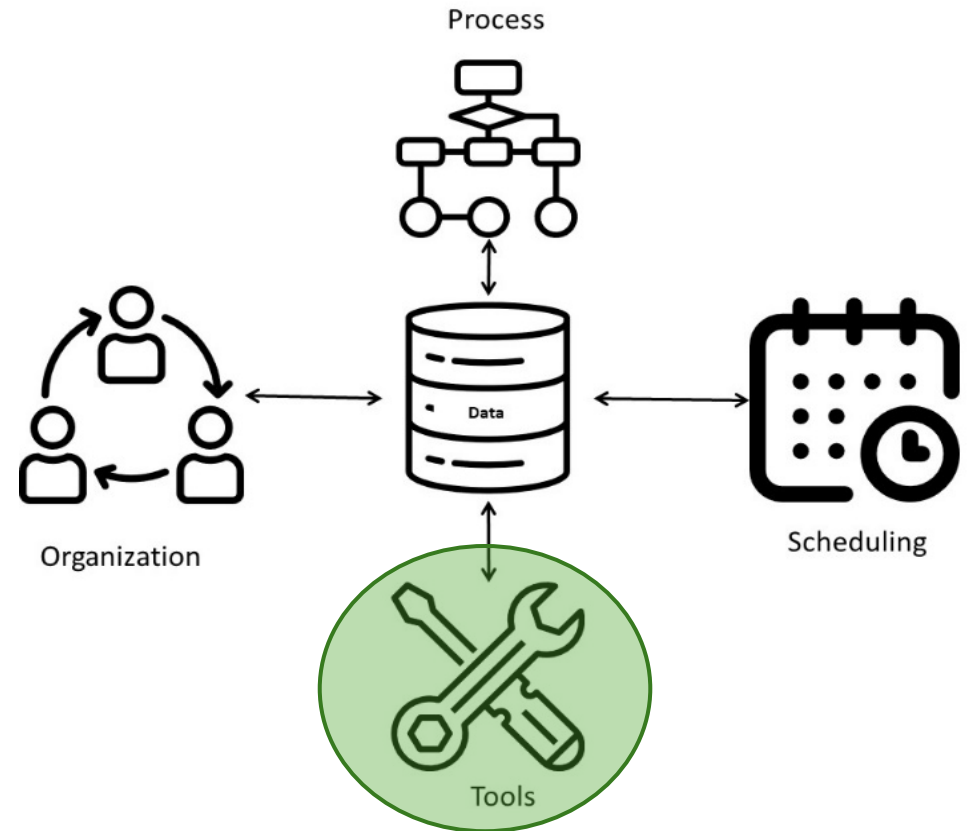
Scheduling



Scheduling

[illegible]

Tools



Tools

[illegible]

Tools

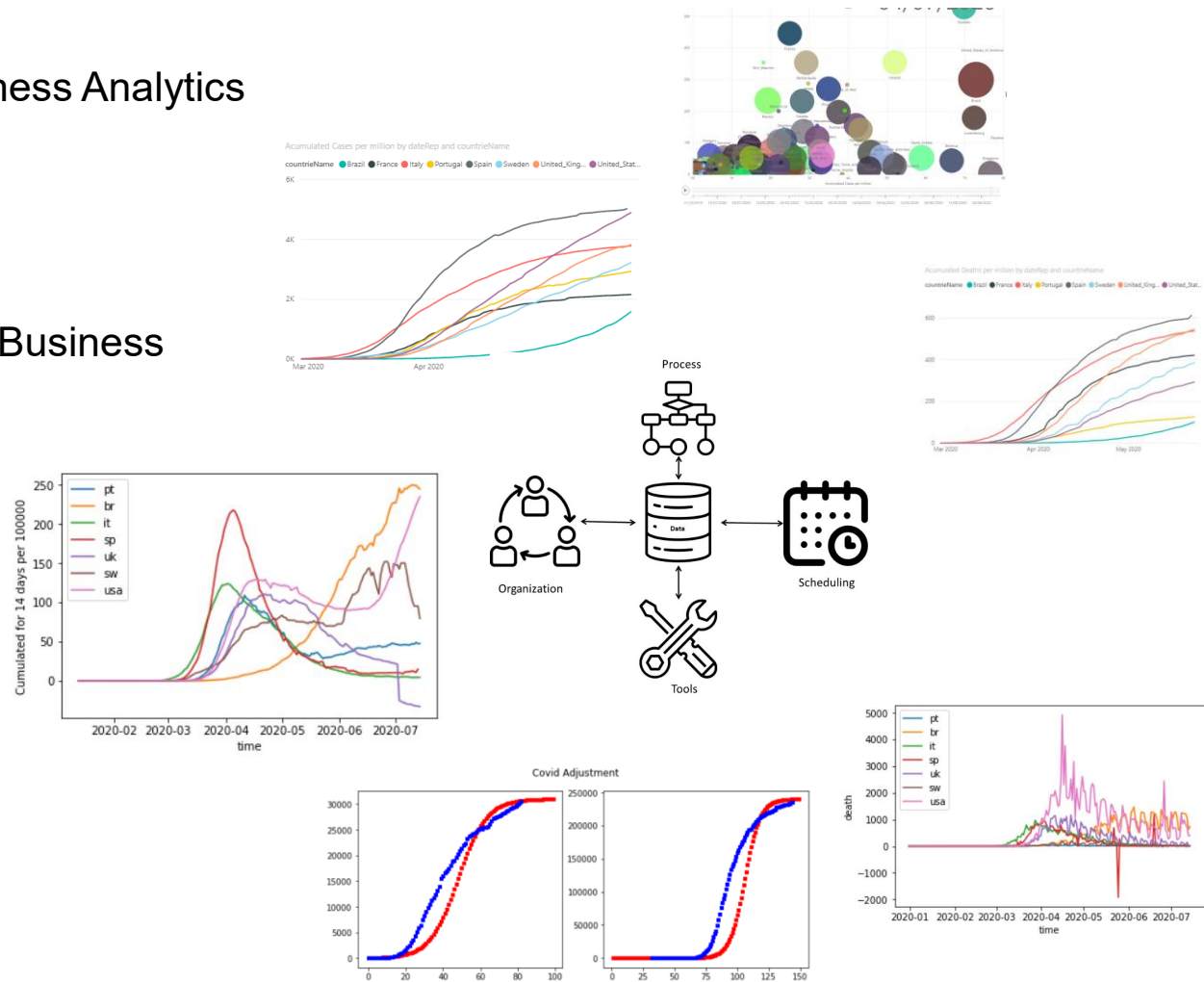
- Charting approaches
- Modeling concepts: supervised algorithms and non supervised algorithms
- Techniques
- Programming Languages

POST-DS

[illegible]

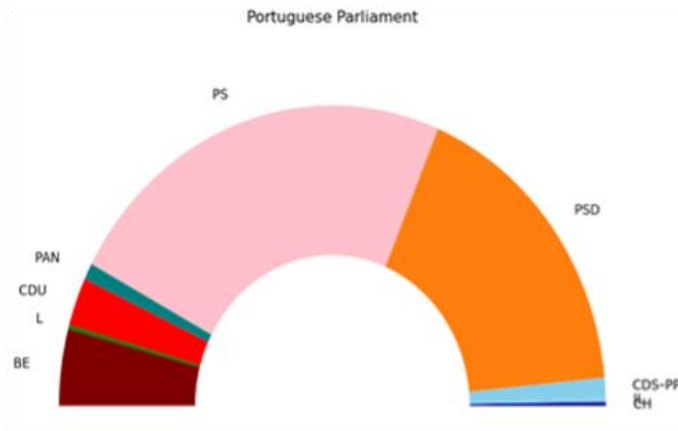
Use Cases

- Data Science and Business Analytics
- Covid
- Financial market
- Software Development Business
- Academic context
- Professional work



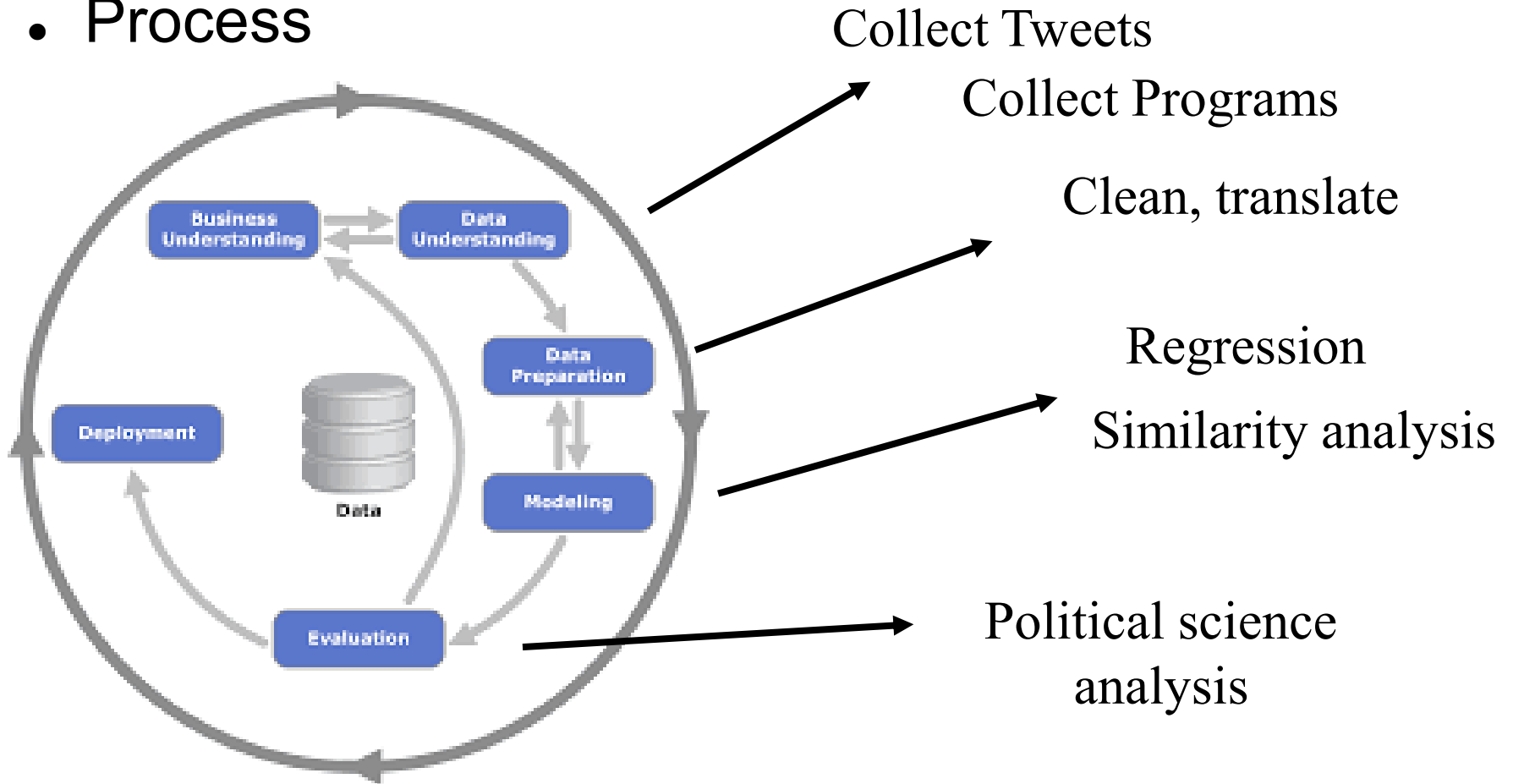
Use Cases

- Emotion analysis of Portuguese Political Parties Communication



Use Cases

- Process



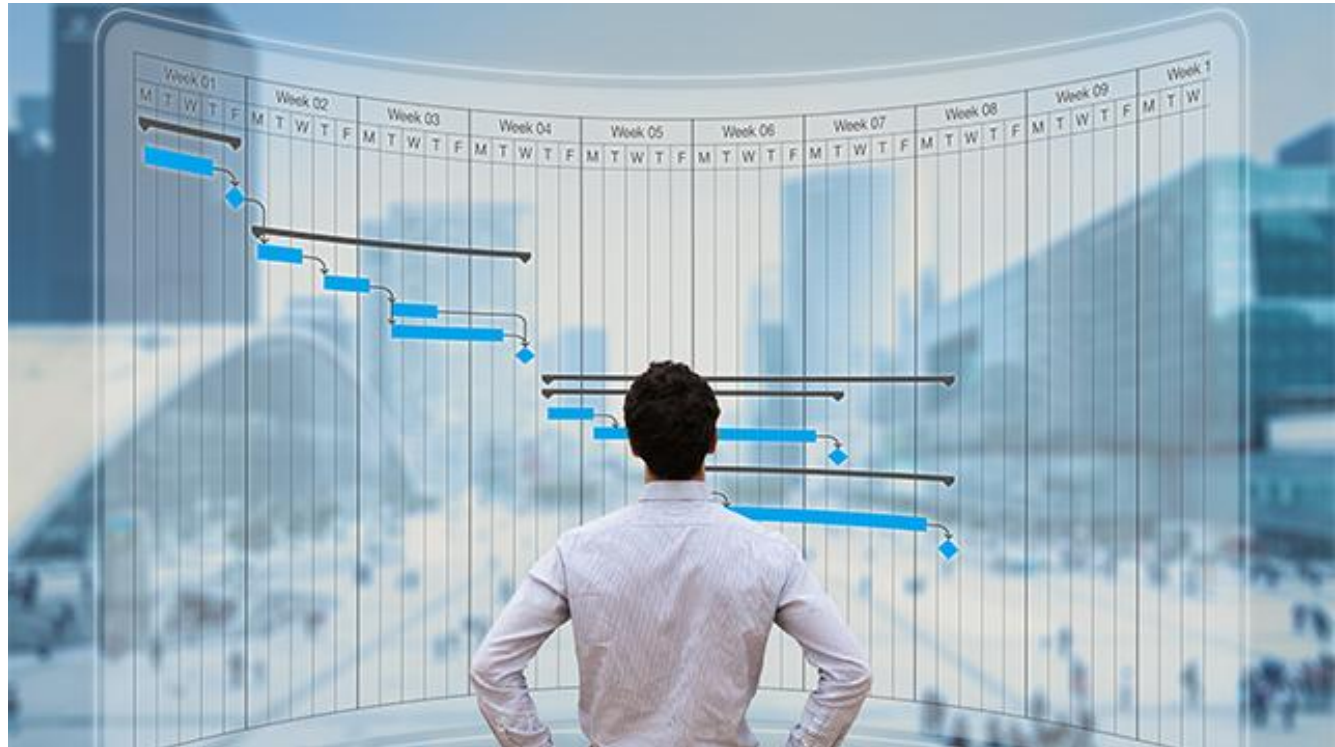
Use Cases

- Process
- Organization
- Scheduling
- Tools



Use Cases

- Process
- Organization
- Scheduling
- Tools



Use Cases

- Process
- Organization
- Scheduling
- Tools



	coef	std err	t	P> t	[0.025	0.975]
const	-1.4423	0.415	-3.473	0.001	-2.256	-0.628
Favorites	0.2296	0.001	349.796	0.000	0.228	0.231
neg	2.0056	0.685	2.930	0.003	0.664	3.347
neu	0.3320	0.435	0.762	0.446	-0.522	1.186
pos	-0.4631	0.594	-0.780	0.436	-1.627	0.701
lenTex	-0.0038	0.001	-4.718	0.000	-0.005	-0.002
Hashtags	0.2503	0.049	5.119	0.000	0.154	0.346
Mentions	0.3500	0.086	4.069	0.000	0.181	0.519
Omnibus:	47844.549				Durbin-Watson:	1.638
Prob(Omnibus):	0.000				Jarque-Bera (JB):	738581430.277
Skew:	15.841				Prob(JB):	0.00
Kurtosis:	865.992				Cond. No.	2.59e+03

```

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3.0
In [13]: # Importing libraries
from __future__ import print_function
from IPythonwidgets import interact, interactive, fixed, interact_manual
from IPython.core.display import display, HTML

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import plotly.figure as go
import folium
import plotly.graph_objects as go
import seaborn as sns
import IPythonwidgets as widgets

In [14]: # Loading data right from the source:
death_df = pd.read_csv('https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_data/confirmed_df = pd.read_csv('https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_data/recovered_df = pd.read_csv('https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_data/country_df = pd.read_csv('https://raw.githubusercontent.com/CSSEGISandData/COVID-19/web-data/data/cases_country.csv')

In [15]: confirmed_df.head()

In [16]: recovered_df.head()

In [17]: death_df.head()

In [18]: country_df.head()

```

Test Data

	R2	MAE	MSE
OLS	0.764957	3.444953	104.684972
Ridge	0.764956	3.444938	104.685139
Lasso	0.765343	3.421643	104.512759
BayesianRidge	0.764878	3.438733	104.720064
Polynomial Regression	0.717263	2.814491	125.927009
Neural Network (MLP)	0.746654	2.942042	112.836870

ACM Digital Library

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

RESEARCH ARTICLE FREE ACCESS

Sentiment Analysis of Portuguese Political Parties Communication

Authors: Carlos Costa, Manuela Aparicio, Joao Aparicio [Authors Info & Claims](#)

SIGDOC '21: The 39th ACM International Conference on Design of Communication • October 2021 • Pages

PDF

IEEE.org | IEEE Xplore | IEEE-SA | IEEE Spectrum | More Sites

IEEE Xplore® Browse My Settings Help Institutional Sign In

Conferences > 2021 10th Iberian Conference ...

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

Publisher: IEEE Cite This PDF

Joao Tiago Aparicio ; Joao Salema de Sequeira ; Carlos J. Costa All Authors

20 Full Text Views

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, 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_7](https://doi.org/10.1007/978-3-030-69984-0_7)