Fundamentals of Data Science

Student: **Alessandro Monolo** | 1790210

———

Lecturer: **Erik Hekman** | MA – Data-Driven Design

Hogeschool Utrecht

August 2021

Index

FDS | Alessandro Monolo | 1790210 | August 2021

Tweets analysis of the principal Italian populist leaders

Introduction: This research aims to analyze the *tweets* of the foremost Italian populist leaders in order to observe the different communication methods adopted by Giuseppe Grillo, for *Movimento 5 Stelle* (left-right-wing party), Giorgia Meloni, for *Fratelli D' Italia* (right-wing party), and Matteo Salvini, for *Lega* (right-wing party). Furthermore, this analysis includes studying the contents expressed within the *tweets* through a sentiment and emotion classification. For this reason, the latest tweets of *@beppe_grillo* (3214), *@GiorgiaMeloni* (2781), and *@matteosalvinimi* (3178) were collected for a total of 9173 *tweets*. This work was made possible by using personal *Twitter* API keys and *Tweepy* library for data scraping. Finally, since the possibility of obtaining *Twitter* data is significantly limited, this study expresses limitations of timeline nature but simultaneously hopes to outline valuable insights for further investigations regarding the different communication methods of the Italian populist leaders and their main similarities/differences.

Theoretical Framework: Technological change played an essential role in the political world during the Information Age. Consequently, today's politics is online and has its exponents with it (Miller, 2020). Along with this, in recent decades, Western society has seen the rise of new political movements called "populists". They appeared together with the Internet Age and resulted from the growing social schism arising from the Network Society (Castells, 1996/2000). To the detriment of ordinary political parties, founded on specific ideologies and principles, populist movements aim to unite voters under the same party, prioritizing electoral success through a mainstream line. Consequently, defining the populist parties and their actual political orientation is complicated. There are populist parties both on the right-wing (such as *Lega* and *Fratelli d' Italia*) and on the left-wing, as well as parties that both have a right-and-left-leaning manifesto, such as the *Movimento 5 Stelle* (Miller, 2020).

One of the main reasons populism has become so central in today's political world is its leaders' massive use of social media (Miller, 2020). *Twitter*, in particular, plays a crucial role in the communication and strategies of populist movements, allowing its leaders to call attention to specific social and political elites by creating 'injustices frames'. In addition, *Twitter* provides a suitable environment (forums) to develop criticism and ambiguity for those in the opposition (Van Kessel and Castelein, 2016). According to Engesser et al.(2016), right-wing populist parties are more active than center/left ones, while center/left parties are more successful on social media. Castels (1996/2000) and Miller (2020), define that parties tend to be negative and promote misinformation, while Golbeck et al. (2010), describes *Twitter* as a self-promotion vehicle for political leaders who use their profile to self-sponsor opinions/events. Finally, according to Van Kessel and Castelein (2016), the *tweets* of some politicians classified as populists do not always aim to criticize the political actions but rather to promote specific ideas or concepts, precisely as other political figures not expressly belonging to populist movements do.

Problem Definition: With the significant use of social media platforms for political issues, the Italian populist leaders were able to gain on Twitter among the last years over 4,9 million followers (*@beppe_grillo* 2.4M, *@GiorgiaMeloni* 1.1M, *@matteosalvini* 1.4M). Their use of *Twitter* occurs for uncommon (or common) reasons and comparable or similar ways. The study seeks to prove whether or not the mentioned populist exponents eventually represent the above-cited guidelines over *Twitter* and *Twitter* usage. This project also explores when *Twitter* followers engage with populist *tweets* and what parameters can influence that engagement.
- What are the principal approaches in using *Twitter* for the Italian populist leaders?
- According to what criterion do interactions with different kind of *tweets* take place?

Methodology: After obtaining personal access keys (*Twitter* API keys) and synchronizing my notebook files with the *Tweepy* library, I scraped the *Twitter* data using the IDs of @GiorgiaMeloni, @beppe_grillo, and @matteosalvinimi as key parameters. For each *Twitter* ID, I stored the latest posts available, collecting a total of 9173 *tweets*. I split them into three datasets, one for each populist leader. Every dataset comprises the *tweet* texts (in extended mode, using the method tweet_mode = 'extended'), the retweet counts, the favorite counts, and the related timestamps. Subsequently, I dropped the retweets for each dataset using the .str.contains('RT') method. Using the same method, have been tracked down where any hashtags (#) or mentions (@) were present, discovering that the proportion between tweets with or without hashtag is balanced for all populist leaders, while the number of mentions is much higher for *@GiorgiaMeloni* (381)

and @*matteosalvinimi* (292), while only 7 for @*beppe_grillo*. Afterward, the length of the text in characters for each *tweet* was calculated using the .apply(len) method, discovering @*GiorgiaMeloni* is by far the one who writes longer with an average *tweet* length equal to 247 characters, greater than 205 by @*matteosalvinimi* and 189 by @*beppe_grillo*. Furthermore, to discover the most mentioned words within the *tweets*, the Stopwords library from NLTK module was adopted to drop the common ones (setting Italian as dictionary language). Then, using the CountVectorizer library from sklearn, it was calculated the total number of words. Thanks to this process, I further analyzed which of the three politicians had a more or less wide vocabulary within their *tweets*. Discovering that, @*matteosalvinimi* can count on a total of 3178 words, @*GiorgiaMeloni* only 2781 words, while @*beppe_grillo* around 2936. The same has been done with the tweet_hashtag data frame column. To obtain timeline data and further insights into the tweet history of each populist leader, using the Datetime library, I extracted the year, the month, and the day of the week for each tweet and created new data frames columns. Finally, to perform a sentiment and emotion classification among all the populist tweets, FEEL-IT, a sentiment and emotion classifier (Bianchi et al., 2021), was adopted to organize them into four emotion classes (joy, fear, anger, sadness) and two sentiment classes (negative and positive). In order to complete the cleaning phase, each dataset was reordered, renamed, and saved in the final csv files, ready for the visualization phase, which was performed using Plotly and Dash.

Results: We can observe how @*beppe_grillo* promotes different themes within its *tweets*: blockchain (32), Cannabis (33), and Alzheimer (22) are mentioned in total 87 times. @*GiorgiaMeloni* and @*matteosalvinimi* instead mainly name their surnames or their parties' names. For example, Salvini appoints himself 539 times, while *Fratelli d'Italia* for @*GiorgiaMeloni* as many as 106. A common theme for both @*beppe_grillo* and @*GiorgiaMeloni* is basic income. @*matteosalvinimi* frequently mentions TV programs, where he is often interviewed or guest, indeed 13 out of 25 hashtags are aimed at television programs (see appendix). As for the use of words, the high frequency stands out for @*matteosalvinimi* (Salvini, 591 times) and Meloni (*Fratelli d'Italia* or *FDI*, 220 and 258 times) self-naming.

Furthermore, again for both of the two right-wing exponents, naming the words 'Italia' and 'Italiani' seems crucial, with an overall frequency equal to 710 times (Italia 333, Italiani 377) for @*matteosalvinimi* and 722 times (Italia 344, Italiani 378) for @*GiorgiaMeloni*. According to emotion and sentiment classification, the *tweets* of @*GiorgiaMeloni* and @*matteosalvinimi* depict a similar situation, where most of the *tweets* are classified as negative and angry (1428 *tweets* out of 2781 for the first one, while 1405 out of 3178 for the second one). @*beppe_grillo*'s *tweets* are again mainly classified as negative but principally joyful from the emotional class (1169 out of 3214). We can also say that @*GiorgiaMeloni* tends to write more in terms of *tweet* text length when angry or sad, with an average length of 263 in the first case and 262 characters in the second one. @*beppe_grillo* tends to write more in sad *tweets* (212 characters), while @*matteosalvinimi* tends to be more balanced according to this type of analysis. Overall, we can see that the feeling of joy characterizes relatively short *tweets* in the number of characters. The level of engagement demonstrated through the retweet and favorite mean counts for each emotional class shows how the *tweets* classified as angry are the most engaged for all the populist exponents.

Finally, regarding the distribution of *tweets* along the timeline, we see a constant and balanced value for @*GiorgiaMeloni* and @*matteosalvinimi* (with a higher density between January and February 2021). @*beppe_grillo*, on the other hand, has decreased the *tweets* since April 2019. In addition, we can observe that the frequency of *tweets* is greater for Salvini, who published 3178 *tweets* from November 2020 to August 2021, @*GiorgiaMeloni* instead published 2781 *tweets* between December 2019 and August 2021, while @*beppe_grillo* 3214 between May 2018 and August 2021.

Conclusion: As Van Kessel et al. (2016) suggests, opposition exponents tend to spread criticism and ambiguity, as we can see from the tweets of @*GiorgiaMeloni* (leader of the main political force currently in opposition), which are mainly cataloged by emotional classes of anger or sadness. Furthermore, we can confirm that right-wing populists interact more than others, with a higher frequency of tweets than those of the center or left *(@beppe_grillo)*.
In addition, we can affirm that @*beppe_grillo* tends to promote different themes than purely political ones, @*giorgiameloni* and @*matteosalvinimi* sponsor their party and above all, @*matteosalvinimi* sponsors itself both online and on TV, observing the large number of TV references mentioned in his tweets and thus confirming the theory expressed by Golbeck et al. (2010). Finally, a shared parameter that defines the level of interaction (retweet and favorite counts) is represented by the 'anger' emotional class, which manages to engage more interactions with users than the other classes.

Bibliography

Bianchi, F., Nozza, D., & Hovy, D. (2021). *FEEL-IT: Emotion and Sentiment Classification for the Italian Language*. WASSA.

Castells, M., (1996/2000). *Rise of Network Society - The Information Age - Economy, Society, & Culture*, vol 1. Oxford: Blackwel.

Engesser, S., Ernst, N., Esser, F., & Büchel, F. (2016). *Populism and social media: how politicians spread a fragmented ideology*. Information, Communication & Society, 20(8), 1109–1126.

Golbeck, J., Grimes, J. M., & Rogers, A. (2010). *Twitter use by the U.S. Congress*. Journal of the American Society for Information Science and Technology.

Miller, V. (2020). *Understanding Digital Culture* (Second ed.). SAGE Publications Ltd.

Van Kessel, S. and Castelein, R. (2016). *'Shifting the blame. Populist politicians' use of Twitter as a tool of opposition'*, Journal of Contemporary European Research. 12 (2), pp. 594 - 614.

Appendix

English translation for the most 25 mentioned words of @beppe_grillo, @GiorgiaMeloni, @matteosalvinimi

Top 25 Words - @beppe_grillo	Count	Top 25 words - @GiorgiaMeloni	Count	Top 25 words – @matteosalvinimi	Count
Ecco – there	837	Governo – government	666	Salvini - salvini	591
Leggete – read	502	Italiani – italians	378	Lega - lega	461
Mondo – world	369	Italia – italy	344	Governo – government	391
Cosa – thing	340	Oggi – today	280	Italiani – italians	377
Qui - here	336	Fdi – fdi	258	Italia – italy	333
Ogni – each	200	Sinistra – left-wing	234	Lavoro – work	267
Città – city	197	Stato – state	231	Oggi – today	254
Essere – to be	196	Fratelliditalia – fratelliditalia	220	Grazie – thank	248
Persone – people	194	Conte – conte	203	Amici – friends	209
Lavoro – work	175	Anni – years	198	Dell – of	179
Futuro – future	175	Grazie – thank	179	Anni – years	176
Dell – of	170	Solo – only	168	Vita – life	171
Anni – years	169	Ancora - again	165	Paese – country	165
Vita – life	158	Litalia – the Italy	162	Diretta – live recording	153
Può – can	155	Essere – to be	159	Roma – rome	150
Senza – without	150	Senza – without	151	Conte – conte	147
Milioni - million	144	Sempre - always	148	Milioni - milion	146
Solo - only	140	Ora – now	140	Sempre – always	143
Oggi – today	138	Lavoro – work	136	Prima – before	139
Sempre - always	133	Cittadini – citizens	132	Salute – health	139
Nuovo – new	132	Dopo – after	130	Tutta – all	136
Prima - before	122	Ogni - each	129	Solo – only	136
Grande - big	117	Fare - to do	128	Referendumgiustizia - justicereferendum	130
Tempo - time	114	Italiano - italian	127	Sicurezza- security	124
Modo - way	113	Nazione - nation	126	Presidente - president	122

English translation for the most 25 mentioned hashtags of @beppe_grillo, @GiorgiaMeloni, @matteosalvinimi

Top 25 hashtags @beppe_grillo	Count	Top 25 hashtags @GiorgiaMeloni	Count	Top 25 hashtags @matteosalvinimi	Count
Roma - rome	58	Fratelliditalia – fratelliditalia	106	Salvini - salvini	539
Redditouniversale – universal income	39	Coronavirus – coronavirus	100	Referendumgiustizia – justice referendum	130
Anvediroma - anvediroma	39	Blocconavale – naval blockade	31	Senato – senate	95
Cannabis - cannabis	33	Meloni – meloni	30	Vociitaliane – voci italiane (tv)	63
Lavoro – work	32	Bastacoprifuoco – stop curfew	28	Portaaporta – porta a porta (tv)	54
Blockchain – blockchain	32	Emiliaromagna – emilia romagna	23	Quartarepubblica – quarta repubblica (tv)	42
Torino – turin	28	Quartarepubblica – quartarepubbloca (tv)	21	Primalitalia – italy first	40
Plastica – plastic	26	Bastasbarchi – stop disembarks	20	Nocoprifuoco – no curfew	36
Mobilitá – mobility	25	covid19 – covid19	20	Fuoridalcoro – fuori dl coro (tv)	36
Futuro – future	24	Cassaintegrazione - payroll subsidies	20	Decretoclandestini – illegal immigrants decree	32
Redditodicittadinanza - basic income	23	Foibe - foibe	17	Cartabianca – carta bianca (tv)	31
Alzheimer – alzheimer	22	Elezionisubito – elections now	16	Crisidigoverno - government crisis	26
Ubi – ubi	21	Calabria - calabria	15	Lariadidomenica – l'aria di domenica (tv)	26
Diegofusaro – diego fusaro	20	Decretorilancio -	15	Lariachetira – l'aria che tira (tv)	23
Cina – china	19	Redditodicittadinanza - basic income	14	Nonstopnews – non stop news (tv)	23
Acqua – water	16	Difendiamoilmadeinitaly – defend made in italy	13	Non - not	23
Energia – energy	15	Conte - conte	13	skytg24 – skytg 24 (tv)	23
Clima – clime	14	Nomes - no mes	13	tg2post – tg2post (tv)	22
Ambiente – environment	14	Nutriscore - nutriscore	12	Mezzorainpi - mezzorainpi	21
Tecnologia – technology	13	covid_19 – covid_19	12	Drittoerovescio – dritto e rovescio (tv)	19
Europa – europe	13	Coraggioitalia – be brave italy	11	Morradimettiti - Morra resign!	19
Robot – robot	12	Recoveryfund – recovery fund	11	tokyo2020 – tokio 2020	18
Basicincome	12	Puglia – puglia	11	Staseraitalia – stasera italia (tv)	18
Virginiaraggi – virginia raggi	12	Dpcm - dpcm	10	Stampaestera – foreign press	18
Rifiuti - garbage	12	Bastatasse – stop taxes	10	Noneladurso - non è la d'urso (tv)	18

FDS | Alessandro Monolo | 1790210 | August 2021