LUISS

Department
of Business and Management

Bachelor's Degree in Management and Computer Science

Course of Artificial Intelligence and Machine Learning

Using Sentiment Analysis for politics: the case of the Italian Political Elections

Supervisor:

Prof. Giuseppe Francesco Italiano

Candidate:

Elio Simone La Gioia

ID No: 257751

Academic Year 2022/2023

Table of Contents

Introduction

Language is the basis of the society we live in; it would be impossible to carry out most of our daily activities. Certainly, much of what we express comes through social networks in an interconnected society such as ours.

The basis of this thesis is therefore a study of language, particularly political language. The past parliamentary elections in September 2022 were a valuable opportunity to collect data from the Twitter profiles of Italy's leading politicians.

There are many motivations for me to delve into the use of language and the analysis of feeling. First and foremost, an interest in political participation, which I believe is more than ever necessary at a time of economic crisis, the war between Russia and Ukraine, and the many concerns that the youths experience in our country. After collecting data and reading up on studies conducted in Europe and the United States, I tried to apply the knowledge I learned in programming and artificial intelligence courses to extrapolate the sentiments of politicians' tweets.

The objective of this thesis is to provide a general overview of sentiment analysis applied to political elections. In addition, there are specific insights derived from the results of the analysis and possible future perspectives of AI in the service of politics in the thesis.

This thesis provides a unique contribution analyzing not only the linguistic and formal aspects of political tweets, but through a broader reasoning attempts to join linguistic and sociological implications to understand in a clear and concise way how a polarized communication has an impact not only on the propagation of the message, but also on the reception of potential electors.

The thesis consists of three chapters: the first contains a general introduction on sentiment analysis with insights into the various possible approaches to conducting the analysis and some of the elements that can hinder this type of analysis. The second chapter, on the other hand, is the more operational one, focusing more on the actual analysis, starting with some examples of sentiment analysis and then going into detail on the Italian parliamentary elections of 2022. The second chapter sets out the entire methodology used for the pre-processing of the data and the subsequent analysis, and at the end of the chapter some results by way of example. The third and last chapter instead deals with the use of social media by the various actors involved in the process of political elections and analyzes - starting from the results of the sentiment analysis - some ethical aspects regarding language and the possible role that AI will play in the future of politics.

Thanks to this thesis, it was possible to analyze some of the most important factors influencing political communication; the results will be detailed in the final conclusions.

Chapter 1
Sentiment Analysis Introduction

- ## What is Sentiment Analysis?

The concept of sentiment analysis involves identifying and classifying subjective opinions contained in source materials (such as documents or sentences) through the application of natural language processing, computational linguistics, and text analytics. An analysis of sentiment usually seeks to determine the attitude of a writer regarding a certain topic, or the overall polarity of the context in which the text is written.

There are several types of attitudes that an author may have. They may be judgments, evaluations, affective states (i.e., the authors' emotional state while writing), or the intended emotional communication (as in, the emotional impact the author wishes to have on the reader).

For sentiment analysis to be of use, it is crucial that it can identify and extract factual information and opinions from the source text materials. Unfortunately, this can be a difficult feat to accomplish.

- ## Sentiment Identification

Sentiment Identification is the procedure by which a piece of text (e.g., a document, a sentence) is evaluated to determine whether it expresses opinions. Generally, sentiment identification relies on the following two assumptions:

(1) the text expresses opinions regarding a single object;

(2) the opinions are held by a single individual.

As part of sentiment analysis, the primary objective is to identify opinion words, which is very important. Opinion words are dominant indicators of sentiment, especially adjectives, adverbs, and verbs that express feelings, such as 'I absolutely love Luiss Guido Carli! The most expensive university in Rome!'

It is possible to divide opinion words into two general categories:

(1) positive words, such as wonderful, elegant, amazing;

(2) negative words, such as horrible, disgusting, poor.

These are also known as polarity words, sentiment words, opinion vocabulary, or opinion-bearing words. A set of opinion words must be generated in advance for us to identify opinion words from a given text. The sentiment identification process is very easy to achieve if we have a rich set of opinion words available, which only requires an analysis of each opinion word in the source text.

Three types of methods are available for generating opinion words:

(1) manual generation, in which opinion words are collected manually; however, this method is very effective (i.e., accurate) but very expensive;

(2) dictionary-based generation methods, using seed lists and growing the lists, such as the ones we used during our bachelor

(3) corpus-based generation methods, relying upon syntactic or co-occurrence patterns in large text corpora to generate words.

In the following section, we briefly examine the two most common types of methods for generating opinion words (i.e., dictionary-based, corpus-based) that can be applied.

- ## Dictionary-based approach

To use a dictionary-based approach it is recommended to prepare in advance a list of seed opinion words, which will be used to help in the process of growing the list of seeds, and then use a dictionary to help in the process of growing the list to generate more opinion words.

It is possible to describe dictionary-based opinion words generation in more detail as follows:

(1) In the first stage, a small seed set is collected manually which consists of opinion words that are known to have positive orientations. For example, [“good”];

(2) As a result, the seed set of opinion words is developed by searching for their synonyms and antonyms in an online dictionary. If we add synonyms to good [‘incredible”, ‘great”, ‘‘wonderful”, ‘‘marvellous”], we obtain a set of positive opinion words; if we add antonyms to ‘‘good”, we obtain a new set of negative opinion words: ‘‘bad”, ‘‘pathetic”, ‘‘awful”;

(3) To eliminate new opinion words, the above two steps would be repeated until the online reference dictionary could no longer provide any more suggestions;
(4) Lastly, it may be necessary to conduct a manual inspection for correction.

In general, dictionary-based opinion word generation is a straightforward process, but it is not able to identify context-dependent opinion words, i.e. words whose opinion orientations heavily rely on the context of the sentence. Let us examine two examples to clarify this concept:

(1) the opinion word "terrible" is presented in the following sentences: "Today I will take a terrible exam" and “The exam was not so terrible as we expected". In the first sentence, the opinion word "terrible" indicates a negative opinion orientation, but in the second sentence, it indicates a positive opinion orientation, demonstrating that opinion orientation is dependent upon the context;
(2) considering the opinion term "long", along with two statements "studying macroeconomics takes a long time" and "my MacBook’s battery lasts a long time", we can determine that this situation is similar to the first phrase.

- ## Corpus-based approach

Unlike dictionary-based opinion word generation, corpus-based opinion word generation is based on syntactic and cooccurrence patterns within large text corpora. The main advantage of corpus-based opinion word generation over dictionary-based opinion word generation is the ability to obtain domain-specific and/or context-specific orientations. In the following example, we illustrate how sentiment orientation can be identified for adjectives within a context:

(1) the first step should be to create a list of seed opinion adjectives;
(2) based on selected corpus, identify additional adjective opinion words and their orientations by using linguistic constraints on connectives, including:
    i. sentiment consistency, which is based on the observation that conjoined adjectives usually have similar orientations. For example, in a sentence such as ‘‘This car is beautiful and spacious’’, if the word “beautiful” is positive also “spacious” is a positive word.
    ii. it is possible to specify different rules for different connectives, such as AND, OR, BUT, EITHER-OR, NEITHER-NOR.

(3) to determine whether two conjoined adjectives have the same orientation or are of different orientations, we will use a log-linear model;
(4) The next step would be to use clustering to develop two sets of opinion words, i.e., positive words and negative words.

The process described above allows us to identify opinion word groups. However, just finding opinion word groups is not sufficient since one opinion word may indicate two distinct opinions within the same domain, for example, ""studying macroeconomics takes a long time " versus ""my MacBook's battery lasts a long time". A basic idea for resolving this problem is to identify all opinion words associated with an object attribute. This process is described in more detail in:
(1) identify pairs of object attribute and opinion word, i.e., [object attribute, opinion word];
(2) then, identify opinion words and their orientations in conjunction with the object attributes;
(3) then, repeat the process.
As a result of the above procedure, context dependency of opinion words can be addressed.

- **Sentiment Orientation Classification**

As the name indicates, sentiment orientation classification is used to determine if a given piece of text has a positive, a negative or a neutral opinion orientation based on the opinion words identified by sentiment identification from that text.

According to the assumption that all opinion words identified in the given text act on one object, based on the assumption that the given text is opinionated on a single object, it is straightforward to classify sentiment orientation by counting both positive and negative opinion words, as in the example 'I absolutely love Luiss Guido Carli! The most expensive university in Rome!'. In the text, the number of positive opinions about the object ‘‘Luiss Guido Carli” outweighs the number of negative opinion words, so the text can be considered positive in nature. As well as counting opinion words, we can also use machine learning approaches for sentiment orientation classification. Below, we will briefly discuss two types of approaches to sentiment orientation classification.

- ## Counting Opinion Words approach

An easy method of sentiment orientation classification consists of counting opinion words, which are generated by sentiment identification based on predefined opinion words. The first step in this process is to assign orientation scores (+1, -1) to every opinion word:
(1) positive opinion words (+1), such as wonderful, amazing, and fantastic;
(2) negative opinion words (-1), such as terrible, bad, ugly.
We can also use a strength value between [0, 1] to evaluate the opinion strength of opinion words in this process. The opinion orientation score of as given piece of text is derived by adding the scores of all opinion words. According to this model, if we assign (+1) to positive words and (-1) to negative words (expensive, for example), then the text about Luiss Guido Carli has a positive opinion orientation score.

Nevertheless, counting opinion words as in the example above is clearly insufficient, since if one were to write, 'There is not one thing I hate about Luiss Guido Carli', it would be assigned a negative opinion orientation score, when in fact it is positive. The problem caused by negation words can be overcome by creating some basic opinion rules to increase the accuracy of simply counting opinion words. For example, it is possible to manually create two simple rules: both ‘‘never ... negative’’ and ‘‘not ... negative’’ represent positive opinions.

- ## Supervised Learning approaches

It is generally the case that counting opinion words can only identify a small number of opinion words, and that only a small number of patterns can be created. Can this task be automatized with a limited amount of manual effort (e.g., finding opinion words and their orientations automatically)? It is recommended to apply supervised learning approaches into sentiment orientation classification to resolve this problem. Supervised learning consists of training and obtaining an opinion classifier containing some target opinion classes, such as positive or negative. To accomplish the objective (i.e., to obtain an opinion classifier), most existing methods of supervised learning can be employed. Now, popular methods include the following:
(1) Support Vector Machine (SVM), the examples are viewed as points in space, with support vectors determining the best classification of the points/examples;
(2) Logistic regression model (LR): This type of model predicts the classes based on a set of variables, either continuous, discrete, or mixed.

Classifiers obtained through supervised learning often exhibit domain dependency, i.e., they perform poorly when tested on documents from a different domain. There are two main reasons for this problem:

(1) language used in different domains can differ significantly

(2) some words mean opposites in different contexts.

The approach to solving this problem can be summarized as follows:

(1) to identify features, labeled data from one group should be combined with unlabeled data from both the source and the target group;

(2) the pivot features should be those that occur often in both groups;

(3) create linear pivot predictors to predict the occurrences of each pivot in the unlabeled data for both groups, and then model correlations between the pivot features and all other features.

In-depth study

Enemies of Sentiment Analysis

- **Synonymy and Polysemy**

There can be several ways in which the same information can be described by different users depending on their context, their knowledge, or their linguistic habits (synonyms). Studies over the time demonstrated, only twenty percent of people use the same keyword to describe well-known objects. As a result, relevant content is often overlooked as searchers and authors often use different words. Additionally, some people use the same words to refer to different things (polysemy). It is possible for the same term to take on varying meanings depending on its context or when it is used by different individuals. Synonymy and polysemy result from the possibility of variability in word usage. These issues pose difficult challenges for sentiment analysis and opinion mining.

Synonymy and polysemy challenges have been addressed through the following approaches:

(1) stemming, which normalizes a certain level of surface variability. It has the purpose of bringing variant forms of a word together (to their morphological roots). As a result of stemming algorithms, the words "stemmer", "stemming", "stemming" and "stems" are reduced to a single root word, "stem". The field of computer science has been studying

stemming algorithms since the 1960s. However, stemming does not address cases in which related words are not morphologically related (e.g., professor and student);

(2) controlled vocabulary, variable word usage has been shown to be an effective method of dealing with these issues. In controlled vocabulary, however, terms must be restricted to a predetermined list of words (it is not feasible to restrict contributors to limit their vocabulary to a predetermined list) and as such is not applied often. For example, there is no feasible way to enforce a particular set of words on websites since contributors can express their opinions in any way they choose;

(3) latent Semantic Analysis (LSA) approach is another method for resolving synonymy challenges. As large collections of text data become more abundant, statistical techniques are becoming increasingly popular for determining the relationship between terms and documents. In LSA, words are simultaneously modeled as constituent words and as words are modeled as occurrences within the documents, simultaneously modeling relationship among documents. Using singular value decomposition of term-document matrixes, LSA can be seen as a linear dimension reduction method. This technique induces similarity among words by reducing dimensions and using fewer dimensions than the number of unique words. A semantic space is created by LSA in which terms and documents that are closely related are placed close to each other in the semantic space. Although LSA is a very effective method for solving the synonymy problem, it only offers a partial solution to the polysemy problem. Since a word's meaning is influenced by other words in the document, LSA provides some assistance with regards to polysemy. The failure occurs, however, because every single word only has one representative point in the semantic space of LSA, rather than more than one.

- ## Sarcasm

For machines to perform natural language processing tasks efficiently, they must be able to reliably identify sarcasm in text. It is a very difficult task to identify sarcasm in text. Sarcasm refers to a form of expression that has the literal meaning opposite to what is intended. “The restaurant was great in that it will make all future meals seem more delicious,” refers to a sarcastic sentence where, although there is no negative term in the sentence, it is intended to convey a negative attitude. Clearly, the example illustrates some difficulties in handling sarcastic pronouncements. Understanding the context, the culture, the topic, and the people involved in the sarcastic statement are all critical to handling the situation effectively. It is not easy for a machine to utilize all this information, but it is especially difficult for a machine to utilize it. Despite extensive study in psychology, cognitive science, and linguistics of the phenomenon of sarcasm, there have been very few computational attempts to analyze it. One of the reasons computational analyses of sarcasm is still very new is that there is no dataset with reliably labeled instances of sarcasm and not-sarcasm. The detection of sarcasm within small text phrases like Twitter data has been attempted by some researchers. However, the drawback of using short phrases to detect sarcasm is the disregard for context in sarcastic situations, which plays an important role. Indeed, it has been demonstrated that lexical features alone are insufficient to detect sarcasm; pragmatic and contextual information are necessary to enhance the detection process. The sentiment analysis research community needs to conduct more in-depth research on analyzing sarcasm.

- **Compund Sentences**

There are two independent clauses in a compound sentence. Two independent clauses may be linked with a coordinating conjunction (e.g. "and", "or", "but", "for") or a semicolon. When dealing with compound sentences, sentiment analysis is challenging. Sentiment analysis presents challenges with sentences such as: "The students enjoyed the school trip, however we did not," or "Despite a pleasant experience, I cannot justify the many positive reviews that it was a great restaurant." Research on sentiment analysis that deals with compound sentences remains largely unexplored.

- **Opinion spammers**

Social media is characterized by the ability to facilitate independent expression by anyone from anywhere in the world without disclosing their identity or fearing negative consequences. It is therefore extremely valuable to have these opinions. This anonymity, however, is not without its costs. In this way, people with hidden agendas or malicious intentions can easily manipulate the system to give the appearance that they are independent citizens and to post fake opinions promoting or discrediting a particular product, service, organization, or individual without disclosing their true intent or the person or organization they are working for in secret. People who engage in opinion spamming are referred to as opinion spammers.

The issue of opinion spam has become increasingly prevalent. Furthermore, there are commercial companies that write fake reviews and bogus blogs for their clients, in addition to individuals who post fake opinions in reviews and forum discussions. It has been reported in the news that there have been several high-profile cases of fake reviews. Such spamming activities must be detected to ensure that Internet opinions are reliable sources of information. It is not exclusively a natural processing language problem that is involved in opinion spam detection, as it also involves analyzing the posting behavior of individuals. In this regard, it may also be considered a data mining problem.

Chapter 2
Sentiment Analysis in action

- ## Some examples of Sentiment Analysis

We have described Sentiment Analysis so far, but now it is time to see in concrete terms what applications it can find in the contemporary world. Thinking about the world of business and companies, especially the larger and more developed ones, we can see that they all have an almost obsessive focus on the customer. A McKinsey & Company paper entitled "*The CEO guide to customer experience*" estimates that 25% of customers abandon a brand or product precisely after a bad experience with customer support. This should make us think hard about how to analyse a user's sentiment from their chat to provide the agent working at that moment with the best way not only to solve the customer's problem, but also to be able to exploit the current sentiment as a business opportunity. Understanding customer behaviour can also be useful in analysing the different stages of purchase and being able to improve the customer journey where needed.

But we must not only focus our attention on the customers, sentiment analysis has in fact repeatedly proven to be particularly useful in the study of competitors. Think of two companies competing directly against each other, it is interesting to note that when one of them does something the public does not like, the other gains in terms of positive sentiment. As an example, we can consider McDonald's, Wendy's and Burger King, the three leading fast food chains in the United States. If McDonald's has a particularly negative sentiment following the release of a particular news story or event, at the same time, Wendy's and Burger King will have a positive increase.

A final example, highly relevant to the analysis we will propose below, is the use of sentiment analysis techniques for political purposes. Sentiment analysis is useful to study attitudes of the public towards political parties, candidates, and issues. This can help political parties and candidates to understand the public sentiment and tailor their campaign messages accordingly. For example, sentiment analysis can be used to analyze the public sentiment towards the policies of the ruling government, which can help the opposition parties to form their strategies.

- ## The case of Italian parliamentary election of 2022

Before moving on to the actual analysis of the tweets of Italian political exponents during the election campaign, we must necessarily explain the political framework that led Italy to the vote on 25 September 2022.

(1) *A little background history*

After the parliamentary elections held on 4 March 2018, there was a parliament without a majority. According to the results, M5S received the largest number of votes among all the party lists competing. Meanwhile, the center-right coalition comprising Forza Italia (FI), Lega (LN) and Fratelli d'Italia (FdI) won a majority in the Senate and Chamber of Deputies.

A lawyer and university professor, Giuseppe Conte, originally proposed by M5S, was commissioned by the President of the Republic, Sergio Mattarella, to form an interim government after almost three months of negotiations between the main political parties. The yellow-green coalition - the party colors of M5s and LN, respectively - helped him appoint a self-declared 'government of change' on 1 June.

A record 34% of the vote for Matteo Salvini's LN in the European Parliament elections of May 2019 marked an exponential growth for the party, while popular support for the M5S declined (15% compared to the previous year's general election). Salvini called for new elections following the League's exit from the government majority due to rising tensions between the two coalition partners. M5S reached an agreement with PD, again under the leadership of Giuseppe Conte, for a new government in the following weeks. Conte II became the new government on 5 September. The formation of the yellow-red government (with PD's red replacing the green of LN) will have immediate consequences, among which is the spillover of two political formations from PD, namely, Italia Viva (IV), led by former party leader Matteo Renzi, and Azione (A), led by Carlo Calenda. IV will remain within the government majority while A will join the opposition.

Early in 2020, when the Covid-19 pandemic was sweeping across the country, the government declared a state of emergency. As a result of disagreements regarding Conte's economic recovery initiatives, Renzi's ministers resigned and the government collapsed. Prime Minister Giuseppe Conte resigned on 26 January. A high-profile government not identifiable with any existing political formula was urged by President Mattarella to be formed by all parties represented in Parliament. Mario Draghi, ex-president of the European Central Bank, was mandated to form such a government. Almost all the parliamentary groups except Giorgia Meloni's FdI supported his government when it took office on 13 February. In terms of parliamentary support, the Draghi government ranked third in the history of the Italian Republic after Mario Monti's 2011 cabinet and Giulio Andreotti's 1978 cabinet, with 535 Deputies and 262 Senators supporting it.

In addition to this change, Enrico Letta replaced Nicola Zingaretti as PD leader and Giuseppe Conte replaced Luigi Di Maio as M5S leader in 2021. The election of the President of the Republic in January 2022 resulted in renewed tensions among the parties supporting Draghi's national unity government. Having been unable to agree on a common candidate, the parliamentary groups decided to re-elect Sergio Mattarella, who was not interested in a second term.

Through 2022, while the Covid-19 emergency eased, the Draghi government was faced with an economic crisis due to the Russian invasion of Ukraine. M5S's increasingly critical attitude toward the Draghi government was exemplified by its failure to support a decree (Decreto Aiuti) concerning economic stimulus to combat the ongoing energy crisis in the Senate on 14 July. Sergio Mattarella rejected Mario Draghi's immediate resignation, asking him to provide substantiation that an alternative majority existed in Parliament. LN, FI, and M5S all voted against the government in the ensuing vote of confidence on 20 July, illustrating the need for an alternative majority. Draghi renewed his resignation the following day. The XVIII legislature was dissolved before the end of its natural term (March 2023) due to this situation, forcing the President to call for new elections on 25 September.

(2) *An overview of the parliamentary elections*

Summer was the time of the 2022 election campaign (for the first time in Italian history). Due to the highly majoritarian spin of the Italian mixed electoral system, the crucial moments of the campaign took place in the very first days, when parties had to bring their respective coalition strategies into play.

While its partners' positions towards the Draghi government had been inconsistent until the previous weeks, the center-right coalition was the first to reveal its strategy. Berlusconi, Meloni, Salvini and Lupi (Noi Moderati: NM) announced their intentions on 27 July to run together as a coalition with the leader of the most popular party serving as coalition leader and prime ministerial candidate.

In the other political corner, things proved to be more complex. As a consequence of the downfall of the Draghi government, PD decided to end its collaboration with M5S, which began under the yellow-red government. This choice also indicated that PD would have continued the Draghi Agenda in its coalition strategies. Therefore, it is important to understand the agreement that was reached on 2 August between PD and the federation made up of Azione (A: Calenda) and +Europa (+E: Della Vedova). During the same time, Letta also negotiated an electoral alliance on 6 August with Alleanza Verdi Sinistra (AVS) to strengthen its left flank. As a result of this agreement, tensions between Letta and Calenda emerged. He considered it impossible to coexist with the left-green alliance in the same electoral coalition because of the latter's support for nuclear power and regasifiers. Calenda informed PD on 7 August that the alliance would be broken. To avoid gathering signatures for a new party, Calenda joined forces with Renzi's IV to develop a centrist alliance (also known as Terzo Polo). Rather than joining the center-left camp, Della Vedova's+E stayed somewhere in the middle, alongside IC (Impegno Civico), the new party headed by Luigi Di Maio.

(3) *Political slogans*

Both major coalitions' slogans accurately reflect the overall campaign dynamic. Meloni's FdI presents itself as being in a position of advantage from the outset, emphasizing its readiness for the campaign (‘Ready’). In contrast, Letta's PD is faced with the challenge of polarizing the campaign by framing it in terms of us-versus-them (‘Chose’), while also resorting to rhetoric that urges voters to choose the 'lesser of two evils' to secure strategic voting.

(4) *Time of results*

Compared to the predictions of political polls, the results of the 2022 elections showed no major surprises. Although Lega declined and M5S recovered, most parties performed as expected, but overall, the number of votes was contained.

In terms of results, we can clearly say that:

- The success of FdI was fueled by significant influxes from M5S and Lega compared to 2018, while the recovery from abstention was relatively limited. The flow of voters from

Lega to FdI from the 2019 European Election is even more evident (almost 4 out of 10 Lega voters opted for Giorgia Meloni's party this time).

- The depletion of Lega, compared to the success of the 2019 European Election, fueled not only FdI but also the pool of abstentions: more than 1 out of 4 voters stayed at home on this occasion.

- Compared to 2018, M5S votes were more than halved (in relative terms). Less than one-third of 2018 voters confirmed their vote for the party led by Giuseppe Conte four years later. One quarter took refuge in abstention (like Lega voters), but 1 in 6 shifted to FdI.

- The Democratic Party, both compared to 2018 and 2019, essentially drew from its own pool of voters. It lost some ground to Italia Viva and Azione (which mainly fed on this flow), but above all, it failed to attract new voters: 3 out of 4 current voters were already PD voters at the time of the European Election (where Zingaretti's party did not shine either).

- The growing abstention was fueled mainly by voters who had voted for M5S in 2018, but also by disillusioned voters from Lega, PD, and Forza Italia.

• Sentiment analysis on the tweets of political leaders

Now to the operational part of our analysis, in this section we will look step by step at how the sentiment analysis was conducted.

(1) *Gathering and importing data*

This is the initial phase of the sentiment analysis process where raw data is collected usually with techniques such as web scraping or via APIs. For this thesis, the data was provided by the *Centro Italiano Studi Elettorali*.

The *Centro Italiano Studi Elettorali* was established at the University of Florence in June 2004, and in 2011 became an inter-university research centre between Luiss Guido Carli and the University of Florence. Its main activity is the study of elections and related institutions and is addressed to an audience ranging from the scientific community to information operators, from the world of politics to all citizens interested in elections.

The source dataset contained 5937 tweets written by leading Italian politicians (whom we have already mentioned extensively in the previous paragraphs) between 28 August and 25 September 2022.

Given the amount of posts and the diversity of topics, it would have been difficult for the purposes of analysis to use the entire dataset. It was therefore decided to filter the rows of the dataset by keywords. The KNIME software was used for this process, in particular the Row Filter node. The node allows for row filtering according to certain criteria. It can include or exclude: certain ranges (by row number), rows with a certain row ID, and rows with a certain value in a selectable column (attribute).

The result is three smaller datasets composed as follows:

- *lavoro* composed of 301 tweets and filtered according to keywords: salario, minimo, precariato, mensilità, stage, lavoro, assu, job, sottopagat;
- *energia* composed of 501 tweets and filtered according to the keywords: gas, caro, energia, rinovvabil, nucleare, emission, Eni, consum, bollett, WATT, elettric, price, cap, prezz, Cingolani;
- *istruzione* consisting of 291 tweets and filtered by keywords: scuol, universit, insegnant, istruzione, ricerca, studio, student, test, facoltà, ragazz, docent, class.

At this point the work proceeds with the import of the data into Python. The initial code serves as a facilitation for a text analysis project in Python by means of installing and importing essential libraries. Specifically, the installation of the "feel-it", "transformers", "nltk", "openpyxl", and "wordcloud" libraries are captured and executed systematically with the aim of extending and optimizing the functionalities of the text analysis project.

In order to further extend the capabilities of the text analysis project, the code consists of importing a spectrum of key libraries. The numpy library provides support for the handling of mathematical functions, while scipy.special facilitates for the computation of special functions. The csv library supports the manipulation of comma-separated value (CSV) file formats in the context of the text analysis project. Similarly, the urllib.request library facilitates for the opening and reading of URLs, and the pandas library provides support for the manipulation and analysis of data structures.

Moreover, the code imports torch, which provides support for machine learning functionalities, re, which enables the usage of regular expressions in Python, and string, which provides a collection of commonly-used string operations. In the context of the text analysis project, the nltk.corpus and stopwords libraries provide a collection of stopwords that can be utilized for text cleaning and processing purposes. Finally, the SnowballStemmer library serves as a tool for stemming, which is the process of reducing inflected or derived words to their base or root form.

Next, the code loads the first Italian sentiment data from an excel file named "Energia.xlsx" into a variable named 'data'. This is achieved by utilizing the pandas library to read the excel file. After loading the data, the code prints the number of rows and columns in the data by calling the 'shape' attribute of the 'data' variable. Then, the code removes the rows containing null values from the 'text' column. This is accomplished by calling the 'dropna' method from pandas, which is used to remove rows with NaN or missing values. The 'subset' argument specifies that only the 'text' column should be considered for removing rows. Finally, the code prints the new number of rows and columns in the data after the null values have been removed.

(2) *Labeling the data*

This part of the code uses the Hugging Face transformers library to perform sentiment analysis on Italian text. Specifically, it loads a pre-trained model called "feel-it-italian-sentiment" from the MilaNLProc model repository. The pipeline function is used to define a text classification pipeline that processes the input text and generates predictions or outputs. The pipeline function takes a model configuration for the chosen task, in this case, text classification. By setting the model parameter to 'MilaNLProc/feel-it-italian-sentiment', the pipeline function loads the pre-trained model with that specific name.

The top_k parameter is set to 2, which causes the pipeline function to return the top 2 predictions with their associated scores. Next, two empty list positive and negative are created to store the predicted positive and negative labels for further analysis. A function called sentiment_analysis(text) is defined that takes text as an input and generates the sentiment prediction based on the loaded pre-trained model.

The classifier object that was created earlier is called inside the sentiment_analysis() function to process the input text and generate the sentiment analysis prediction. The prediction variable is a list of dictionaries that stores the prediction results with their associated scores. Since the top_k parameter was set to 2 earlier, the prediction list contains two predictions with their scores. The sentiment_analysis() function then loops through the prediction list and checks the label for each prediction. If the label is "positive," the score associated with the prediction is appended to the positive list. Similarly, if the label is "negative," the score is appended to the negative list.

(3) *Preprocess and text cleaning*

To make the analysis as effective as possible, it is not enough just to divide the tweets into categories by topics. It is necessary to eliminate many of the elements that are present in politicians' messages such as URLs, numbers, punctuation and stop words. To do this, the code downloads the Italian stopwords and sentence tokenizer from the nltk corpus.

After this, the code defines a function called clean_italian_text which takes in a text input, removes URLs first using regular expressions, then removes numbers and punctuation, and converts the text to lowercase characters. Next, the text is tokenized into individual words using the sentence tokenizer from nltk. The function then removes stopwords from this list of tokens, which are basically common words like 'il/la', 'e', 'è', etc. that are not generally helpful in determining the sentiment of a text.

The remaining tokens are then stemmed using the Italian SnowballStemmer, which reduces the words to their root form. For example, the word 'carburi' would be stemmed to 'carb'. Stemming helps in achieving a higher degree of similarity and reduces the number of distinct words in the text, thus making it easy to analyze sentiment. The stemmed tokens are rejoined into a single string before being returned as the cleaned text.

Once the clean_italian_text function has been defined, the code uses it to create a new column in a dataframe 'data', which contains the cleaned version of the Italian text messages. This cleaned_text column is then passed to another function called sentiment_analysis, which uses TextBlob, a Python library used for processing textual data, to calculate the sentiment polarity of each text.

(4) *Starting the sentiment analysis*

Now we can apply the sentiment_analysis function to each row of the data frame and create a new column called 'sentiment'. This column contains numerical values between -1 to 1, where -1 represents the most negative sentiment and 1 represents the most positive sentiment. Positive score and negative score are divided.

	from_user	text	clean_text	positive score	negative score
0	LegaSalvini	++ #OGGIVOTOLEGA PER IVA ZERO SU PANE, PASTA, ...	oggivotoleg iva zer pan past ris latt frutt ve...	0.0005	0.9995
1	LegaSalvini	++ #OGGIVOTOLEGA PER L'INDIPENDENZA ENERGETICA...	oggivotoleg lindipendt energet nucl sicur sì p...	0.8437	0.1563
2	bendellavedova	Questa campagna elettorale è stata una sfida e...	campagn elettoral stat sfid prov coragg dat ma...	0.9997	0.0003
3	SI_sinistra	Filippo Sestito: “Eni intensificherà lo sfrutt...	filipp sest “ eni intensific sfrutt risors ben...	0.0015	0.9985
4	Mov5Stelle	🔴 | LIVE #DallaParteGiusta\n\nL’intervento di ...	🔴 liv dallapartegiust ’ intervent giuseppecont...	0.9995	0.0005

(5) *Defining the sentiment*

At this point the aim is to compute the score difference of our dataset with positive and negative scores, and create a column for it, as well as a column of sentiments based on the output. First, the code creates an empty list called "score_diff". The for-loop iterates over the indices of the "positive" dataset, and computes the score difference by subtracting the negative score from the positive score for each corresponding index. The result is appended to the "score_diff" list.

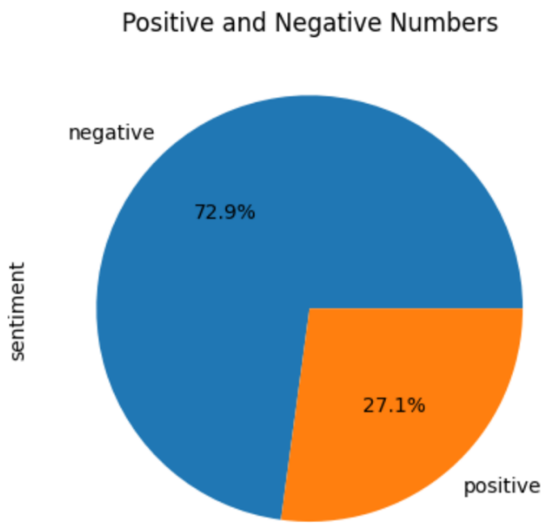
The second for-loop computes the sentiment using the same dataset. The loop iterates through the indices of the "positive" dataset, and checks if the positive score is greater than the negative score, in which case the sentiment is classified as "positive". Similar comparisons are made if the negative score is greater than the positive score, in which case the sentiment is classified as "negative". If, however, the positive and negative scores are equal, then the sentiment is classified as "neutral". The results are appended to the list "sentiment".

Finally, two columns are created in the dataset, "score difference", which contains the score difference computed earlier, and "sentiment", which contains the sentiments computed in the second loop.

	from_user	clean_text	positive score	negative score	score difference	sentiment
0	LegaSalvini	oggivotoleg iva zer pan past ris latt frutt ve...	0.0005	0.9995	0.9990	negative
1	LegaSalvini	oggivotoleg lindipendt energet nucl sicur sì p...	0.8437	0.1563	0.6874	positive
2	bendellavedova	campagn elettoral stat sfid prov coragg dat ma...	0.9997	0.0003	0.9994	positive
3	SI_sinistra	filipp sest “ eni intensific sfrutt risors ben...	0.0015	0.9985	0.9970	negative
4	Mov5Stelle	🔴 liv dallapartegiust ’ intervent giuseppecont...	0.9995	0.0005	0.9990	positive

(6) *Results and visualization*

First, the code imports the Pandas and Matplotlib libraries using the "import" keyword and gives them the aliases "pd" and "plt", respectively. Next, the code calculates the count of positive and negative values in the "sentiment" column of a Pandas DataFrame called "data". This is done using the "value_counts()" function, which counts the number of unique values in a column (in this case, 'positive' and 'negative'). After calculating the counts of positive and negative values, the code creates a pie chart of the positive and negative counts using the "plot.pie()" function. The "autopct" parameter is set to '%1.1f%%', which means that the percentages of each slice of the pie chart will be displayed to one decimal place. The title of the pie chart is set to 'Positive and Negative Numbers' using the "plt.title()" function. Finally, the pie chart is displayed using the "plt.show()" function.

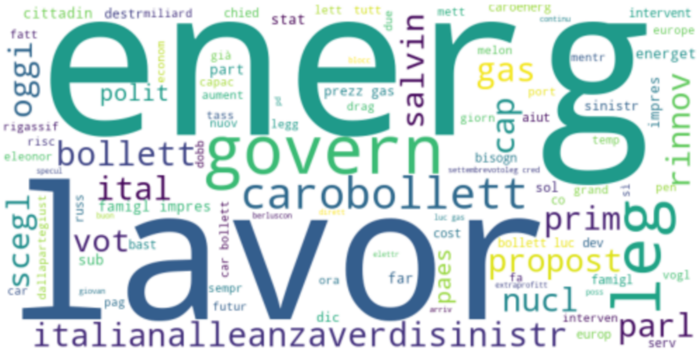


Lastly, the code creates two new Pandas DataFrames called "positive_data" and "negative_data". These DataFrames contain all the rows of "data" where the value in the "sentiment" column is 'positive' and 'negative', respectively. This is done using the "loc[]" function, which filters the rows of a DataFrame based on a certain condition (in this case, the value in the "sentiment" column).

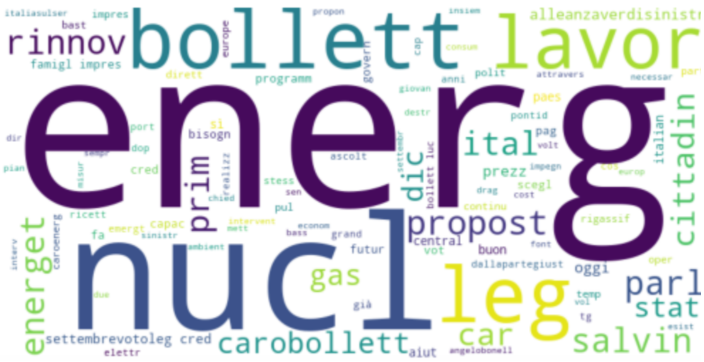
For completeness, it is useful to create three wordclouds: one general, one for positive words and one negative. At this point the code imports the necessary libraries for creating wordclouds, specifically 'wordcloud' and 'WordCloud'. The text that is going to be used to create the word cloud is obtained by concatenating all of the text in a specific column of data into a single string. This is done using the 'join' method and passing in the results of a pandas dataframe's 'tolist()' method that is applied to the specified column.

A WordCloud object is then created by calling the 'WordCloud' function, which takes various arguments such as 'background_color', 'width', 'height', and 'max_words'. The 'background_color' argument sets the background color of the wordcloud, 'width' and 'height' set the dimensions of the object, and 'max_words' is the maximum number of words to be displayed. Stopwords, which are common words that typically do not add much meaning to the word cloud, are defined as a set, and passed in as an argument to the 'stopwords' parameter of the 'WordCloud' function.

The word cloud is generated by calling the 'generate' method on the WordCloud object, and passing in the previously defined text string. Finally, the word cloud is displayed using 'imshow', 'axis('off')', and 'show' methods from the 'matplotlib.pyplot' library.



General wordcloud



Positive wordcloud

Negative wordcloud

In the next chapter, we will describe together the results of the analysis for the three themes. The graph and wordclouds are only to describe the process of the analysis.

Chapter 3
Political implications and negative sentiment

- **Citizen use of social media in politics**

For a moment, let us turn our attention away from Italy, and instead consider the United States, where there are numerous studies regarding how citizens use social media for political purposes.

A growing number of Americans are using social media to engage in political and civic activities. According to a recent survey conducted by Pew Research Center's Internet & American Life Project, 60% of American adults use social networking websites such as Facebook or Twitter, and 66% of these social media users - or 39% of all American adults - have participated in at least one civic or political activity.

The use of social media such as social networking sites and Twitter is characterized by mixed partisan and ideological patterns. People who regularly discuss politics on social media are more likely to use social media for civic or political purposes. Moderates in both parties are less likely than liberal Democrats and conservative Republicans to use social media for these purposes, even when they have stronger ideological ties.

The majority of social media users between the ages of 50 and 54 will engage in some of these activities. As the age of users increases, younger users are increasingly prone to posting opinions on issues, posting links to political material, encouraging others to engage in political action, belonging to political groups on social networking sites, following elected officials on social media, and liking or promoting political materials they have seen posted by others.

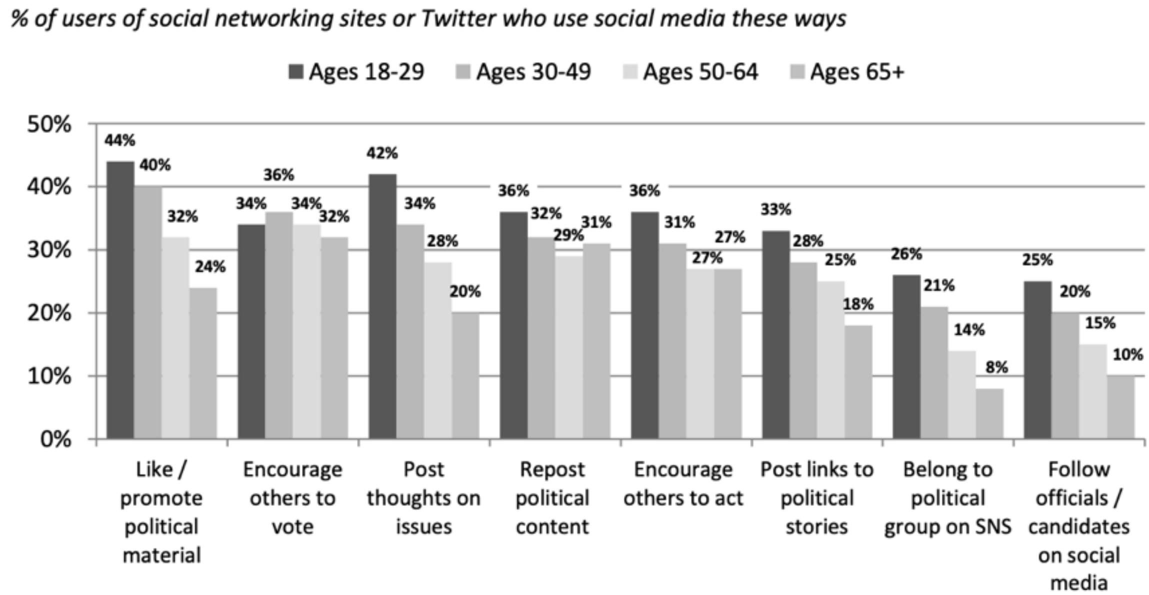
An annual national representative survey of people who use social media such as social networking sites (SNSs) or Twitter revealed the following key findings:

- As the statistics reveal, a significant percentage of individuals- 38% precisely - leverage social media to appreciate or commend contents concerning politics and critical societal matters posted by others. A comparative analysis shows that liberal Democrats utilizing these platforms have adopted the 'like' button in considerable numbers with an estimated count of 52%, while their conservative Republican counterparts only account for approximately 41%.

- The proportion of individuals who have recently taken advantage of social media's tools to champion voting is about 35%. Students, as compared to Republicans and independents, exhibit a higher likelihood in leveraging their use of social media towards advocating for the aforementioned cause. Such behavior is exhibited by approximately 42% amongst students versus that observed from Republican users (36%) and independent users (31%).
- Social media enthusiasts have taken one step further in expressing their opinions by publishing personal remarks and reflections on issues pertaining to politics and the society. Both liberal democratic supporters (42%) as well as conservative republicans followers (41%), stand out for utilizing social media platforms for this very purpose.
- In the domain of online networking, a staggering 33% of participants have partaken in disseminating material that relates to political or social themes by reposting content which had previously been circulated through pre-existing utilities. The statistics reveal Republican members as being more inclined toward this activity compared to their Democratic and independent counterparts with 39% engaging within it while Democrats are present at 34%, followed by independents, trailing behind only slightly at 31%.
- Over the course of three years, a considerable 31% of individuals who utilize social media have taken advantage of said tools in order to motivate and spur others into action regarding politically or socially relevant issues that resonate with them. In terms of political affiliations, Democratic users make up an overwhelming 36%, while their Republican counterparts represent only slightly less at 34%. Comparatively speaking, independents utilizing these platforms lag behind at just shy of one-third of the participation rate (29%).
- Due to the utilization of social media, a considerable 28% of social networking enthusiasts have employed their accounts for disseminating links and articles concerning political news pieces, polls or blogs. Those who are predominantly engaging in this activity happen to be those with liberal Democratic leanings; nevertheless it is conservative Republicans that stand out as being most inclined towards leveraging these platforms for such purposes.

- o As per an estimation, approximately 30% of individuals who frequent social networking or tweeting platforms are affiliated with a group that strives to resolve political or societal matters while utilizing these virtual spaces for promoting their cause. It has been observed that ideological vantage point and political inclinations do not play a significant role in determining one's involvement in using such media outlets as means to support causes.
- o According to statistical approximations, around one-fifth of social media consumers utilize the platform for tracking political representatives and individuals running for government positions. Among those who identify as conservative Republicans engaging with these online channels, approximately thirty-two percent engage in such monitoring practices; on the other hand twenty-seven percent of liberal Democrats use this digital methodology to fulfill their objectives.

Youth use social media tools for civic engagement more than adults



Source: Pew Research Center's Internet & American Life Project Civic Engagement Survey, conducted July 16- August 7, 2012 on landline and cell phones and in English and Spanish. N for social media users ages 18- 29=323. N for social media users ages 30-49=388. N for social media users ages 50-64=323. N for social media users ages 65+=167.

- ## Politician use of social networks in Italy

The use of social media portals has become a crucial constituent in political campaigns globally. In Italy, this inclination is conspicuously conspicuous as politicians and political parties leverage the supremacy of social networks to interact with voters, disseminate their message and consequently impact public viewpoint. In this monograph, we will scrutinize how these digital platforms have been employed within Italy's political arena; expound on noteworthy examples of politician deployment and shed light on the contribution that such networking sites perform toward molding national debates surrounding government policies. Over the last few years, Italian politics has undergone considerable modifications characterized by a surge in nascent political factions alongside variations within citizens' opinions. Social networking sites have been instrumental in this evolution that enables politicians and groups to connect with an extensive audience thereby increasing direct interaction with their followers. Facebook, Twitter, Instagram and YouTube are all significant tools for communicating politically-related messages; they facilitate defining perceptions of politicians as well as designing campaigns effectively across Italy's public realm.

Throughout contemporary history, the highly recognizable and influential figure of Matteo Salvini has been an active participant in Italian political affairs. He presides over the right-wing party named Lega Nord (Northern League). His extensive use of various social media platforms provided him with an opportunity to connect effectively and gain popular support throughout Italy's length and breadth. Throughout his political career thus far, Salvini utilized Facebook, Twitter and Instagram amongst other online forums to voice forth opinions regarding immigration policies, national identity concerns as well as European Union laws.
In the previous general elections, Salvini's social media impact played a decisive role in catapulting Lega Nord to success. The party laid emphasis on several core issues such as territorial security, controlled borders and financial growth that struck a chord with several voters who were anxious about an influx of immigrants alongside compromised national sovereignty. Social media was instrumental for Salvini and his political campaign team by playing host to their message allowing them to gather support from different quarters ultimately making it feasible for the party become third-largest Italian parliamentarian group.

An exemplar exposition of the influence that social networking has had on politics in Italy is evidenced by the ascendancy of a political party known as Five Star Movement (M5S), which is led by Luigi Di Maio, who holds a governmental position. Beppe Grillo, who was once an entertainer, established M5S in 2009 and it quickly gained widespread popularity due to its comprehensive usage of online platforms and social media. Its position against existing authority figures coupled with their call for more honesty and straightforward democracy struck a chord amongst Italians that were disillusioned with traditional political systems. Through employment of social networks during general elections held in 2018, Di Maio along with other members from M5S interactively participated with voters while making public policy proposals thereby uniting support towards his campaign; all thanks to strategically using official blogs as well as Facebook, Twitter & Instagram handles which served significantly in implementing effective communication channels leading up-to performing strongly at polls eventually resulting formation government coalition together partnered alongside Lega Nord Party collectively taking control over governance helm within Italy.

After a political crisis and the fall of Lega Nord-M5S coalition government, in 2019 Giuseppe Conte became Prime Minister of a new amalgamated alliance consisting of M5S and Democratic Party (PD). The previous PM in the fallen regime, he utilized digital social media to engage with people regarding various state initiatives that are underway. He remained active on Facebook, Twitter as well as Instagram throughout his tenure- providing frequent updates on governmental policies by sharing pictures or videos related to his work duties during this period. His effective usage helped bridge an intimate connection between Italian public & governance role thereby personalizing Prime Minister's job description for them via direct communication channel building trust towards it too.

- ## The implications of negative language in politics

In the results of the analysis, 72.9 per cent of the analysed tweets had a negative sentiment, *could it be that politicians prefer to use negative terms because they are more engaging?* There exists a rationale for the usage of pessimistic terminology by politicians. Its effectiveness at seizing public attention is notable; negative verbiage holds considerably more emotional weight than affirmative language, thereby inducing an amplified response among audiences. The root cause behind this phenomenon lies in its tendency to tap into individuals' apprehensions, vexations and anxieties - all substantial catalysts driving decision-making and action-taking behavior.

Furthermore, language with a pessimistic connotation has the potential to sway people's opinions. Whenever politicians employ negative lexicon in their speeches or debates, they aim at persuading individuals that either their adversary or policies proposed by them would prove detrimental for the nation. Politicians can project themselves as superior substitutes and encourage more votes towards them through exposing unfavorable aspects of an opponent's past performance or proposals laid down.

It is imperative to acknowledge that unfavorable verbiage may also give rise to undesirable effects. The use of pessimistic language by politicians can add fuel to the fire, contributing significantly towards an already polarized and divided political landscape we witness today. Furthermore, such adverse phrasing could be perceived as a form of dishonorable politicking which in turn has the potentiality for discrediting any politician employing it.

Might politicians favor negative phrases for they are more captivating? The response lies in a complex dichotomy. Although unfavorable language could foster considerable appeal, it is not invariably the most favorable option among all leaders seeking elections. There exist certain candidates who prioritize espousing upbeat messages which highlight their successes and future prospects instead of vilifying adversaries (this attitude is more common in centre-left parties than in centre-right parties). Ultimately, whether to employ pessimistic rhetoric or opt for an affirmative approach represents a strategic choice that rests on various factors such as personal style and charisma, intended listener demographics along with situational considerations dictated by electoral exigencies prevailing during any given time period within political history's course.

- ## The impact of negative messaging on public debate

Communicating in a way that conveys negativity can exacerbate an already noxious political climate where courteousness and deference are lacking. The utilization of disparaging messages has the potential to generate greater isolationism and discordance among citizens, which renders it more complex for lawmakers to collaborate in resolving issues plaguing their constituents. In addition, negative messaging is inclined towards diminishing voters' confidence levels regarding government transparency and democratic establishments wherein disillusionment surfaces with respect to the electoral procedure adopted by such institutions.

The use of negative messaging can considerably diminish the level of political discussion. A significant outcome is that politicians who utilize such techniques are less probable to participate in meaningful policy conversations; instead, they resort to character defamation and personal offense rhetoric which could deflect from genuine concerns. Consequently, this practice poses a difficulty for voters attempting to make informed judgments on whom or what policies deserve their support.

The usage of negative messaging can result in a deterrent impact on political engagement. The bombardment of pessimistic messages to voters might lead them towards disinterest regarding politics and consequently reduce their likelihood to engage with the political process, thus causing an unfavorable consequence for democratic well-being - reduction in voter turnout.

The democratic system faces significant consequences as a result of negative messaging in public discourse. The overarching dominance of unfavorable messages can trigger erosion in the trust placed by society on government and other related democratic institutions, engendering disillusionment with the political process among citizens. Such dynamics heighten resistance to governance efforts from politicians who would find it more challenging to implement policies effectively amidst an atmosphere that fosters disenchantment towards democracy's ideals.

The dissemination of pessimistic messages can also exacerbate the degradation of democratic integrity. The utilization by political agents of negative messaging as a tool to discredit their adversaries inevitably leads them astray from engaging in constructive debates surrounding public policies, enervating attempts at mitigating multifarious challenges that confront today's societies - such as environmental crises, economic inequities and social equity issues.

Further, the usage of pessimistic communication could add to a reduction in political decorum and admiration. In situations where leaders utilize unfavorable words to discredit their rivals, they are giving out an impression that it is okay for one to make personal accusations or slander against others in politics. This can enhance the toxic ambiance surrounding politics resulting from lack of respect and civility which makes collaborative work among politicians daunting towards finding solutions aimed at tackling issues affecting society as a whole.

- ## The crucial role of media

It is incumbent upon the media to deliver reliable and impartial reporting of political aspirants and topics. Nevertheless, in actuality, negative assaults get more airplay than optimistic postures from such news outlets. This trend carries on because pessimistic appraisals are usually over-the-top or provoke intense feelings when compared to upbeat reports which have a subdued effect thereby leading upticks in viewership numbers translating into greater profits for said services providers.

Whilst it is true that broadcasting negative news falls under the mandate of media outlets, they are also burdened with a responsibility to exhibit impartiality when framing and presenting such information. A marked lack thereof can contribute towards a malignant political climate where decorum and courtesy are scarce resources. What further compounds this issue is that the portrayal of adverse messaging by media outlets holds sway over public sentiment as well as influences how candidates or topics are perceived in society at large.

The broadcast industry's dissemination of adverse communication can wield an immense influence on the political topography. In cases where the mass media accentuates negative messaging, it could foster a reduction in confidence that people have towards their government and democratic establishments; subsequently making it harder for politicians to effectively govern since they would encounter greater opposition from voters who are disenchanted with how politics work.

Additionally, the media's portrayal of detrimental messaging can lead to a deterioration in politeness and admiration. In instances where the media provides an excessive amount of attention towards negative accusations as opposed to affirmative statements, it creates a perception that engaging in personal affronts or insults within politics is tolerable. Consequently establishing an atmosphere which lacks decorum and respect contributing to problems faced by politicians while attempting cooperation on dealing with challenges prevalent among their communities.

The impact of negative messaging in media coverage can lead to electoral implications. The emphasis that the press places on unfavorable assaults towards political figures might mold how citizens perceive them, thus swaying public opinion. This could result in significant obstacles for candidates who are targets of such attacks and may hinder their ability to win an election, regardless of possessing an impressive background or persuasive agenda.

This raises an important ethical question about the use or not of negative content by politicians on social media. Citizens can be divided into two sides:

- against negative messaging contend that the utilization of negative messaging is morally questionable since it hinges on instilling fear and vilifying rivals. The assertion stands that negative messaging may compromise the uprightness of political proceedings by dwindling meaningful policy deliberations into malicious accusations and character defamation. Negative communication has the potential to foster a pernicious, hostile atmosphere in politics where courteousness and admiration are void, thereby impeding politicians' efforts aimed at tackling their region's problems together. Moreover, it has been contended by certain individuals that negative messaging is unethical on the grounds that it capitalizes upon human emotions and apprehensions. By tapping into a person's distress or disillusionment, pessimistic rhetoric may render them more malleable to outside influence. These citizens interpret this method as morally objectionable since it seeks to take advantage of individuals' vulnerabilities rather than participate in substantial deliberations concerning policy-related issues.
- in favour of negative messaging with citizens of this side supporting a common argument which asserts that negative communication holds a valid position in the field of politics, and politicians hold discretion to employ any messaging deemed effective. Advocates maintain that adverse rhetoric exists as an integral facet within political correspondence, providing opportunity for candidates to emphasize crucial issues while simultaneously distinguishing themselves from their competing opponents. It is contended by certain individuals that unfavorable messaging assumes a crucial role in circumstances where there exist pronounced contrasts with regard to policies between nominees or parties. During such instances of significant disparity, denigrating communication can enable electors to get an insight into the varied disparities among contestants and thus formulating an astute decision on who should be bestowed their vote forensically. It has been contended by certain individuals that the utilization of unfavorable messaging is a vital instrument in ensuring that politicians are held responsible for their actions. Negative communication can serve to bring forth awareness regarding inadequacies and flaws exhibited by political candidates while also illuminating regions where they have failed to deliver on commitments made during electioneering campaigns. This approach may be viewed as an indispensable element of effective political discourse which functions as a means of ensuring transparency and accountability among lawmakers towards those whom they represent within society at large.

- **What AI could do for politics and possible future scenarios**

  Various scenarios can be expounded on regarding how artificial intelligence has the capability to metamorphose politics in diverse approaches. We could summarize a few of the conceivable courses for AI's potential utility that could enhance governance at some point:

  - predictive analytics: through the cognitive capacity of AI to analyze copious amounts of information and uncover trends for predicting upcoming occurrences, predictive analytics are rendered feasible. Its applicability in politics lies within its potentiality to anticipate election outcomes, detect plausible policy quandaries and foretell social as well as economic tendencies.
  - personalized campaigns: utilizing artificial intelligence, one can devise bespoke communication approaches tailored to the psyche of every voter by analyzing their social media conduct and other pertinent data. This particular technique could aid political campaigns in delivering highly germane and efficacious messages that would potentially influence each discrete elector’s decision-making process.
  - fact-checking: with the aid of Artificial Intelligence, political statements and claims can be automatically scrutinized for factual accuracy. This application is instrumental in averting misinformation dissemination or broadcasting falsified content.
  - speech analysis: the implementation of artificial intelligence in examining political speeches and debates presents an opportunity to pinpoint regularities in usage, inflection and attitude; thus providing a foundation for prognosticating electoral results as well comprehending voter partiality.
  - voter analysis: A profound use of artificial intelligence lies in the ability to dissect voter data, detecting regularities and inclinations that can aid political campaigns grasp constituents' predispositions and fabricate more impactful messaging (sentiment analysis in this thesis is only a starting point).
  - decision-making: utilizing Artificial Intelligence, decisions can be automated in the realm of politics. For example, AI could potentially determine propitious policy solutions or focus on assigning priority to governmental expenditures.

Conclusions

This paper focused on the study of sentiment analysis in the September 2022 parliamentary elections. The main objective of the thesis was to assess the impact language has on political communication through the technique of sentiment analysis. In this case, the elections proved to be an excellent opportunity to collect real data from the tweets of political figures.

In terms of the results achieved, it is safe to say that the prevailing sentiment in the messages is negative. A negative sentiment usually leads to a polarization of users as well as to an increased virality of the content. As ethically questionable as voter polarization is, we must say that it is necessary for a successful political communication campaign.

Certainly, such an analysis was not easy to conduct. Especially for the Italian language, there is still a lot of work to be done starting with the basic elements that allow this type of analysis. For example, the lack of complete dictionaries in Italian with the respective positivity and negativity scores constituted a major slowdown in the work and certainly also had an impact on the reliability of the results, which however on the whole proved to be in line with expectations.

The analysis obviously has its limitations due to the model used, which, however accurate and elaborate it may be, carries with it all the difficulties due to the study of language discussed at length in the first chapter of this thesis: synonymy, polysemy, sarcasm, compound sentences. The Italian language is full of these peculiarities that limit the predictive capacity of the score assigned to the term in the dictionary, precisely because the same term in different contexts can drastically change its meaning from positive to negative or vice versa.

In the future, there will be much to question about the role and impact that powerful analyses such as sentiment will have on political life and thus on the lives of each and every one of us. Some possible developments for future studies might include:

- *Exploring the Impact of Sentiment Analysis on Political Campaign Strategies: A Comparative Study of International Elections*

- *Deepening Understanding of Voter Behavior: Leveraging Sentiment Analysis and Social Media Data*
- *Algorithmic Bias in Sentiment Analysis: Assessing Its Effects on Political Discourse and Public Opinion*
- *The Role of Sentiment Analysis in Fighting Disinformation and Fake News in Political Campaigns*
- *Sentiment Analysis and the Rise of Political Polarization: A Causal Investigation*
- *Real-time Sentiment Analysis for Political Debates: Enhancing Voter Engagement and Decision-making*

Bibliography

- Luo, T., Chen, S., Xu, G., & Zhou, J. (2013). Sentiment Analysis. In *Springer eBooks* (pp. 53–68). https://doi.org/10.1007/978-1-4614-7202-5_4
- Farhadloo, M., & Rolland, E. (2016). Fundamentals of Sentiment Analysis and Its Applications. In *Studies in computational intelligence* (pp. 1–24). Springer Nature. https://doi.org/10.1007/978-3-319-30319-2_1
- *A Survey on Sentiment Analysis and Opinion Mining*. https://docplayer.net/44230052-A-survey-on-sentiment-analysis-and-opinion-mining.html
- *Social Media and Political Engagement*. https://docplayer.net/15760594-Social-media-and-political-engagement.html
- *Row Splitter — NodePit*. Iacus, S. M., Porro, G., Curini, L., & Ceron, A. (2012). *Every tweet counts? How sentiment analysis of social networks can improve our knowledge of citizens' policy preferences. An application to Italy and France* (No. 2012-19). https://ideas.repec.org/p/mil/wpdepa/2012-19.html
- Ramteke, J., Shah, S., Godhia, D., & Shaikh, A. (2016, August). Election result prediction using Twitter sentiment analysis. In *2016 international conference on inventive computation technologies (ICICT)* (Vol. 1, pp. 1-5). IEEE. https://ieeexplore.ieee.org/abstract/document/7823280
- Neri, F., Aliprandi, C., Capeci, F., & Cuadros, M. (2012, August). Sentiment analysis on social media. In *2012 IEEE/ACM international conference on advances in social networks analysis and mining* (pp. 919-926). IEEE. https://ieeexplore.ieee.org/abstract/document/6425642
- Garzia, D. (2023). The Italian parliamentary election of 2022: the populist radical right takes charge. *West European Politics*, *46*(5), 1038-1048. https://www.tandfonline.com/doi/abs/10.1080/23248823.2022.2150934