Gender Politics - Text Mining Final Project

Alice Iordache

Università degli Studi di Milano, Italia

Abstract. The purpose of this project is to evaluate gender distribution in different legislatures of the Italian Republic and quantitatively analyze the implications of women representation in the Parliament, by the frequency of their interventions during assemblies. We will perform a temporal analysis and qualitatively asses the development of women's presence in politics. Lastly, a further insight will be provided by comparing political parties and their gender composition, both with respect to time and the number of speeches made in the periods evaluated. We hope this work would be beneficial to provide a more conscious picture of the actual women representation in the Italian political sphere.

1 Introduction

The Italian Parliament is the national parliament of the Italian Republic. It is the representative body of Italian citizens. It is a bicameral legislature with 945 elected members and a small number of unelected members. The Italian Parliament is composed of the Chamber of Deputies (with 630 members or elected on a national basis) and the Senate of the Republic (with 315 members elected on a regional basis, plus a small number of senators for life, either appointed or ex officio). The two Houses are independent from one another and never meet jointly, except under circumstances specified by the Constitution. By the Republican Constitution of 1948, the two houses of the Italian Parliament possess the same powers; perfect bicameralism has been coded in the current form since the adoption of the Albertine Statute, and resurged after the dismissal of the fascist dictatorship of the 1920s and 1930s.

The main prerogative of the Parliament is the exercise of the legislative power. The Chamber comprises 630 members elected by universal suffrage and by citizens who are 18 or older. Out of the total number, according to the Rosatellum bis, 232 are directly elected in single-member districts, 398 are elected by proportional representation on a national basis (of whom 12 are elected in overseas constituencies). The Chamber is composed of all members meeting in session at Montecitorio. The assembly also has the right to attend meetings of the Government and its ministers. If required, the Government is obligated to attend the session. Conversely, the Government has the right to be heard every time it requires. The President of the Chamber of Deputies performs the role of speaker of the house and is elected during the first session after the election.

2 Research question and methodology

Starting from a collection of public speeches in the Italian parliament, we want to address the issue of gender representation in political discussions. In particular, we divide between two periods of time from which the corpus has been extracted, namely: the I Legislature of the Italian Republic, in office from June 25 1946 to January 31 1948; the V Legislature of the Italian Republic, in office from June 5 1968 to May 24, 1972; the VI Legislature of the Italian Republic, in office from 25 May 1972 to 4 July 1976; the VII Legislature of the Italian Republic, in office from 5 July 1976 to 19 June 1979. The other period includes two legislatures, namely: the XVII legislature of the Italian Republic, in office from 15 March 2013 to 22 March 2018 and the XVIII legislature of the Italian Republic, in office since 23 March 2018 and currently in progress.

2.1 Dataset

The dataset is made of a data frame displaying four columns: the `id` of the text, the `convocation_id` which is the identification code for each session and the text. Our main challenge is to extract each speech with the corresponding speaker from each session. In order to do that, we define three important functions that we will use: one needed for finding the name of the president of each session, the second one needed for dividing each session from the next one, and the third one needed for extracting the single speech made by each deputy. After this steps, we discard all the speeches made by presidents, since they are not particularly relevant for our analysis, given the tasks of presidents during assemblies. Next, we apply the classical text cleaning function on the whole data set we obtained. At this point, we extracted the whole text of each session and the name of its president. The next step is to separate each speech with the corresponding speaker. In order to find this combination, we download the available data from *dati.camera.it* in order to get the legislature for each speech, the gender of each deputy and the party to which she/he belongs to, accordingly.

2.2 Data pre-processing

We want to approach the first research question issue: establishing the gender of the speakers. In order to do that, we used the API endpoint of the website https://dati.camera.it and performed the necessary queries through `sparql` in order to obtain the names of deputies for each legislature we are analyzing and their gender, accordingly. We download each list as a csv file and filter it by dropping out duplicates (i.e. one deputy could have been present in more than just one legislature). We obtain a unique csv file with the list of all deputies and their gender. Further, we have to find the correspondence between the gender of

each speaker from the csv file and their name in our dataframe, in order to be able to recognize her/his gender. Moreover, we considered the different combinations of names and surnames that the OCR could have possibly found: there is the possibility that some deputies may have the same surname, but different gender. In addition, some of them are noted down either by surname or fully by name and surname in our dataframe column `deputy`. Performing this operation, we end up with 200000 speeches to which we cannot attribute the gender of the speaker. This can be due to many reasons; among the most plausible ones we can think of,it could happen because there is a misspelled name resulting from the OCR and all of their repetitions afterwards. We drop all those speeches where we could not find the associated gender to the speaker and convert the labels "female" and "male" to "1" and "0", respectively.
Finally, the gender distribution of deputies can be seen in the following graph:

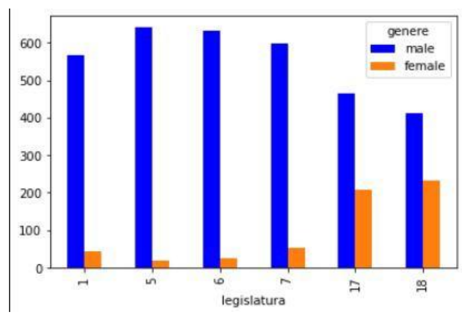

Fig. 1. Gender distribution per legislature

From this, we can clearly see that labels are highly unbalanced.
Moreover, we can count them by the speeches they made in all the considered legislatures as follows:

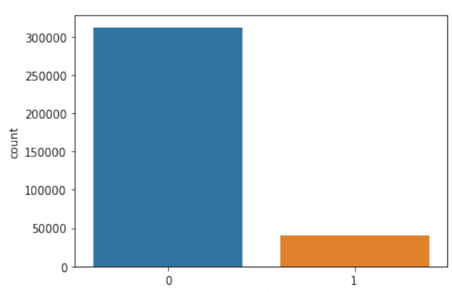

Fig. 2. Gender distribution per speeches

As we can see from the bar chart, labels are highly unbalanced: there are far more speeches made by men than their counterpart. In the third section of this work we will explore the deeper implications of this immediate result.

Continuing with our analysis, we want to delve into a more informative study, namely both an temporal one and an inter-party one. The first step is to associate the single speeches we obtained to their corresponding legislature, in order to see how the gender composition and the amount of interventions in the Chamber has changed over time. We match each speech to its corresponding session, and therefore extract the date for that speech. We only consider the different ranges of legislatures and from them we can attribute the correct one to each intervention. From this, we also count the gender distribution of speakers for each speech for each legislature. The final data frame will look like this:

gender	female	male	total
legislature			
1	915	67422	68337
5	1645	61623	63268
6	1653	47436	49089
7	5453	39896	45349
17	20448	72500	92948
18	10303	23451	33754

Fig. 3. Gender distribution per speeches per legislature

As we expected, if we look in particular at the female distribution, we can see that the speeches made by them is steadily increasing over the years; the lower value of the last legislature could be due to the fact that this is an on-going legislature, therefore we don't have all the speeches that have been made yet. However, looking at the global trend, we can expect that it will display a higher value than the previous legislature.

The next step is to take a glimpse at the changes in the political forces that have presided over thee span time we considered. In particular, we want to obtain the belonging party of each deputy and the period of time in which she/he belonged to it in a clearer way (i.e. : we want to create a column in the data frame which contains the name of the party for each speaker). However, we will raise an important issue that needs special attention to deal with in the analysis, for which deputies change party during the very same legislature or among different legislatures.

Also in this case, in order to get the party for each deputy, we will create a function from scratch. We look at which day a speech has been made; the range we consider here is the one coming from the official records where we have the period of belonging to a party of each deputy; next, we know the name of the deputy, we match it with the corresponding row in the official records and if we find more rows in the data frame it means that the person has belonged to more than one party. We choose the party in which we see the match between the date of that speech and the one of the party for that period. The data frame we obtain will be:

	convocationid	deputato	discorso	genere	date	legislatura	partito
0	14593	VIVIANI LUCIANA	onorevole mini stro rivolgo visto d onorevoli ...	female	1948-10-09	1	COMUNISTA
1	14593	SCELBA	minis t ro interno mini stero interno entra af...	male	1948-10-09	1	DEMOCRATICO CRISTIANO
2	14593	VIVIANI LUCIANA	circolare co munque emanata ministero interno ...	female	1948-10-09	1	COMUNISTA
3	14593	SCELBA	mini s t ro interno vigore tutte funzioni pass...	male	1948-10-09	1	DEMOCRATICO CRISTIANO
4	14593	VIVIANI LUCIANA	passaggio av venuto molto recentemente determi...	female	1948-10-09	1	COMUNISTA
...	...	...	...	...	...	...	...
352740	25970	DAVIDE CRIPPA	grazie presidente colleghi concittadini oggi d...	male	2020-03-11	18	MOVIMENTO 5 STELLE
352741	25970	NUNZIO ANGIOLA	misto presidente cari colleghi cari ministri u...	male	2020-03-11	18	MISTO
352742	25970	NUNZIO ANGIOLA	misto avvio concludere ecco caro presidente ca...	male	2020-03-11	18	MISTO
352743	25970	VITTORIO SGARBI	mniuseic ac certamente onorevole presidente cr...	male	2020-03-11	18	FORZA ITALIA - BERLUSCONI PRESIDENTE
352744	25970	LAURA CASTELLI	sottosegretaria stato leconomia finanze presid...	female	2020-03-11	18	MOVIMENTO 5 STELLE

352745 rows × 7 columns

Fig. 4. Data frame with parties

From the data frame above we found that some deputies have made speeches when belonging to different parties (in the very same legislature or not). This is a problem when they made more than one speech in the same legislature but in different parties, because the count of gender distribution in the legislature and the female/male representation in the party is counted for how many times they have switched party. In order to solve this problem and not alter the global gender distribution, we decided to keep only one party-match for each deputy: the criterion used for this choice is based on the maximum number of speeches made by each speaker while belonging to a given party. Therefore, a deputy will only have as party belonging the one in which they have made most interventions, even though they have switched parties in the same legislature. The problem does not apply when they changed party in different legislature.
Finally, we extract the most speeches per party as we explained above and get (here only a part of the data frame is depicted because of spatial constraints):

party	ALLEANZA NAZIONALE	ALTERNATIVA POPOLARE-CENTRISTI PER L'EUROPA-NCD-NOI CON L'ITALIA	ARTICOLO 1-MOVIMENTO DEMOCRATICO E PROGRESSISTA-LIBERI E UGUALI	CIVICI E INNOVATORI	COMUNISTA	COSTITUENTE DI DESTRA - DEMOCRAZIA NAZIONALE	DEMOCRATICI DI SINISTRA - L'ULIVO	DEMOCRATICI DI SINISTRA-L'ULIVO	DEMOCRATICO CRISTIANO	DEMOCRATICO CRISTIANO - PARTITO POPOLARE ITALIANO	...	PARTITO SOCIALISTA ITALIANO DI UNITA' PROLETARIA	PARTITO SOCIALISTA UNITARIO	POPOLO DELLA LIBERTA'
deputy														
ABBIATI	0	0	0	0	0	0	0	0	0	0	...	0	0	0
ABBIATI DOLORES	0	0	0	0	0	0	0	0	0	0	...	0	0	0
ABELLI	0	0	0	0	0	0	0	0	0	0	...	0	0	0
ABRIGNANI	0	0	0	0	0	0	0	0	0	0	...	0	0	0
ACCAME	0	0	0	0	0	0	0	0	0	0	...	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
ZOSO	0	0	0	0	0	0	0	0	0	0	...	0	0	0
ZUCCHINI	0	0	0	0	0	0	0	0	0	0	...	36	0	0
ZUCCONI	0	0	0	0	0	0	0	0	0	0	...	0	0	0
ZUECH	0	0	0	0	0	0	0	0	0	0	...	0	0	0
ZURLO	0	0	0	0	0	0	0	0	0	0	...	0	0	0

3666 rows × 56 columns

Fig. 5. Most speeches per party

The implications and final results of this analysis will be evaluated in the "Experimental results" section.
In the next section we will discuss the algorithms we employed for performing gender classification.

2.3 Support Vector Machine

SVM is a learning algorithm for linear classifiers which, once fixed a linearly separable training set $(\boldsymbol{x}_1, y_1), ..., (\boldsymbol{x}_m, y_m) \in \mathbb{R}^d \times \{-1, +1\}$, generates the linear classifier corresponding with the single solution $\boldsymbol{w}^* \in \mathbb{R}^d$ of the following optimization convex problem with linear constraints:

$$\min_{w \in \mathbb{R}^d} \frac{1}{2} \|\boldsymbol{w}\|^2 \quad \text{s.t. } y_t \boldsymbol{w}^T \boldsymbol{x}_t \geq 1, t = 1, ..., m \tag{1}$$

where $\boldsymbol{w}^*$ geometrically represents the separating hyperplane with maximum margin. The *support vectors* are those $\boldsymbol{x}_t$ on which $\boldsymbol{w}^*$ has margin equal to 1, that is $y_t(\boldsymbol{w}^*)^T \boldsymbol{x}_t = 1$.

In the non linearly separable case, the SVM problem becomes:

$$\min_{w \in \mathbb{R}^d} \frac{\lambda}{2} \|\boldsymbol{w}\|^2 + \frac{1}{m} \sum_{t=1}^{m} h_t(\boldsymbol{w}) \tag{2}$$

where $h_t(\boldsymbol{w}) = \left[1 - y_t \boldsymbol{w}^T \boldsymbol{x}_t\right]_+$ corresponds to the hinge loss. A regularization parameter λ is introduced in order to balance the two components. Also in this case, as for the linearly separable case, the solution $\boldsymbol{w}^*$ belongs to the subspace of the linear combinations of example of training set multiplied by their labels. In order to solve the above minimization problem, we can apply OGD, that is the online version of the Stochastic Gradient Descent.

2.4 Convolutional Neural Network

A Convolutional Neural Network is a deep learning algorithm that has recently gained a lot of attention because of its suitability into many fields because the pre-processing required in a CNN is much lower as compared to other classification algorithms; in fact, it provides a reduction in the number of parameters involved and the re-usability of weights.

The structure of a CNN consists of stacking convolutional and pooling layers with dense ones. In the convolutional layer, a convolution multiplies a matrix of features with a filter matrix (or *kernel*) and sums up the multiplication values. Then the convolution slides over to the next feature and repeats the same process until all the features have been covered.

The stride represents the number of feature shifts over the input matrix. Padding is needed when the filter does not perfectly fit the feature matrix. We have two possibilities: we may add zeros to the features such that the filter will fit, or we may drop the features where the filter did not fit (*valid padding*). The *pooling layer* decreases the computational power required to process the data by reducing the number of parameters when there are too many features. Although there are different kinds of pooling, we used *max pooling*, that takes the largest element from the rectified feature map.

Flattening simply converts the last convolutional layer into a one-dimensional layer that we can then connect to dense one(s). We may repeat this process by stacking by stacking convolutional and pooling layers.

At the end, the flattened output is fed to one or more fully-connected layer(s). In the last layer the network handles classification and outputs the predicted class since the last fully connected layer computes the probability that the input belongs to each class by exploiting an activation function (the softmax in our case) that provides the probability distribution.

3 Experimental results

First of all, we can consider the final complete information about what we have evaluated so far and comment on the results:

legislatura	female	male
genere		
1	7.352941	92.647059
5	2.731411	97.268589
6	3.945372	96.054628
7	8.294931	91.705069
17	30.700447	69.299553
18	35.869565	64.130435

Fig. 6. Final gender distribution per legislature (in percentage points)

legislature	female	male
gender		
1	1.333413	98.666587
5	2.600678	97.399322
6	3.357497	96.642503
7	12.261359	87.738641
17	22.086731	77.913269
18	30.524723	69.475277

Fig. 7. Final gender distribution per speeches per legislature (in percentage points)

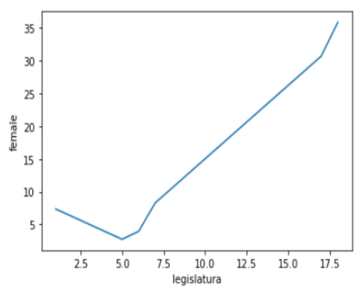

Fig. 8. Graph of final gender distribution per legislature (in percentage points)

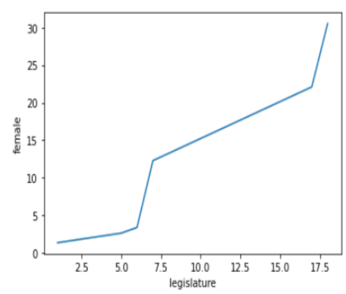

Fig. 9. Graph of final gender distribution per speeches per legislature (in percentage points)

From fig. 6 we can firmly state that there has been a positive trend regarding the female presence in the Italian parliament during the last 70 years. Starting from a 7% in the I legislature, the percentage has currently increased to 35%; this is an outstanding result and it shows how women have been involved in politics at a constant ever-increasing rate. Their higher presence can certainly be explained by social factors: during the V Legislature ('68-'72) Italy was settling after the economic boom and it has experienced social and political unrest; it is therefore understandable that the female presence in politics was at its lowest, since women were protesting in those very same days for what is today known as the period of "feminism", mainly concerning topics like divorce and abortion. However, we can see that in just 4 years it tripled, showing the results of the previous battles and their ambitions. In particular, it is worth to note that in the current legislature female representation is just a little more than half of the male one.

However, if we look at fig. 7 and consider female representation by the percentages of the speeches they have made, results may look a bit different: also in this case we see a positive trend with an ever-increasing percentage of female interventions during time, but the numbers are globally lower than their only-gender counterpart. By this result we can state that even though women are physically present in the Parliament, they are not so actively participating in the political life (in this case evaluated only by the number of speeches they made during official assemblies). Indeed, for what concerns the current legislature we see that only the 30% of speeches in the Chamber of Deputies are made by women, while more than the double of them are made by their opposite gender. This can be interpreted as somewhat difficult for women to be vocal during these assemblies and perhaps make a difference (however, we are not assessing the implications of policies discussed during these assemblies).

For what concerns the differences in the political forces' evaluation, we considered four graphs:

In the first one we can see the female representation per party with respect to the deputies (both males and females) of all legislatures in the Chamber of Deputies. Each line represents a party; when one party hits bottom it is generally because it disbands from that legislature on, or vice versa it did not exist until a certain year. In this case we are evaluating the gender over the total of legislatures: for what concerns the coalition that is currently in office, we see that 15% of the deputies in the current legislature are women belonging to the M5S, the 6% of them to the PD and lastly the 0.6% to the LeU. There are parties that are actually more representative than the last one (LN, FI, IV and others). Looking at the past, it is quite interesting to take a glimpse at the same percentages for the PCI during the V, VI and VII Legislatures: the female composition of the entire Chamber increases steadily from 1.3% up to 5.6%.

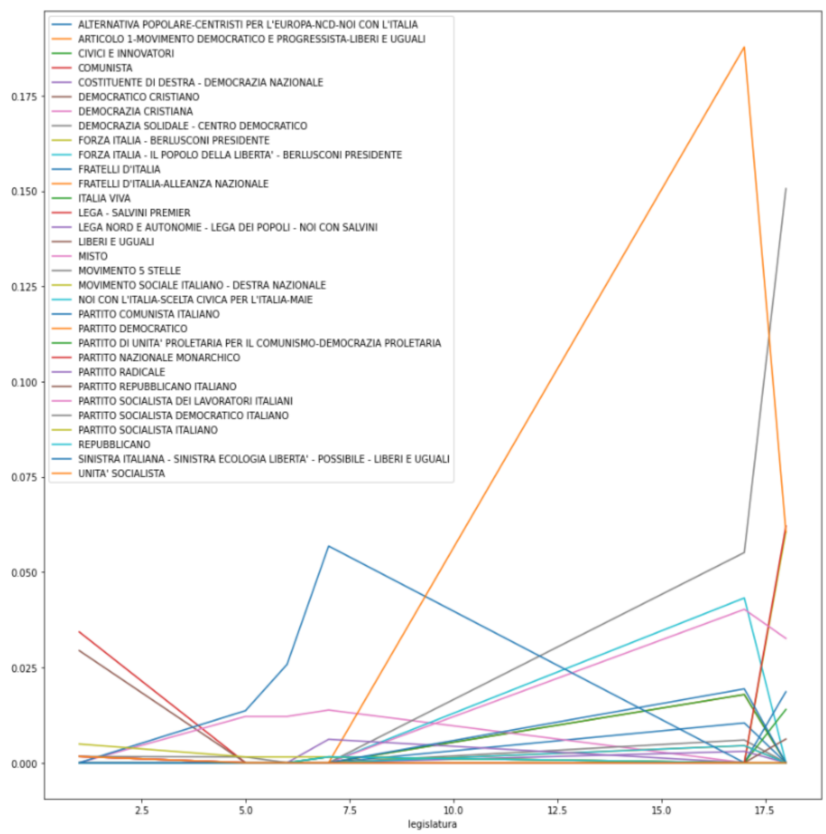

Fig. 10. Number of fem per party / tot dept per legis

The second graph displays the number of females per party over the deputies of the party per legislature. Also in this case we can firstly focus on the female composition of the government in office: it is the coalition formed by Movimento 5 Stelle, PD and LeU; the female representation with respect to the current legislature is 43%, 33% and 28%, accordingly. From this we see that the leading party (M5S) is highly inclusive. It can also be noted that with respect to the previous legislature, both M5S and LeU have increased their female representation in the party, while the PD is the only one for which it actually decreased (from 37% to the current 33%). This decrease can, at least in part, be explained by the internal frictions that the party experienced after the failed Referendum of 2016. We know from local newspapers that some members have left the party and established a new one (IV) or joined other ones. On the other side, if we want to make a polarized analysis and see a more general picture, we can say that the parties that place themselves to the right-wing all exceed the threshold of 30% of female representation (in the current legislature), while those on the left-wing can be placed under this arbitrary threshold. This can be a challenging point for a further discussion: are right-wing parties somehow a trend nowadays (also given the Italian political landscape, for which we know that left-wing parties are experiencing a deep crisis from quite some years now) and for this reason they attract more women too; or is it just a temporary phase and things will stabilize more in the next years?
Lastly, we can also take a look at the numbers from the past: one outstanding result is the female composition of the Communist party in the very first Italian legislature with 15% of female representation. Instead, skipping to the legislature "in the middle", it is worth highlighting the numbers of DC and PCI; the last one, in particular, has grown from an initial 4.9% up to 16.5%. As we explained before, this may be the result of the socialist and communist social movements that originated in those year and led to an ever-increasing presence of women in the political sphere.

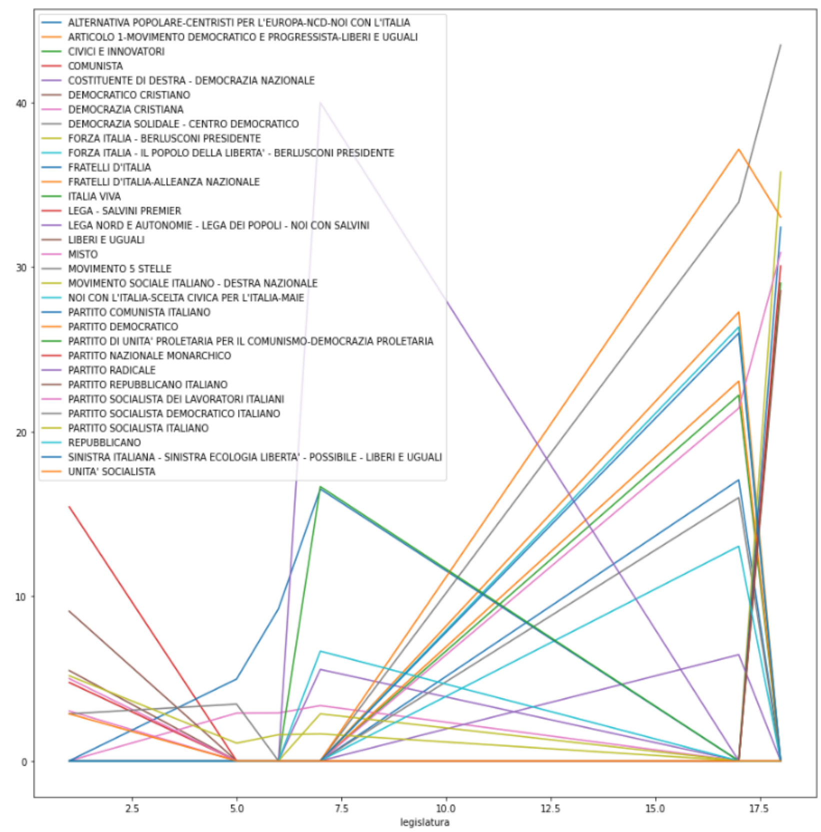

Fig. 11. Number of fem per party / tot dept per party per legis

The third graph shows the number of speeches made by females per party over the total amount of speeches of the Chamber per legislature. In this case we want to understand how much women of a given party talk with respect to the speeches of all members of the Chamber in a given legislature. Let's consider again the composition of the government in office: 7.8% of the speeches in the current legislature have been made by the women of the M5S party, 8.5% for PD and no speeches displayed by any member of the LeU party. This last result can be derived from three reasons: either there was no intervention by any member of the LeU deputies in the current legislature yet (therefore we discarded it as a whole), or because someone who actually belonged to this party switched it before and made more speeches in the "new" party (therefore the speeches she/he made are not displayed here anymore); or lastly, there has been an error in the reading of the OCR and unfortunately we don't get any meaningful result. For what concerns the number of PD, this higher value than the leading party (M5S) can be corroborated by empirical evidence: since the party has elected his new Secretary Nicola Zingaretti, they have intensively worked alongside the leading party and supported each other in many occasions. This is a good indicator, since it means that the coalition sustains frequent debates and there is no party prevailing on the other one inside it. For what concerns the parties at the Opposition instead (FI, FdI, LN), if we sum their interventions in the current legislature we get a percentage of 10.8% of total speeches in the legislature, against 16.3% of the governing coalition. Also in this case we can see that the Opposition is very vocal during assemblies and constantly interact in debates.

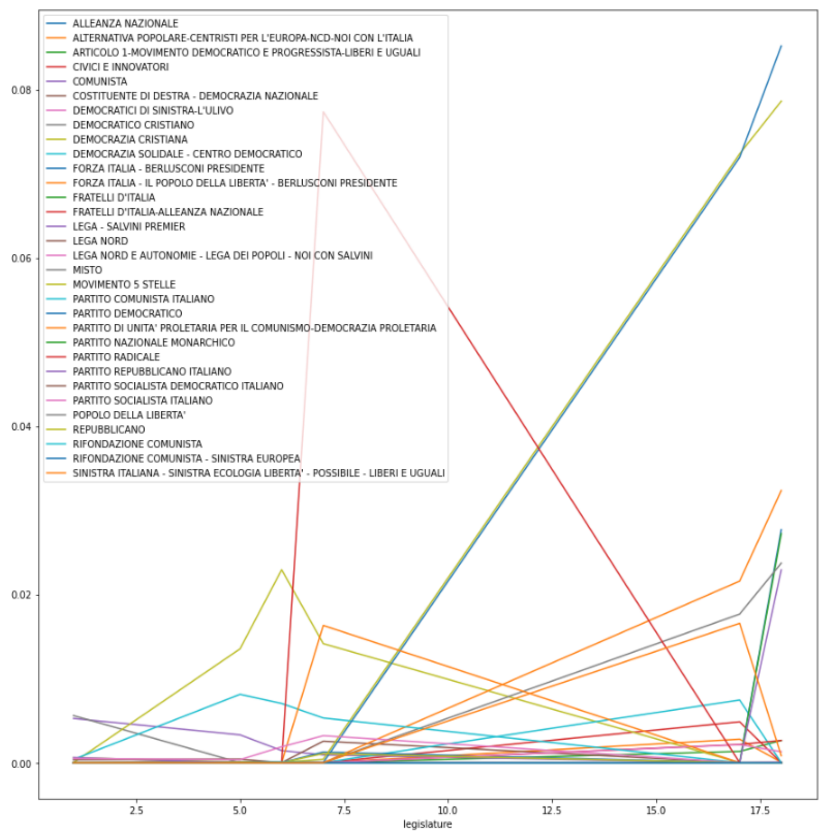

Fig. 12. Number of fem speeches per party / tot speeches per legis

Lastly, in the fourth picture we can see the number of speeches made by females for a given party over the total number of speeches of that party for a given legislature. We want to evaluate how much women talk with respect to their male counterpart in their own party. We want to remind that also in this case some results will not make sense and they need to be critically assessed because of errors deriving from OCR, mainly. Moreover, also in this case we do not see LeU displayed for the same reasons aforementioned. Starting from the governing coalition again, we can distinguish between: 44% of M5S speeches inside the party are made by women and 32% for PD. Note here that the percentage of both M5S and PD is perfectly consistent with the result 2, in which the percentage of female representation inside the M5S party is exactly 43% and for the PD 33%: this means that women in the governing political coalition in Italy are perfectly represented by their interventions during assemblies. This is a robust result in the sense that from this we derive that women are both physically and verbally present in the Italian governing coalition. Lastly, if we want to make a comparison with respect to the Opposition coalition, the total amount of speeches done by women in the right-wing parties is actually on the same level as the total amount of speeches done by women in the one in-office. Therefore, also the women belonging to Opposition are well-represented during the political debates inside their own parties.

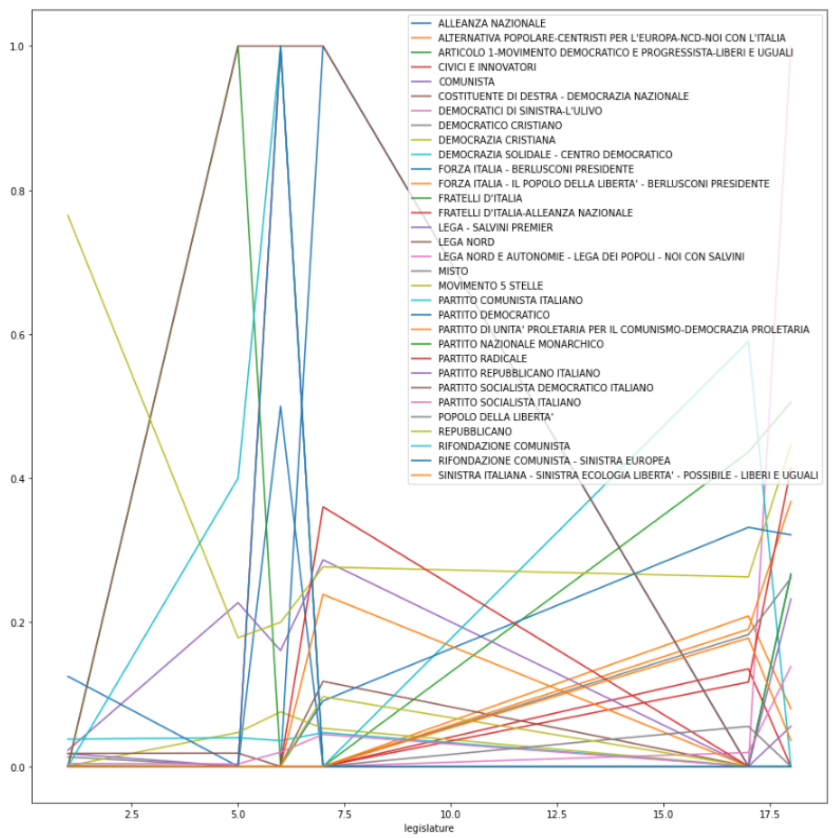

Fig. 13. Number of speeches of fem per party / tot speeches per party per legis

The last part of our analysis focuses on gender classification through the use of two algorithms: the SVM and the CNN.
The experiments were run on a free Google Colab instance running Ubuntu that provided a dual-core Intel Xeon E5-2650 v3 that run at 2,30 Mhz and 12GB of RAM. Moreover, it provided also a NVIDIA Tesla K80 as GPU that we used to train the Neural Networks using keras with Tensorflow as backend. We set a random seed to ensure the reproducibility of the results.
For what concerns the SVM, in particular, we used the `tf-idf` document matrix representation. Moreover, we can directly look at the confusion matrix in order to evaluate the results:

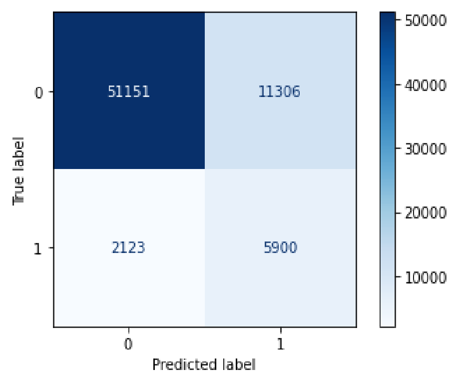

Fig. 14. Confusion matrix for SVM

Let's say that we consider the case in which positive are speeches made by females, namely label 1: the dark blue square represents the number of speeches performed by males that are correctly predicted (51151) by the algorithm (namely, the true negative). Diametrically opposite we find the true positive, namely the number of correct predictions of speeches made by females (5900). 2123 are the speeches that are truly made by women, but the SVM predicted as made by males, therefore they are the false negative (also called error of second degree). By elimination, 11306 are the false positive: the speeches made by men, but which the algorithm mistakenly predicted as made by women.
We also report the metrics used to evaluate the results: we obtained an accuracy of 80%, precision of 34%, recall of 73% and finally the `f1_score` 46%. The large difference between the accuracy and the `f1_score` is justified by the large unbalancement in the labels.

The CNN model is constructed as follows: we used the word embedding technique on our speeches in order to feed them to the network. Because of lack of memory, we first needed to train the network on a sample of 1500000 speeches and we also set the maximum number of features selected to 200 and not equal to the

actual longest sentence length.
The model starts with a first embedding layer that embeds padded speeches to 50 dimensions and then we apply a dropout of 20% for regularization purposes. This is followed by two stacked `Conv1D` layers with ReLU activation, 200 filters each, a kernel size of 3 and padding set to "same". After, we apply `GlobalMaxPooling` and flatten the result, so that we are able to connect it to the output dense layer that has a sigmoid activation function.
We trained the model for 10 epochs using `binary_crossentropy` as loss and `adam` as optimizer. We also implemented `early stopping` in order to prevent overfitting.
After 6 epochs, the final results are far more satisfying than the SVM: both evaluated on the test set, we reach an accuracy of 90% and a `f1_score` of 90%, too. However, also in this case the problem of the unbalancement in data persist, but we also need to face the problem of dealing with a very large data set. In order to fix the first problem, one solution could be to look at the `f1_score` instead of accuracy; for the second one, one solution could be using a sample of the data set or finding another document feature representation, in order to feed the neural network by converting only the batch each time, and not the whole matrix inside. This last problem is left to be developed for a further discussion.

4 Concluding remarks

This project is intended to provide a picture of the Italian political sphere by specifically addressing the gender issue in the Chamber of Deputies. Some of the results were unexpected and proved that Italy is not so far behind other countries that are worldwide known for their female representation in politics. In particular, the evolution that female inclusion has experienced over time for many parties is evident. Special attention has also been devoted to the polarization effects that may emerge when dealing with this kind of analysis.

There are still some issues that need to be addressed: we did not make any manual adjustment to the data set in order to clean it from the OCR errors. This is a thorny problem that unfortunately outputs some level of errors in our work.
The second one is the inclusion of all the data set and not sampling it, in order to feed the CNN with more datapoints.

Finally, in order to have a deeper knowledge of the subject and gain more insights on this project, topic modelling techniques could be employed and dig into themes that women address during the assemblies hold in the Chamber of Deputies compared to those issues addressed by men. This type of work could be beneficial in order to develop also a deep policy analysis of the Italian Parliament and see its performance through the gender indicator.

References

1. Papavero, L., & Zucchini, F. (2018). Gender and party cohesion in the Italian parliament: A spatial analysis. Italian Political Science Review/Rivista Italiana Di Scienza Politica, 48(2), 243-264. doi:10.1017/ipo.2017.
2. EU Parliament, Women in politics in the EU.
3. Burger, J. D., Henderson, J., Kim, G., & Zarrella, G. (2011, July). Discriminating gender on Twitter. In Proceedings of the conference on empirical methods in natural language processing (pp. 1301-1309). Association for Computational Linguistics.
4. Cheng, N., Chandramouli, R., & Subbalakshmi, K. P. (2011). Author gender identification from text. Digital Investigation, 8(1), 78-88
5. Neural networks and deep learning - C. C. Aggarwa