Predicting Twitter Users' Political Orientation: An Application to the Italian Political Scenario

Matteo Cardaioli
Department of Mathematics
University of Padova, GFT Italy
Padova, Italy
matteo.cardaioli@phd.unipd.it

Pallavi Kaliyar
Department of Mathematics
University of Padova
Padova, Italy
pallavi.kaliyar@phd.unipd.it

Pasquale Capuozzo
Department of General Psychology
University of Padova
Padova, Italy
pasquale.capuozzo@phd.unipd.it

Mauro Conti
Department of Mathematics
University of Padova
Padova, Italy
conti@math.unipd.it

Giuseppe Sartori
Department of General Psychology
University of Padova
Padova, Italy
giuseppe.sartori@unipd.it

Merylin Monaro
Department of General Psychology
University of Padova
Padova, Italy
merylin.monaro@unipd.it

2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) | 978-1-7281-1056-1/20/$31.00 ©2020 IEEE | DOI: 10.1109/ASONAM49781.2020.9381470

***Abstract*—Recently, the increasing spread of Online Social Networks (OSNs) provided an unprecedented opportunity of analysing online traces of human behaviour to get insight on individuals and society. Among the others, the possibility of predicting users' political orientation relying on data extracted from OSNs received growing attention. In this study, we introduce and make publicly available a dataset composed of 6.685 unique Twitter users and 9.593.055 Tweets. Differently from most of the dataset currently available in the literature, here, each user was manually labeled according to their political orientation by a pool of human judges, using strict inclusion criteria. Further, we address the feasibility of the automatic classification of Italian Twitter users' political orientation based on their Tweets content. Our analysis focuses first on implementing a series of classifiers with the aim of predicting users' political preference as right- or left-oriented. The built models were then evaluated for inferring the political orientation of those users supporting "Movimento 5 Stelle" (M5S), an Italian political party with a still unclear political leaning. Results show high performances on the left-right classification task, with accuracy rates up to 93%. Finally, classification performances obtained on M5S supporters and possible applications of our findings are discussed.**

***Index Terms*—Political Orientation Prediction, Natural Language Processing, Twitter Dataset, Italian Politics.**

I. Introduction

In recent years, the massive popularity of Online Social Networks (OSNs) provided the opportunity of studying digital records of human behaviour as a source of information about individuals and society [1]. Data acquired from social network platforms (e.g. textual posts, likes and profile pictures) showed to be suitable for inferring and predicting a variety of users characteristics [2], [3], rising also issues related to privacy and data ownership [4]. For instance, recent evidence shows that data gathered from OSNs can be successfully employed for predicting users' personal information like personality traits [5], sexual orientation [6] and ethnicity [7]. Among the personal information that can be deduced from OSNs data, the automatic prediction of users' political orientation received growing attention from scholars.

The possibility of predicting users' political orientation relies on the fact that the introduction of OSNs has changed not only the way people passively receive information but also how they express personal opinions about news and political events. Indeed, according to the Pew Internet and American Life Project [8], a substantial part of the US adults internet users (73%) used the web to get news or information about politics in 2010, with the 22% of them using OSNs for political purposes. Moreover, it's important to note that this trend can only be increased given the recent growth in political expenditures for online campaigns [9], with candidates spending more on digital advertising each election cycle [10]. Thus, the massive shift of political campaigns and deliberations on the web and the interactive nature of OSNs environment provided the unprecedented opportunity of analyzing online human behaviour to gather insights on the population political leaning.

The majority of studies in scientific research focused on OSNs data extracted from the US population [11], [12], [13], researchers from many other countries including Germany [14], Spain [15], France [16], United Kingdom [17] and Belgium [18] tried to face the same research question with encouraging results. The primary ambition of the above-cited studies was to assess the predictive power of data gathered from OSNs to forecast electoral results, a goal far to be free of limitations. That papers analysed Tweets corpora by applying mostly sentiment analysis, where the underlying assumption is that the audience sentiment extracted from politics-related tweets can be indicative of vote intentions [19], [20]. However, the fact that a given user shows a certain political orientation does not imply a vote during the general elections [21]. Moreover, some researcher highlighted how a random selection of OSNs users is not representative of the whole electorate [22], [23], lowering the predictive power of that kind of studies. Alternatively, a valid and equally ambitious direction is to predict users' political orientation restricting results to the considered sample, without comparing predictions against electoral results or election polls [24].

Another significant contribution due to the increasing attention received by OSNs data for predicting users' political orientation is represented by the introduction of several sets of data collected for studying this phenomenon. Previously introduced datasets differ from each other based on the country, the language, the data labeling procedure, and the time window in which data are collected. Furthermore, a crucial difference is the social network platform under investigation (e.g., Facebook, Twitter).

Although prior works introduced datasets extracted from various OSNs [25], [3], the majority of studies focused on on a specific social network platform, Twitter [26], [12], [14]. Twitter is a popular microblogging platform with around 330 million active users worldwide[1]. It allows users to share and express in real-time their opinions and tastes in the form of typewritten text, the so-called Tweets. The massive popularity of Twitter as a source of information for this kind of studies is given by the fact that most of the Twitter accounts are public and the privacy policies do not impose hard restrictions on data gathering. Thus, Twitter allows scholars to analyze its data without breaching any user's privacy. Another advantage of using Twitter is that it facilitates the data extraction process about its public accounts through its developer's web page and available APIs (Application Program Interface), which makes the process of collecting data much faster, legitimate, and more accessible than in other social network platforms. Another factor worth noting is that Twitter is heavily politicized [27], and its users are more likely to express their political views there than on any other platform.

A. Contribution

In this paper, we introduce a new manually-annotated dataset for predicting political orientation of Italian Twitter users, labeled by a pool of human judges. The proposed dataset is composed of 9.593.055 Tweets belonging to 6.685 unique Italian Twitter users labeled for their political orientation. Based on this data, we implement a set of models to predict the left-right political orientation of Twitter accounts. Finally, we investigate the political inclination of the supporters of M5S, which in 2018 general election resulted the biggest party in Italy, even if its collocation on the left-right line is unclear.

B. Organization

The organization of the rest of our paper is as follows. In Section II, we briefly described state-of-the-art and background work that has been done by some of the previous researchers on political orientation using Social Media. In Section III, we present and explain the methodologies we used to create our dataset, including a description Machine Learning methods. In Section V, we present the performance and results of our models. Finally, in Section VII, we conclude our work.

[1]https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/

II. Related Work

As described in Section I, the automatic prediction of users' political orientation from OSNs data is receiving growing attention from different scholars worldwide. This paragraph is aimed into providing an overview of the main findings obtained in previous works addressing Twitter users' political orientation prediction.

Prior studies can be divided into those relying on Twitter non-textual information [28], [29], [30], those analysing the content of Tweets [31], [7] and those assessing the effectiveness of a mixture of them [32], [33].

Regarding the employment of non-textual information, one of the first studies in this field focused on classifying the political orientation of the audience of media outlets such as Fox News and Washington Times as liberal or conservative to improve the users' experience [34]. After assigning to a sample of Twitter users a political orientation corresponding to the members of Congress they follow, the authors characterized each media outlet audience as more conservative or liberal with clearer correspondences for conservative audiences than liberal ones. Moreover, encouraging results were achieved by analysing non-textual information such as user-party interaction [30], the degree of Tweets and Retweets each user publishes regarding a given party account [28] and the similarity regarding ideological stances between users and certain political orientations [29].

Closer to our work are those studies considering the content of Tweets as a valuable source of information to predict users' political orientation. Following this approach, an interesting result was achieved by comparing the performance of classifiers relying on Tweets content with those based on network-level features [32]. The authors employed the TF-IDF computation, the relative frequency of hashtags and the Latent Semantic Analysis of hashtags as Tweets content features. Results showed quite high performances, with accuracy rates ranging from 79.2% to 90.8% in the classification of users' political preference as left or right oriented. Another significant contribution analysed the predictive power of both non-textual and content-based feature sets in classifying Twitter users as Republican or Democratic [33]. Referring to the content-based analysis, the combination of the five linguistic features investigated achieved a 77% accuracy rate, with Domain-specific Latent Dirichlet Allocation as the best performing singular feature (76% accuracy). Although the focus was not mainly on Tweets content, the evidence just reviewed seems to suggest a pertinent role of content-based analysis in predicting political orientation.

Among the studies relying mainly on linguistic, content-based analysis, two deserve mention. In an investigation aimed into measuring political homophily of Twitter users [31], the authors automatically classified a sample of Twitter users as Republican or Democrats by means of the TF-IDF measure. They obtained a quite high user-based classification performance, with an accuracy rate of 79%. Another significant contribution aimed at studying both political engagement and political leaning of Twitter users by means of content-based features [7]. The authors asked participants to self-declare their political ideology on a 7-point Likert scale going from "very conservative" to "very liberal". The prediction of the extremes political leanings was based on a set of linguistic features including unigrams, word clusters and emotions. Results exhibited a correlation between the predictions and self-reported ideologies of 0.37 when considering the whole feature set while the best performing singular feature was the Word2Vec clusters (correlation = 0.3).

In all the studies reviewed here, Twitter is recognised as the most studied social media platform for politics-related topics. As a consequence, previous works introduced different datasets for the automatic prediction of political preferences relying on Twitter data. Since most of the studies focused on forecasting electoral results, several datasets are composed of a collection of Tweets mentioning candidates name [35], specific political parties [17] or hashtags associated to electoral campaigns [36]. Nonetheless, datasets compiled following that methodology are not suitable for user-based political orientation prediction because the data cannot be linked to a specific user. Among the studies employing a dataset suitable for user-level classification, only a few made them publicly available for research purposes [7]. Furthermore, to the best of our knowledge, none of them focuses on the Italian political scenario.

III. Dataset

In this section, we present the methodology we have followed for the creation of a huge dataset from Italian Twitter accounts. Our work mainly consists of two phases: (i) collecting legitimate and labeled Italian Twitter user accounts with some political influences to use as the ground truth, and (ii) downloading Tweets using Twitter APIs.

Currently, the Italian political landscape consists of five main parties. Three parties are considered center-right parties: Forza Italia (FI), Fratelli d'Italia (FDI), and Lega. One party, Partito Democratico (PD), is considered as a left-wing party. Finally, Movimento 5 Stelle (M5S) represents the third pole. The collocation of M5S is not clear yet. It derives from the fact that the M5S made government alliances with both the right-wing and the left-wing parties since 2018. Moreover, M5S does not show a clear position on some political issues. This uncertainty is causing a loss in preference for M5S, moving its electorate to the right and left-wing parties. Finally, the Italian parliament includes many minor parties, especially on the left-wing, which we do not consider in this study as they represent less than 5% of the Italian population each (see Figure 1).

For our dataset, we started with collecting legitimate (non bot) Twitter user accounts. We labeled Twitter accounts of Italian users by their political orientation for the aforementioned parties. We applied two labeling methodologies. The first methodology consisted of scrolling the official Twitter account of each party and identifying, between the followers of the account, which are the users who expressed their political orientation by retweeting and commenting on the posts created by political parties Twitter accounts. One-thousand users account for each party (FDI, Lega, FI, M5S, PD) were collected manually using this methodology. Then, each account was analyzed by four to five independent judges who evaluated, according to the user's profile contents which party the user belonged. In the second methodology, we randomly selected the Italian user accounts from Twitter. Again, four to five independent judges analyzed the contents of the user's profile and classified it as supporter of one of the five parties or as "not classifiable" when there were not consistent cues of political orientation. Using the second methodology, three-thousand accounts were collected.

The labeling of the accounts was limited to the Italian Twitter user accounts. Moreover, all the bots were discharged. We set few conditions to identify legitimate accounts and also to remove Twitter bot accounts, as it is very common to find bot Twitter accounts which are prone to some specific political parties over social media[2]. The conditions we set to remove bots are (i) the Twitter user has at least 20 Tweets which can be political and non-political, (ii) the user must be an Italian national and his/her Tweets can be written in Italian or in English language, (iii) we also checked the activeness of the user account with respect to how frequent the user tweets, retweets, and comment (e.g., only the active users in the last six month).

Finally, we obtained 6.685 unique Italian Twitter user labeled for their political orientation. They were organized in a file reporting: (i) Numeric anonymized ID (ii) User Gender (Male, Female, and Undefined), (iii) the opinion of the five independent judges, and (iv) the data collection methodology we used to label the accounts.

We used the Twitter Application Program Interface (APIs)[3] to download the Tweets of the labeled accounts. For each Tweet, we stored: (i) Tweet ID, (ii) time of creation, (iii) total number of likes, (iv) number of retweets, (v) the source (e.g., Android, iOS, web), and (vi) the full text of the Tweet. The full dataset consists of 9.593.055 Tweets. The dataset is available at https://spritz.math.unipd.it/projects/politicalorientation/.

IV. Methods

This section introduces the techniques we used for data pre-processing and feature extraction. Moreover, we describe the machine learning classifiers we used to model our data.

A. Data Pre-processing

As the text obtained from Tweets contains a lot of noise and cannot be directly used as input for model training, a data pre-processing is needed. The cleaning and pre-processing we applied on the data frames includes the following steps:

- Removing urls (uniform resource locator). We removed all the urls from the tweets as urls don't have any predicting significance.

[2]https://spectrum.ieee.org/computing/software/how-political-campaigns-weaponize-social-media-bots

[3]https://developer.twitter.com/en

Fig. 1: Distribution of seats in the Italian Parliament after the 2018 general elections. Left: Chamber of Deputies. Right: Senate

- Removing punctuation. We removed characters such as "?", "!", ";" from the tweets.
- Upcase/Downcase of Letters. All the words, for example Lega and lega must to have the same predicting power, so for this reason we converted all the words in downcase format.
- Removing emoji's. We removed all the emoji's from the gathered text, as it is not possible to analyze them.
- Removal of Italian and English stop words (e.g.,"the", "what").

B. Feature Extraction and Learning Algorithms

In order to use a Machine Learning algorithm, we need a way to represent the text in numerically, identifying characteristics and relations between the words. For our work, we used the TF-IDF (Term Frequency – Inverse Document Frequency) [37] approach to represent the Tweets.

TF-IDF is one of the most popular term-weighting approaches. It's a numerical statistic that highlights the importance of each word in a document belonging to a larger corpus of documents. The TF-IDF value increases proportionally and determines how important a word is by looking at how frequently it appears in the document. It consists of two main components, which are: (i) Term Frequency measures the local importance of the word (if a word appears a lot of times, then the word must be important), and (ii) Inverse Document Frequency used to measure how much information the word provides (it calculate the weight of rare words across all documents). We used TF-IDF features to train the classifiers, firstly reducing the dimensionality by applying Truncated Singular Value Decomposition (SVD) [38]. To infer the political orientation of the labeled accounts and to predict the tendency of M5S supporters, we decided to test five different classifiers.

- Non-linear SVM with radial basis function kernel (RBF).
- Logistic Regression, with both L1 and L2 regularization.
- Stochastic Gradient Descent (SGD), with both L1 and L2 regularization.
- Random Forest.
- eXtreme Gradient Boosting (XGBoost) with binary logistic objective.

V. Experiments

As the Italian political landscape is very complex compared to that of other countries, our analysis was focused on the classification of each user as right-wing supporter or left-wing supporter. For this reason, the supporters of the three right-wing parties have been grouped into one unique class (right-wing supporters). The second class (left-wing supporters) was made up just of PD supporters. We considered 2.177.897 Tweets collected in a 6 month time window (June 1 to December 1, 2019). Only the accounts that reached 75% of agreement between the five judges were considered for the analysis. Out of the 2.945 accounts that fitted these constraints, 1.120 were labeled as left-wing, 1.825 as right-wing, and 801 as M5S.

To further investigate the political inclination of M5S supporters, whose position is currently unclear in a traditional left / right bipolar scenario, we divided our analyses into two steps. In the first step, we evaluate different ML models (see Section IV-B) for the classification of left-wing and right-wing Twitter accounts. In the second step, we leverage these models to infer the political tendency of M5S supporters.

1) Left-Right wing Prediction: To predict the political orientation of left-wing and right-wing supporters, we balanced the two classes by randomly undersampling the right labeled accounts, resulting in 1.120 Twitter accounts labeled as right and 1.120 accounts labeled as left-wing. We splitted the balanced dataset in training and test set with the proportion 80/20. We extracted TF-ID features as described in Section IV-B, applying a sublinear scaling.

2) M5S Tendency: The aim of the second step is to assess the political tendency of the M5S electorate. In particular, we want to predict if a Twitter account labeled as M5S, manifests an attitude closer to the left-wing or to the right-wing of the Italian political landscape. As for Subsection V-1, we undersampled the right labeled accounts, using all the 2240 accounts labeled as left-wing or right-wing as training set, while the test set consisted in the 801 accounts labeled as M5S. We extracted the features as described in Section IV-B, applying a sublinear scaling for TF-IDF.

In both steps, we performed a feature reduction with SVD [38], varing the dimensionality of output data in $\{2, 5, 10, 20, 30, 40, 50, 70, 100\}$. We selected a-priori a set of hyperparameters to perform the model validation using grid search with 5-fold cross-validation. In particular, for the SVM we considered C $\{1, 2, 5, 10, 20, 50, 100, 200\}$, and γ $\{2^{-5}, 2^{-4}, \ldots, 2^{0}\}$. For Logistic Regression classifier we varied C in $\{2^3,\ 2^4, \ldots, 2^7\}$. The α value of SGD was evaluated in $\{10^{-6}, 10^{-5}, \ldots, 10^{-2}\}$. For the Random Forest we varied the $n\_estimators$ in $\{400, 500, \ldots, 1000\}$, and the max_depth in $\{50, 80, 110, 140\}$. Finally, for XGBoost we considered

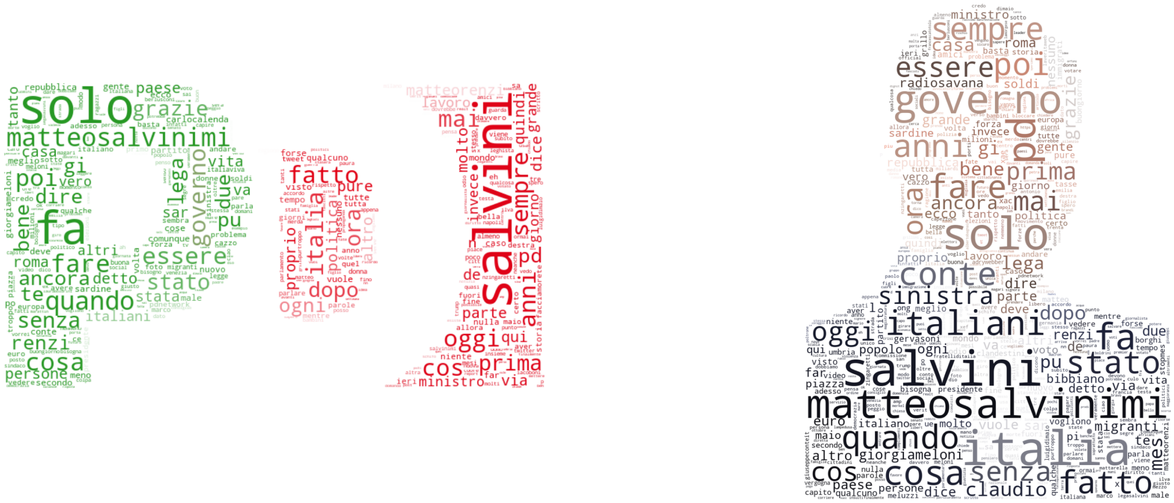

Fig. 2: Left: cloud of words of left-wing supporters. Right: cloud of words of right-wing supporters.

$n\_estimators$ in $\{2^7, 2^8, \ldots, 2^{11}\}$, and $learning\_rate$ in $\{0.01, 0.05, 0.1, 0.5\}$.

VI. Results

In this Section, we discuss the results obtained by our models for the settings described in Section V. First, we present an exploitative analysis based on the term frequency (TF) for left-wing and right-wing supporters (see Figure 2). For both left-wing and right-wing supporters, "salvini" (head of Lega) resulted the most frequent word, with 35.996 and 56.094 total appearances, respectively. Normalizing these values on the number of supporters of each wing, we find that left-wing supporters use more often the word "salvini" (i.e., 32 times per user) than the right-wing supporters (i.e., 30 times per user). Moreover, "pd" (the main left-wing party) is the sixth most common word between the right-wing supporters with 47.487 appearances, while for the left-wing supporters is the 26th with 13.440 appearances, even after "lega" (25th 13.474 appearances).

For the the prediction of left and right orientation (see SectionV-1), in Table I, we report the hyperparameters of our models and the number of components of SVD with the highest validation accuracy. Moreover, for each model, we report the results of accuracy, precision, recall, and F1-Score. Our results show that our models can predict with more than 90% of accuracy if a Twitter user supports left or right wing. Logistic Regression shows the best accuracy, achieving 93%. XGBoost is the worst classifier in our setting (i.e., 90% accuracy). SVM and SGD achieve the same accuracy (i.e., 92%), but SVM shows a higher discrepancy in precision and recall values compared to SGD.

For what concern the prediction of M5S inclination V-2, we report in Figure 3 the results of our models. Logistic Regression that is the best model for classifying left and right-wing accounts results to be the only classifier to predict a left tendency for the M5S supporters. Random Forest and XG-Boost, show comparable results, predicting M5S inclination between 62-56% for right-wing, and 38-44% for left-wing. SGD is the most polarized classifier (i.e., 76% right-wing and 24% left-wing), while SVM shows more uncertainty in prediction. Finally, we analyzed the agreement between the models in the prediction of M5S inclination. In particular, we have identified a scale of 6 intervals: strong left (all the models classified the account as left), left (4 on 5 models classified the account as left), weak left (3 on 5 models classified the account as left), weak right (3 on 5 models classified the account as right), right (4 on 5 models classified the account as right), and strong right (all the models classified the account as left). Figure 4 show the distribution of the M5S supporters based in relation to the models agreement. In particular, applying the same threshold used to label our dataset (i.e.,75%), 51% of M5S accounts show a right inclination, 32% a left inclination and 17% results not classifiable (i.e., weak left and weak right).

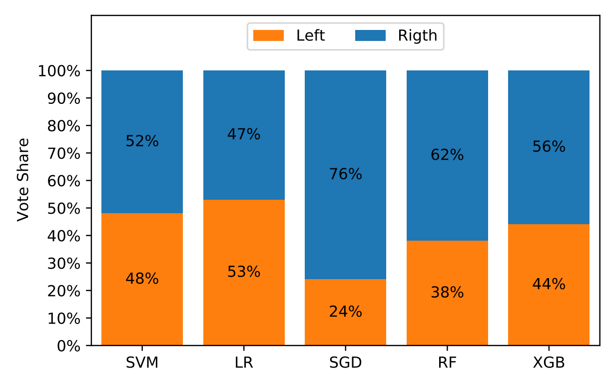

Fig. 3: Prediction of vote tendency for M5S.

TABLE I: Model, hyperparameters, number of components and performance on the classification of left and right-wing supporters.

Model	Hyperparam	SVD components	Accuracy	Label	Precision	Recall	F1-Score	Support
SVM	$C = 200$ $gamma = 0.5$	40	0.92	left right	0.90 0.95	0.96 0.89	0.92 0.92	224 224
Logistic Regression	$C = 128$ $penalty = L2$	70	0.93	left right	0.91 0.94	0.95 0.91	0.93 0.92	224 224
SGD	$alpha = 10^{-7}$ $penalty = L1$	40	0.92	left right	0.92 0.93	0.93 0.92	0.92 0.92	224 224
Random Forest	$max\_depth = 50$ $n\_estimators = 900$	10	0.91	left right	0.92 0.91	0.91 0.92	0.91 0.91	224 224
XGBoost	$learning\_rate = 0.5$ $n\_estimators = 128$	40	0.90	left right	0.92 0.89	0.88 0.92	0.90 0.90	224 224

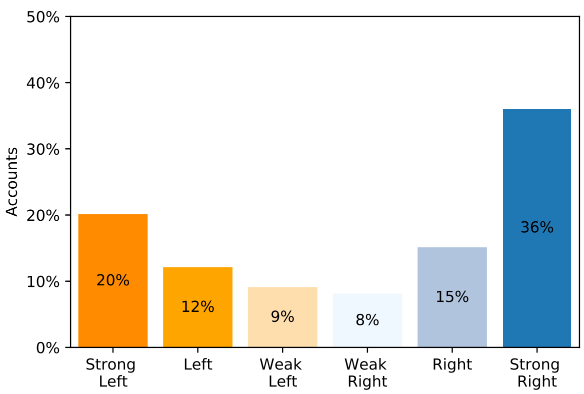

Fig. 4: Agreement of our models on M5S supporters.

VII. Discussion and Conclusion

In this paper, we confirm the results of previous work in literature that demonstrate the possibility to infer the user political orientation based on the social network contents. In particular, we reach an accuracy up to 93% in classify left and right-wing supporters, analyzing the textual content posted by the user on Twitter. Moreover, we investigate the possibility to detect the political inclination of the supporters of M5S (an Italian political party whose collocation on the left-right line is currently uncertain), training ML models on the right and left-wing supporters' data. Specifically, we trained the models on the text contents extracted from the Tweets posted by right and left-wing supporters on Twitter. Then, we tested our models on the Twitter contents of the M5S users. At the best of our knowledge, this is the first time that Natural Language Processing (NLP) and ML techniques are used to infer the inclination of the supporters of a third party in a bipolar scenario. This approach could be very useful in all the situations where the supporters of a third political force are called to express themselves about two representative left-right poles (e.g., during the second rounds of the elections). Further, in the Italian political landscape, the M5S party is now losing consent and it could be useful to predict whether this electorate will move mostly to the right or left-wing. Looking at the distribution of the classification of the M5S supporters, our results seem to be realistic for the current political scenario, as the recent analysis of the vote intentions shows an increment in the popularity of the right-wing parties[4], simultaneously with the drop of the M5S.

More interesting, the Term Frequency analysis shows that the most frequent words used by the left-wing supporters in their Tweets are referred to the right-wing parties (i.e., the most frequent word for the left-wing supporters is "salvini" who is the leader of the right-wing coalition). Similarly, right-wing supporters use the word "pd" (acronym for the largest left-wing party) much more often than left-wing supporters do (i.e., "pd" is sixth for the right-wing, while for the left-wing is outside to the top-20 most frequent words). This seems to confirm the idea that people are more used to criticize the ideas of the out-parties rather than discuss the ideas of their own party [39]. Moreover, this may reflect the fact that political leaders often build their election campaign on the criticism of the opposition rather than on their own ideas on the big political topics [39]. Finally, another interesting contribution of our work, is the strong ground truth on which we base our predictions. Indeed, the labeling process of the Twitter accounts in the dataset was performed by a pool of human judges with strictly inclusion criteria.

References

[1] S.-L. Shaw, M.-H. Tsou, and X. Ye, "Human dynamics in the mobile and big data era," *International Journal of Geographical Information Science*, vol. 30, no. 9, pp. 1687–1693, 2016.

[2] D. Gayo-Avello, P. T. Metaxas, E. Mustafaraj, M. Strohmaier, H. Schoen, and P. Gloor, "The power of prediction with social media," *Internet Research*, 2013.

[4] http://www.sondaggipoliticoelettorali.it/ListaSondaggi.aspx?st=SONDAGGI

[3] M. Kosinski, D. Stillwell, and T. Graepel, “Private traits and attributes are predictable from digital records of human behavior,” *Proceedings of the national academy of sciences*, vol. 110, no. 15, pp. 5802–5805, 2013.
[4] D. Butler, “Data sharing threatens privacy,” *Nature*, vol. 449, no. 7163, pp. 644–646, 2007.
[5] D. Azucar, D. Marengo, and M. Settanni, “Predicting the big 5 personality traits from digital footprints on social media: A meta-analysis,” *Personality and individual differences*, vol. 124, pp. 150–159, 2018.
[6] Y. Wang and M. Kosinski, “Deep neural networks are more accurate than humans at detecting sexual orientation from facial images.” *Journal of personality and social psychology*, vol. 114, no. 2, p. 246, 2018.
[7] D. Preoţiuc-Pietro and L. Ungar, “User-level race and ethnicity predictors from twitter text,” in *Proceedings of the 27th International Conference on Computational Linguistics*, 2018, pp. 1534–1545.
[8] A. Smith, “Twitter and social networking in the 2010 midterm elections,” *Pew Research*, 2011.
[9] C. B. Williams and G. J. Gulati, “Digital advertising expenditures in the 2016 presidential election,” *Social Science Computer Review*, vol. 36, no. 4, pp. 406–421, 2018.
[10] K. Kaye, “Data-driven targeting creates huge 2016 political ad shift: Broadcast tv down 20%, cable and digital way up,” *Advertising Age*, 2017.
[11] S.-I. Chiu and K.-W. Hsu, “Predicting political tendency of posts on facebook,” in *Proceedings of the 2018 7th International Conference on Software and Computer Applications*, 2018, pp. 110–114.
[12] J. Ramteke, S. Shah, D. Godhia, and A. Shaikh, “Election result prediction using twitter sentiment analysis,” in *2016 international conference on inventive computation technologies (ICICT)*, vol. 1. IEEE, 2016, pp. 1–5.
[13] F. M. F. Wong, C. W. Tan, S. Sen, and M. Chiang, “Quantifying political leaning from tweets, retweets, and retweeters,” *IEEE transactions on knowledge and data engineering*, vol. 28, no. 8, pp. 2158–2172, 2016.
[14] A. Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Welpe, “Predicting elections with twitter: What 140 characters reveal about political sentiment,” in *Fourth international AAAI conference on weblogs and social media*, 2010.
[15] J. M. Soler, F. Cuartero, and M. Roblizo, “Twitter as a tool for predicting elections results,” in *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. IEEE, 2012, pp. 1194–1200.
[16] A. Ceron, L. Curini, S. M. Iacus, and G. Porro, “Every tweet counts? how sentiment analysis of social media can improve our knowledge of citizens’ political preferences with an application to italy and france,” *New media & society*, vol. 16, no. 2, pp. 340–358, 2014.
[17] P. Burnap, R. Gibson, L. Sloan, R. Southern, and M. Williams, “140 characters to victory?: Using twitter to predict the uk 2015 general election,” *Electoral Studies*, vol. 41, pp. 230–233, 2016.
[18] S. Praet, P. Van Aelst, and D. Martens, “I like, therefore i am: predictive modeling to gain insights in political preference in a multi-party system,” 2018.
[19] A. Bermingham and A. Smeaton, “On using twitter to monitor political sentiment and predict election results,” in *Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology (SAAIP 2011)*, 2011, pp. 2–10.
[20] V. Lampos, “On voting intentions inference from twitter content: a case study on uk 2010 general election,” *arXiv preprint arXiv:1204.0423*, 2012.
[21] D. Gayo-Avello, “A meta-analysis of state-of-the-art electoral prediction from twitter data,” *Social Science Computer Review*, vol. 31, no. 6, pp. 649–679, 2013.
[22] D. Gayo-Avello, P. T. Metaxas, and E. Mustafaraj, “Limits of electoral predictions using twitter,” in *Fifth International AAAI Conference on Weblogs and Social Media*, 2011.
[23] P. T. Metaxas, E. Mustafaraj, and D. Gayo-Avello, “How (not) to predict elections,” in *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*. IEEE, 2011, pp. 165–171.
[24] P. Ceravolo and S. Guerretti, “Testing social network metrics for measuring electoral success in the italian municipal campaign of 2011,” in *2013 International Conference on Cloud and Green Computing*. IEEE, 2013, pp. 342–347.
[25] E. David, M. Zhitomirsky-Geffet, M. Koppel, and H. Uzan, “Utilizing facebook pages of the political parties to automatically predict the political orientation of facebook users,” *Online Information Review*, 2016.
[26] I. Alfina, D. Sigmawaty, F. Nurhidayati, and A. N. Hidayanto, “Utilizing hashtags for sentiment analysis of tweets in the political domain,” in *Proceedings of the 9th International Conference on Machine Learning and Computing*, 2017, pp. 43–47.
[27] A. Jungherr, “Twitter use in election campaigns: A systematic literature review,” *Journal of information technology & politics*, vol. 13, no. 1, pp. 72–91, 2016.
[28] A. Boutet, H. Kim, and E. Yoneki, “What’s in your tweets? i know who you supported in the uk 2010 general election,” in *Sixth International AAAI Conference on Weblogs and Social Media*, 2012.
[29] S. Gottipati, M. Qiu, L. Yang, F. Zhu, and J. Jiang, “Predicting user’s political party using ideological stances,” in *International Conference on Social Informatics*. Springer, 2013, pp. 177–191.
[30] A. Makazhanov, D. Rafiei, and M. Waqar, “Predicting political preference of twitter users,” *Social Network Analysis and Mining*, vol. 4, no. 1, p. 193, 2014.
[31] E. Colleoni, A. Rozza, and A. Arvidsson, “Echo chamber or public sphere? predicting political orientation and measuring political homophily in twitter using big data,” *Journal of communication*, vol. 64, no. 2, pp. 317–332, 2014.
[32] M. D. Conover, B. Gonçalves, J. Ratkiewicz, A. Flammini, and F. Menczer, “Predicting the political alignment of twitter users,” in *2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing*. IEEE, 2011, pp. 192–199.
[33] M. Pennacchiotti and A.-M. Popescu, “Democrats, republicans and starbucks afficionados: user classification in twitter,” in *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2011, pp. 430–438.
[34] J. Golbeck and D. Hansen, “Computing political preference among twitter followers,” in *Proceedings of the SIGCHI conference on human factors in computing systems*, 2011, pp. 1105–1108.
[35] J. E. Chung and E. Mustafaraj, “Can collective sentiment expressed on twitter predict political elections?” in *Twenty-Fifth AAAI Conference on Artificial Intelligence*, 2011.
[36] S. M. Mohammad, X. Zhu, S. Kiritchenko, and J. Martin, “Sentiment, emotion, purpose, and style in electoral tweets,” *Information Processing & Management*, vol. 51, no. 4, pp. 480–499, 2015.
[37] G. Salton, E. A. Fox, and H. Wu, “Extended boolean information retrieval,” *Communications of the ACM*, vol. 26, no. 11, pp. 1022–1036, 1983.
[38] P. C. Hansen, “The truncatedsvd as a method for regularization,” *BIT Numerical Mathematics*, vol. 27, no. 4, pp. 534–553, 1987.
[39] S. Klar, Y. Krupnikov, and J. B. Ryan, “Affective polarization or partisan disdain? untangling a dislike for the opposing party from a dislike of partisanship,” *Public Opinion Quarterly*, vol. 82, no. 2, pp. 379–390, 2018.