Measuring Populism in a Comparative Perspective using the Comparative Manifesto Project Coding Categories

George Makris, PhD Candidate, Aristotle University of Thessaloniki, Greece

Ioannis Andreadis, Professor, Aristotle University of Thessaloniki, Greece

Defining Populism

During the last decades there has been an intensification on the study of populism. Studies of the populist phenomenon take place both on policy level (e.g. rise of populist parties and movements, effects of those actors on the party system etc.) and on discourse level.

The dominant scientific definition in the field of populism is given by the "ideational approach" (Canovan 1999; Kaltwasser et al. 2017; Laclau 2005; Mudde and Kaltwasser 2017). According to this definition, the populist ideology considers society to be separated into two antagonistic groups: the people and the elite. More specifically, people is a homogeneous group that is morally "good" and "pure", and elite is a group that is "corrupted" and opposes the people placing their own interest over them. Also, according to the populist ideology, politics should be an expression of the people's general will.

Although, there is still a debate in the scientific community regarding a precise conceptual clarification of the term populism, it seems like the aforementioned definition is the most capable to make up a methodological starting point for empirical studies and research on the populist phenomenon (Mudde 2004; Mudde and Rovira Kaltwasser 2018). The reason for that is the fact that with this particular definition, we can facilitate quantification because it is an instrumental one (Aslanidis 2018). So it is easier to classify texts on the populism dimension as well as to detect populism, in general (Mudde and Rovira Kaltwasser 2018). Also, the ideational approach has provided the research community with the capacity to study populism both from the supply and the demand side, using various methods (Hawkins and Rovira Kaltwasser 2017).

However, the ideational approach for the study of populism doesn't take into account other ideologies which can be combined with populism and/or include it, such as socialism, nationalism etc. (Hawkins and Rovira Kaltwasser 2017). For that reason, populism has been characterized as a "thin-centered ideology", because many times it attaches to various ideologies. (Mudde, 2004). This is why there are several kinds of populist parties (e.g. left-wing, right-wing etc.), as well as populist discourses.

There are other approaches to the populist phenomenon as well, which are not discussed in this paper such as the political-strategic and the socio-cultural one (Mudde and Rovira Kaltwasser 2018). The first (Weyland 2017) argues that populism is an organized strategy of some charismatic leaders in order to rise to power and the second (Ostiguy 2017) argues that populism is a style of policy-making where the leader tries to break social "taboos", behaving "weirdly" and against the dominant way of political communication.

Populism and Automated Text Analysis

When it comes to populism and automated text analysis there is a growing literature which take advantage of the recent developments in the field of Natural Language Processing

(NLP) in order to measure populism on political parties using texts (Bonikowski and Gidron 2016; Cocco and Monechi 2021; Hawkins and Silva 2018; Pauwels 2011; Rooduijn and Pauwels 2011). Following the "ideational approach" definition of populism as mentioned earlier, some of them use dictionary methods in order to measure populism at parties and others use supervised machine learning in order to classify parties as populist or not, or scale them on the populist dimension (see Aslanidis 2018, for more on the literature of the use of text analysis on populism). Unfortunately, although the need for a large scale cross-national comparative analysis on populism through automated text analysis methods has been highlighted by previous research (Gidron and Bonikowski 2013) the only two studies existing until now are that of Cocco & Monechi, 2021, which concerns a very limited number of countries, and that of Hawkins & Silva, 2018.

So in this paper, we followed a large scale comparative approach to populism using automated text analysis methods on party manifesto data. More specifically, we developed a continuous measure of the political parties' level of populism through supervised machine learning using cross-national and longitudinal data from the Comparative Manifesto Project. The main advantages of our approach are four. First, by using computational methods to measure populism we offer a cost-efficient alternative to the more resource-intensive methods of measuring populism through human coding. Second, we can easily quantify the level of populism across a large number of countries and parties very quickly and also make our work replicable through the R code that we are going to share.

Third, we develop a continuous measure for populism which avoids a cruel dichotomization of the concept, because it has been shown that when dealing with populism as a dichotomous phenomenon there may be arbitrary classifications by experts or failure to grasp a "strategic" kind of populism. Fourth, and more important, our model was trained on a large scale and cross-national pooled dataset and so the paper is one of the few large scale comparative text analysis projects on populism (the only other being that of Hawkins and Silva 2018). A recent literature review showed that a large part of the existing scientific literature on populism is conflated by country-specific operationalizations and measurements (Hunger and Paxton 2021), so large scale comparative analyses are missing in the field.

Furthermore, our work capitalizes on the effort of (Cocco and Monechi 2021) and takes a step forward. Their paper was the first to use supervised machine learning on a big scale using party manifestos from 6 countries, and we included 26 countries. Also, although our models are similar, ours are trained on a pooled cross-national dataset, while Cocco and Monechi's models are trained on each country separately. This has a result that our model draws information from a common pool of data and not from a specific context only.

Also, our model is more computationally efficient because of the initial text representation which is the relative frequencies of the CMP coding categories. In that way, our representation summarises information into just the 56 CMP coding categories, making the algorithm more computationally efficient, in contradiction to the high-dimensional bag of words representation of Cocco and Monechi. Finally, our model is trained on continuous labeled data and not dichotomous as Cocco and Monechi models do.

Finally, we validate our measure with the "Chapell Hill Expert Survey", the "Political Parties Expert Survey" and the "V-Party" data, showing its high and statistically significant correlation with their populism items. Then, we show some pros and cons in its ability to grasp populism between left-wing and right-wing parties as well as its high correlation with the Euroscepticism phenomenon.

Data

Our data are party manifestos and we drew them from the Comparative Manifesto Project database (the most updated version). Our data set comprises of 486 election manifestos of 190 political parties. The number of countries included are 26, and are almost all of the EU countries and the United Kingdom. The only countries left out is Malta and Luxembourg, because CMP does not have a recent decade manifesto from those countries in its database. In Table 3 (appendix) we present the parties, countries and election years in our dataset.

CMP has a 56 coding category scheme (main categories) that it uses to code each sentence of the party manifestos (see appendix Table 4 for the categories description). Regarding populism, CMP does not include any categories measuring populism and its dimensions, like people-centrism and anti-elitism. For that reason we drew our dependent variable from the Populism and Political Parties Expert Survey (POPPA). The POPPA dataset contains a "populism" variable which is continuous and is derived as weighted arithmetic mean of the other populist items (see Table 5 appendix) included in the survey. It is measured on a scale from 0 (no populism) to 10 (high populism).

The reason why we chose the populism variable from the POPPA dataset is because it is the only measure of populism that is continuous. Other expert surveys like the PopuList are measuring the phenomenon as dichotomous. The reasons to rely on a continuous measure are two, as we hinted before. First, when populism is operationalized as dichotomous there may be arbitrary classifications by experts in "contentious cases", meaning in cases where parties are on a slight verge of being classified as populist or not due to their complex ideology or discourse. Second, several parties may use certain degrees of populism in a strategic manner (to gain electoral or other, support). In that case a continuous measure allows us to estimate the degree of populism a party may exhibit due to strategic -or other- reasons.(Meijers and Zaslove 2021).

The time period under consideration in the study is 2010 until 2021. The reason that we chose this time period is because the POPPA survey took place at 2018 and so we opt to include elections close to that time point. Making the assumption that populism of a given party is something that may change over a long period of time, and given that we don't have any expert measurements before and after 2018 (the next POPPA survey will take place in 2023) we chose a time period close to that point so that populism of parties is the same or almost the same, while simultaneously trying to include as many manifestos as possible in order for the machine learning algorithms to perform well.

Methods

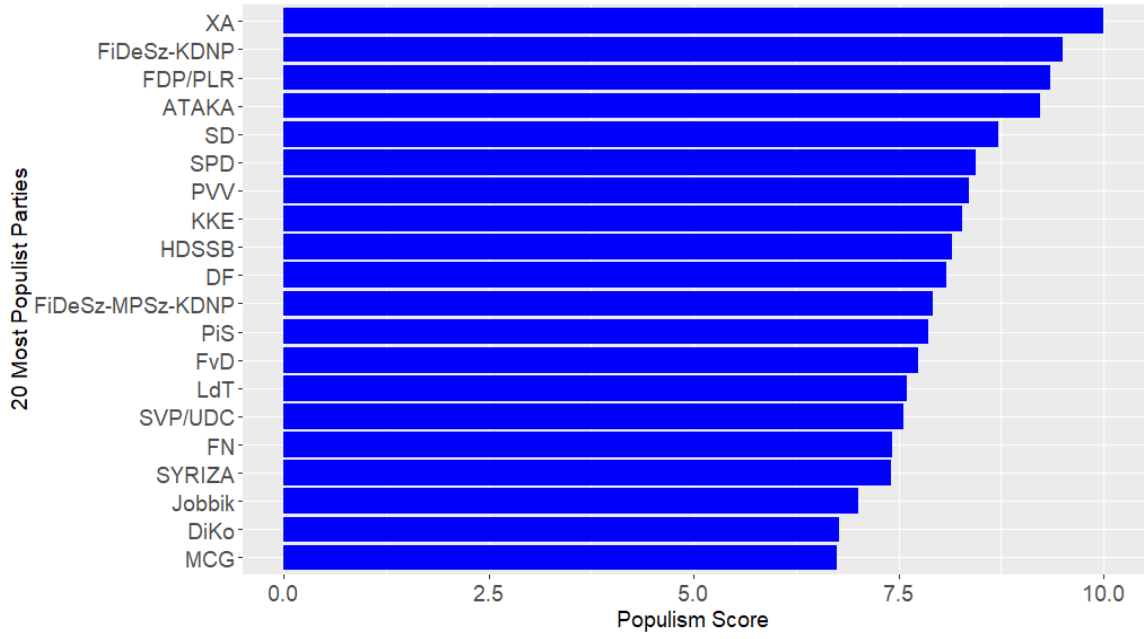
In order to derive the measure for populism we used all of the 56 CMP categories as independent variables and the POPPA variable for populism as a dependent one. Our initial

data representation was a document-term matrix. Each document is a manifesto and the terms are the coding categories. The values in the matrix are relative frequencies and, more specifically, the relative frequencies of each category in each document. Using that matrix, we trained several supervised machine learning models for a regression task, as long as our DV is continuous. More specifically, we trained 7 models on the 70% of dataset and tested them on the remaining 30%. Also, before running the models we preprocessed the data through centering and scaling in order to deal with potential outliers. We evaluated the models through the estimation of the Rooted Mean Squared Error (RMSE). In Table 6 in the appendix we present each model and its performance.
The final model that we reached was a gradient boosting machine model (GBM)[1], a tree-based model (Montgomery and Olivella 2018). The tuning parameters of the model (and of every model) were estimated through the use of 10-fold cross-validation in the training set. We present our finalized tuning parameters at Table 7 in the appendix. Finally, the predictions of the model for each manifesto constitute our populism measure. Then, of course, this measure can be aggregated on higher grouping levels like party level or country level.

Results

As we said above, our new measure is a continuous measure and is derived by a GBM model trained on a pooled dataset. This allows us to estimate the level of populism of each political party under consideration by aggregating the populist scores of its manifestos and also allows comparison of different parties. For example, in figure 1 we see the populist scores of the 20 most populist parties on the test set.

Figure 1: Populism Score of the 20 most populist parties on the test set.



[1] See appendix for a more detailed overview of the Tree Models and the Gradient Boosting Machine.

We remind that the populism score was calculated on the test set which is 30% of the data (meaning 144 manifestos). That being said, the parties presented above constitute only a random sample of the 190 parties, so they are the most populist parties of the test set. The main aim is to show that our ML-based measure produces very reasonable populist scores and can be used for cross-national comparison. For example, we see that some of the top populist parties are some right-wing parties like the Hungarian FiDeSz-KDNP electoral alliance of Orban and Semjen which has ruled Hungary since 2010, the far right-wing neo-nazi Greek party Golden Dawn (XA), the Dutch Party for Freedom (PVV) and the Czech right-wing populist party Freedom and Direct Democracy (SPD). However, we can also see that there are also left-wing populist parties are being scaled. Among them we can notice the left party SYRIZA and the far-left party KKE from Greece. SYRIZA has also been in governmental position as the ruling party of Greece from 2015 until 2019. Also, it is evident that we can also measure populism on centre parties like FDP/PLR and DiKo.

In Figure 2, below we see the most important variables (coding categories) that contributed to the predictions of our model. Tree-based models that combine multiple trees can be pretty difficult to interpret (as opposed to single trees which are directly interpretable) because of their increased complexity (which increases, of course, their predictive power). However, there is a measure called variable importance that can be calculated through summing up the overall decrease of the loss function[2] that a predictor introduces, averaged across all the trees. If this mean decrease index is high then it highlights an important predictor. However, the variable importance doesn't inform us about the direction of the effect of each predictor. That's why in the figure below we included the sign of each predictor which we estimated from a ridge regression model. So, in summary, in figure 2 the Variable Importance axis indicates the predictive importance of each predictor and the sign of each predictor indicates whether it predicts more (+) or less (-) populism.

So it is clear that two of the most important predictors of populism are those related to the European Union. A positive stance towards EU predicts less populist manifestos while a Eurosceptic stance predicts more populist ones. To that extent, it is evident that there is a certain amount of interplay between populism and Euroscepticism and we will showcase this relation later on the paper. Moreover, we see positivity towards national way of life as well as political corruption as two strong predictors of more populism. The first one may be attributed to the nativist attributes of many populist parties. The second is probably a remark of the populist parties' anti-elite element (one of the two core elements of populism, the other being people-centrism).

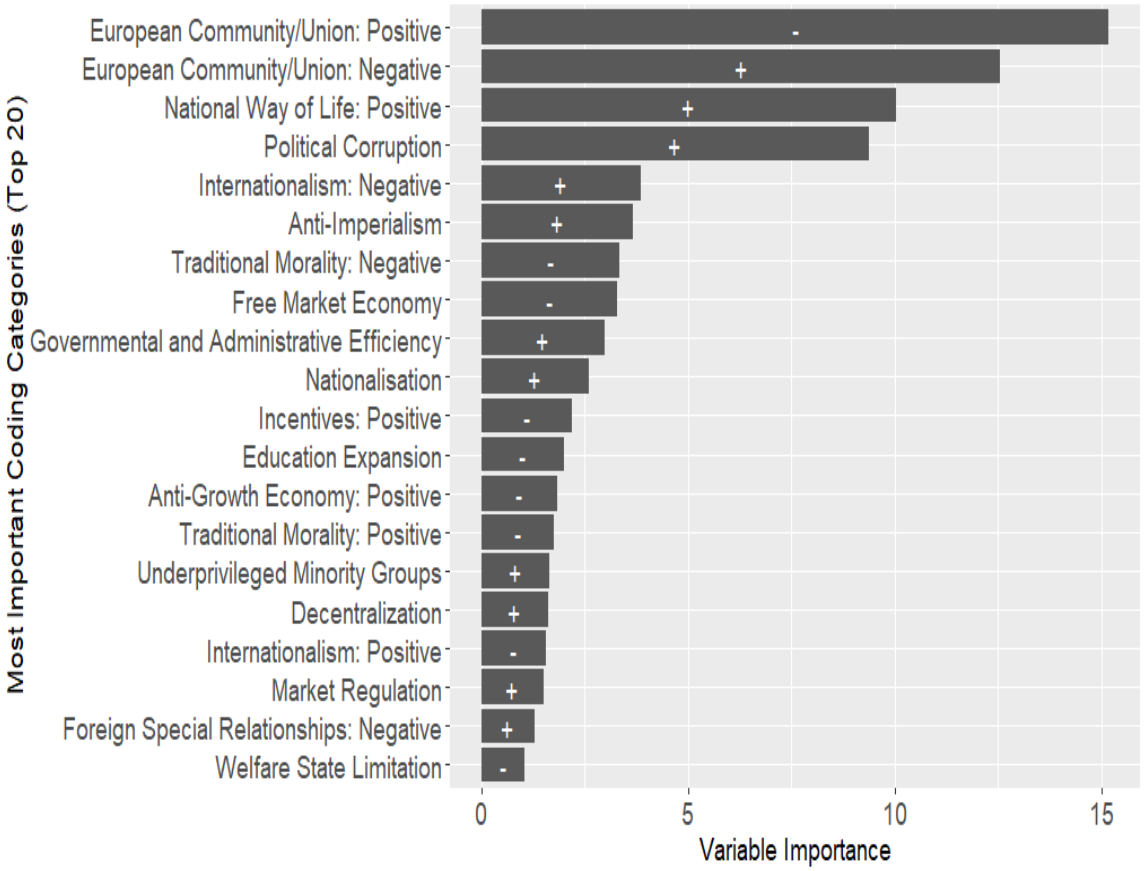
In general, we can see a trend of the populists to be more nation-centered, and following a more nationalistic approach regarding economy and society, while opposing internationalism, the EU, and the corrupt elites. This content of the manifestos aligns with the "ideational approach" operationalization of populism that we described earlier, which proposes a dual scheme (people vs elite), hinting that our measure perform well, although

[2] See appendix for a description of the loss function that the tree models minimize in the regression setting.

we will come back to this in the following chapter. However, the above characteristics may be indicative, mainly, of right-wing populist parties' manifestos. But, it seems that the model can also classify very effectively left-wing and center populist manifestos as well, and we will elaborate on that momentarily.

Again, the interpretation of tree models is not straightforward because all the predictors are taken into account in order for the model to predict a new observation, as well as the various interactions between them (several trees). So there are various factors interacted with each other that can predict a party as much or less populist, not just preferences towards a national way of life or opposition to internationalism. However, in figure 2 below we get a general overview of the most important predictors across trees and their contribution to the predictive performance of the model.

Figure 2: Variable Importance of the top 20 most important coding categories.



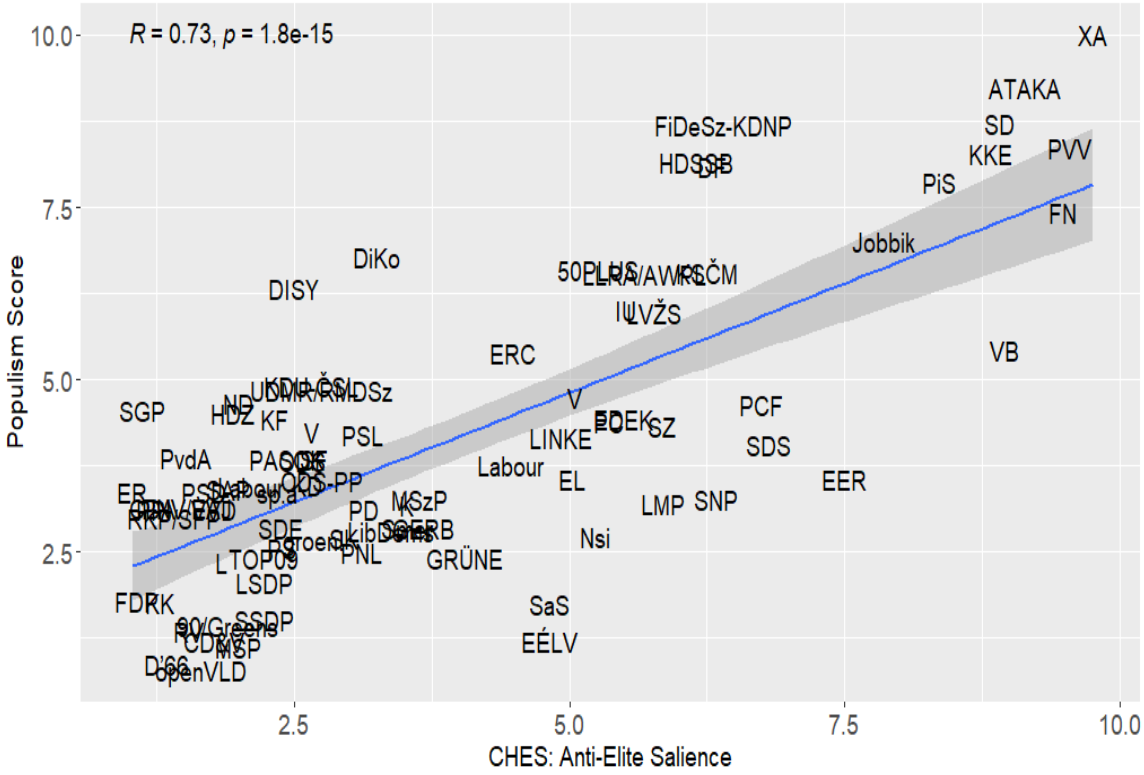
Validation

We validated our measure by estimating its correlation with the "anti-elite" element of the Chapell Hill Expert Survey (CHES)[3]. CHES is a survey that asks experts to place political parties on a scale regarding various ideological dimensions. The "anti-elite" element

[3] See Table 8, Appendix

measures the anti-elite and anti-establishment rhetoric of parties and it is one of the two main components of populism as operationalized in this paper. The "people vs elite" element is the only CHES that is pretty close to the other half of the populism definition, the people-centrism. However, it doesn't grasp exactly that concept, and that is why we don't use it for our validation. It mainly measures a party's stance towards referendums vs representative democracy. Theoretically, it could be a useful index for populism, as long as direct democracy is a key element of many populists political program, due to their lack of support towards representative institutions (Jacobs, Akkerman, and Zaslove 2018), but this is not always the case (Gherghina and Silagadze 2020).

Figure 3[4]: Correlation of populism score with the CHES "Anti-Elite Salience" item.



We see in Figure 3 that our measure strongly, and in a statistically significant way, correlates with the "anti-elite" salience measure. As expected, we observe some high populist parties on the highest scores like XA, KKE, PVV, PiS etc. And also we observe some less populist liberal parties on the lowest scores like openVLD etc. This points out to the fact that our measure is capable to grasp the anti-elitism element of populism.

[4] The parties presented at Figures 3,4,5 and 6 are the parties that matched between our dataset and the CHES dataset.

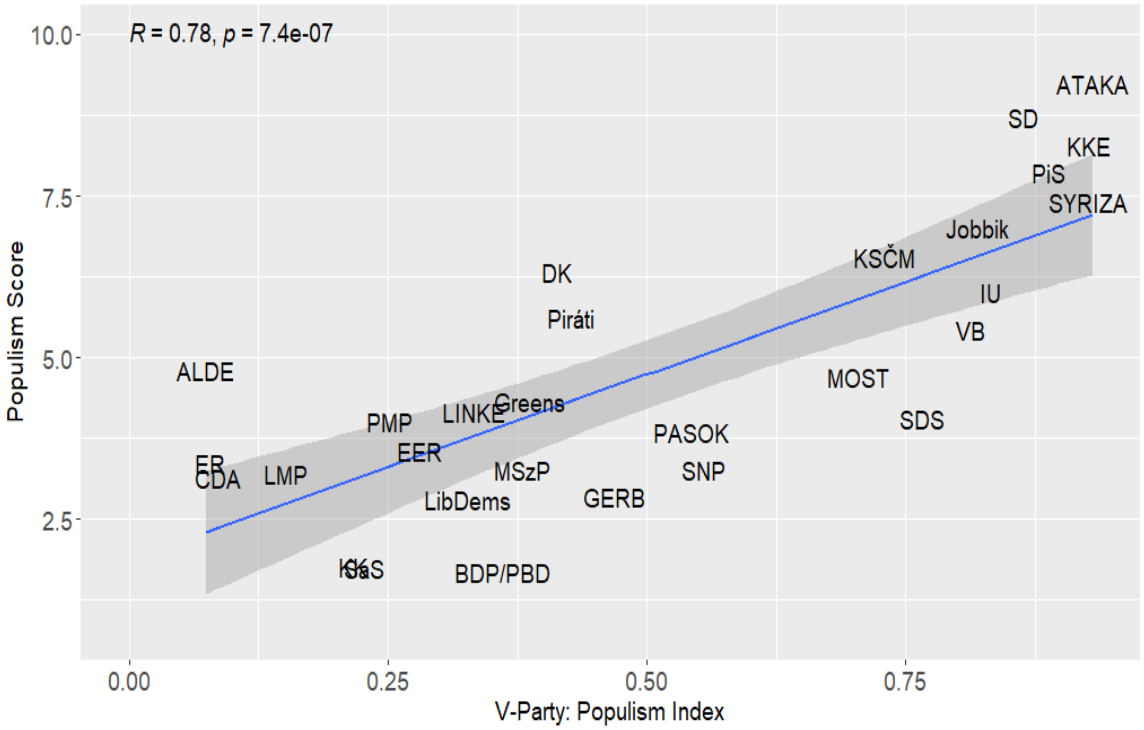
In addition, in Table 1 below we present the correlation coefficients and their p-values between our measure and the different elements of the POPPA survey. As expected, we see that our measure can efficiently grasp both the anti-elitism dimension of populism as well as the people-centrism dimension in its various facets.

Table 1: Correlation between the populism measure and elements of the POPPA survey.

Dimensions	Correlation Coefficients	p-values
Manichean	0.65	3.041e-14
Indivisible	0.65	3.269e-14
General Will	0.65	2.857e-14
People-Centrism	0.66	1.682e-14
Anti-Elitism	0.63	4.332e-13
Populism	0.73	< 2.2e-16

Finally, we validate our measure by correlating it with the populist index of the V-Party dataset. V-Party derives a populism index as a harmonic mean of its people-centrism and anti-elitism items (See Appendix Table 9 for more details). As we see in Figure 4, it correlates very highly with it and in a statistically significant way.

Figure 4: Correlation of populism score with the V-Party "Populism Index".



Validation with Ideology

As we said in the beginning, populism is considered a thin-centered ideology (Mudde, 2004), meaning that it attaches to various, "thick", ideologies in order to exist. It is primarily an ideology that views society separated between the antagonistic groups of people and elite, highlighting the homogeneity of the people, while also proposing the singularity of the people's interest and a Manichean worldview of politics which opposes the "good" people and the "evil" elites. And in order to take its final form it attaches to various other ideologies. That results in left-populist parties, right-wing populist parties etc.

For that reason, it is crucial to detect how well our model can discriminate between different kinds of populism. CHES include an item on Ideology on a scale from 0 (Extreme Left) to 10 (Extreme Right). We used this item for validation after we grouped the parties on the three categories (Left, Center, Right) based on their scores on the scale (Left: 1-4, Center: 5, Right: 6-10). In Figure 5 below we see the correlation of our measure with the "anti-elite" item of CHES, faceted by ideology. We can see that our measure can easily grasp the anti-elitism of every ideological category, since all of the three correlation coefficients are high and statistically significant. The reason that the correlation is slightly higher for right-wing populist parties, as opposed to left-wing, may be attributed to our sample including more right-wing parties. And this happens as long as we draw the manifestos from EU parties, and the EU politics tend to be dominated mainly by right-wing populist parties (as opposed to Latin America where left-wing populism is almost the rule). So it should not be our model or the CMP coding categories responsible for that change.

In order to further prove this point, we ran the same model (GBM) with the same data (the CMP categories), but using a different response variable, the populism index of the V-Party dataset, and then produced a new populism score in the exact same way. Since, V-Party studies parties from all around the world we ended up with a larger dataset consisting of 1057 manifestos and 357 parties from 61 countries around the world from 1970 to 2019, and thus including more left-wing populist parties. In Figure 8 in the Appendix we present the correlation of this new populism score with the CHES "Anti-Elite salience" item, faceted by ideology. In Figure 8 we see that the correlation of our measure with left-wing parties is slightly higher than this with the right-wing ones, indicating that in the first place our inclusion in the dataset of mostly right-wing parties resulted in a measure slightly more capable to grasp right-wing populism, which could easily change if we included (or if POPPA had included) more left-wing parties, as figure 8 indicates. And thus, this should not be a problem of either our model or the CMP coding categories.

That being said in Figure 6 we present again the correlation of our measure with each dimension of the POPPA survey, but this time faceted by party ideology in order to see more analytically which dimensions of populism can our measure grasp between the different party ideologies.

Figure 5: Correlation of Populism Score with the CHES "Anti-Elite salience" item, by party ideology

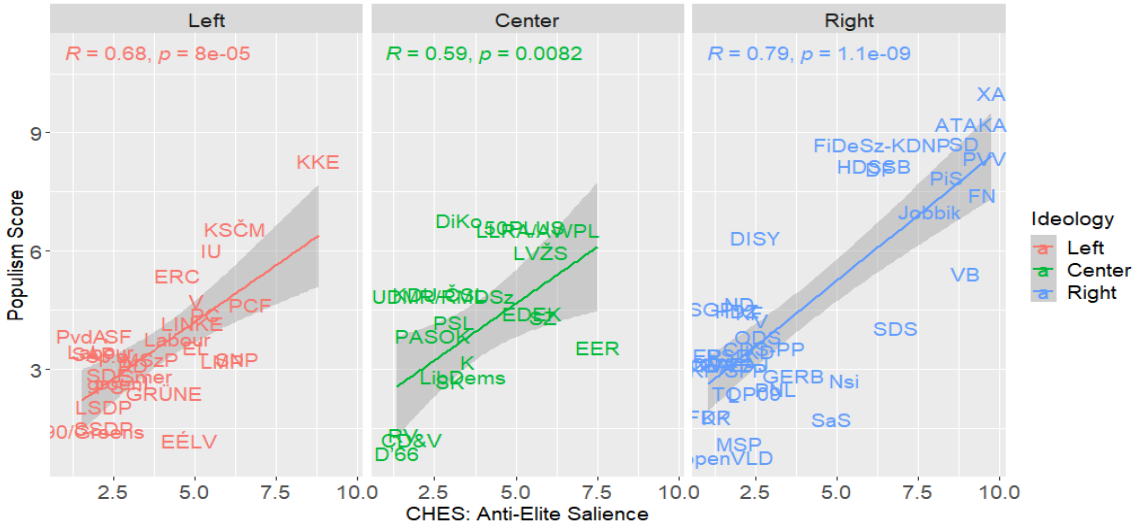
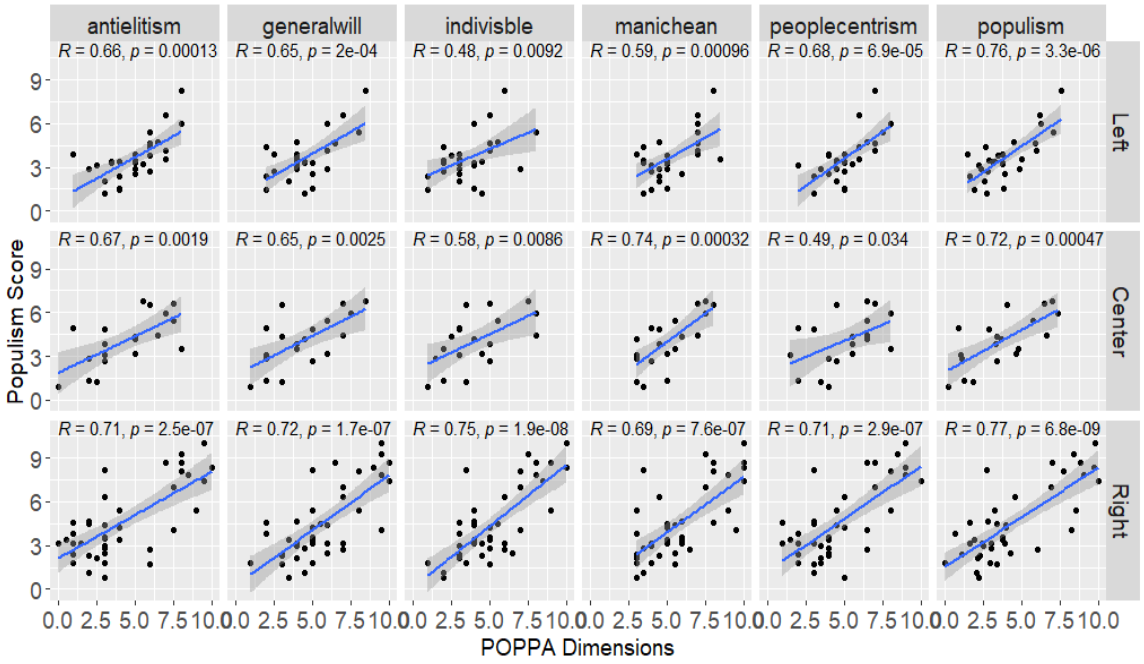


Figure 6[5]: Correlation of Populism Score with the POPPA dimensions, by party ideology



As it is evident from Figure 6, it seems that our measure does not correlate highly with the indivisible element of POPPA, regarding left parties. The same fact has also been highlited by Meijers (2021), regarding left parties, noting that in the "indivisible" element the left scores low. And this is a result in line with the fact that left-wing populist parties have a

[5] The parties presented in this figure are the parties that matched between our dataset and the POPPA dataset

broader definition of the people which is pluralistic, something that has been called inclusionary populism (Mudde and Kaltwasser 2013), and thus does not consider people to be an indivisible or homogeneous entity, but a diverse group of individuals with various interests and needs whose main opponents are the elites.

This is not the case in the right-wing populism idea of the people, which are viewed as a homogeneous sociocultural group, mainly in a tight and nativistic way, which opposes various international elites (and this is proved by the high correlation with the "indivisible" item in Figure 6). This kind of populism has been called exclusionary populism and it mainly characterizes right-wing populist parties in Europe which are also the majority, as opposed to Latin America (Hawkins and Silva 2018; Mudde and Kaltwasser 2013).

As for the center parties, our model doesn't correlate highly and significantly with the people-centrism element which is an expected situation, since center parties are usually not much populist, and this crucial for populism element is missing from their discourse or it is not that strong.

However, the low correlation of our score with the "indivisible" element for the left parties and the people-centrism for center parties, could simply be attributed to the coding categories of the CMP (aside from the fact that our sample includes more right-wing populist parties, as we said before). There may be a few categories missing that could possibly capture populism in left-wing parties as well as center parties (if populism exists at all in their case), or it could be that there is relevant "hidden" meaning behind the already existing categories which could be discovered (Horn et al. 2017). If the problem lies in the first case, then newer categories in the future will be able to solve it. If it lies on the second case, then our machine learning model is doing exactly that ("discovering" meaning) when it "learns" from the data, given that it is also a tree-based model and as such it takes into consideration all the possible interactions between the categories in order to inform its predictions of new values.

Regarding all the other dimensions of POPPA, it seems that our measure performs pretty well since it exhibits a high correlation with each one of them and for all of the ideological families. So, in a summary, by comparing our findings with POPPA and CHES, our populism score seems to grasp the core elements of populism for all types of ideologies. However, it is a little more effective regarding right-wing parties compared to left-wing and centre ones.

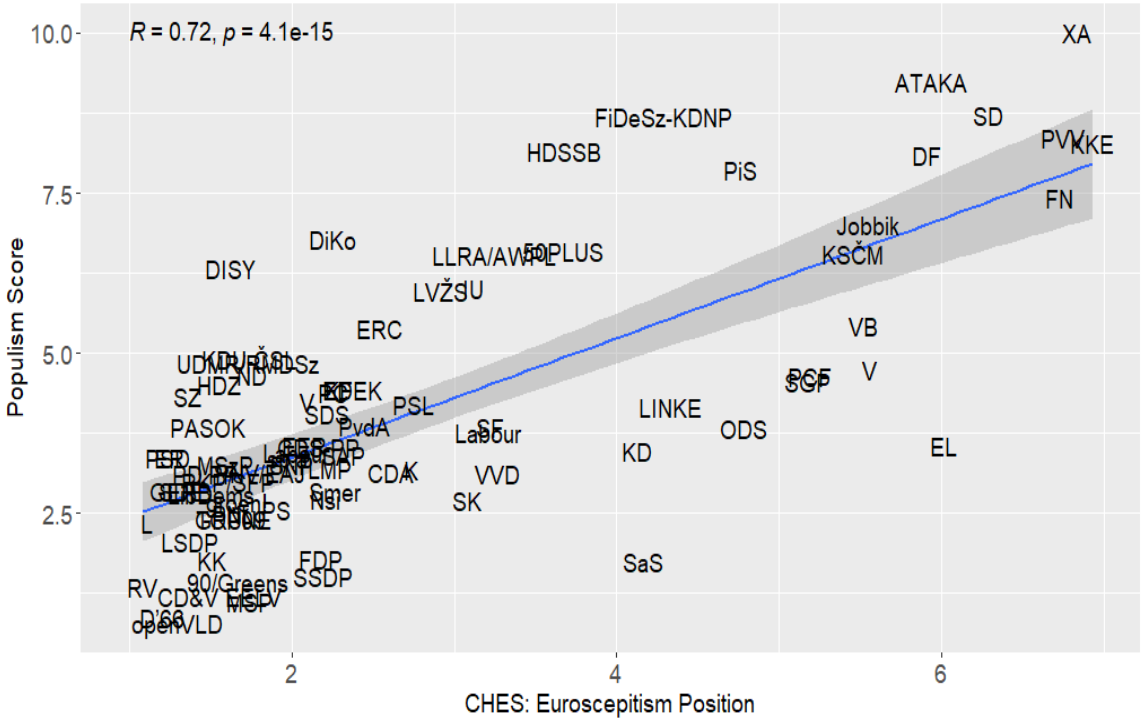
Populism and Euroscepticism

Populism and Euroscepticism have been considered as possibly interrelated phenomena in the literature. For example, while there are some historical differences between the two concepts, they may be interconnected through their common anti-elite discourse (Harmsen 2010; Pirro and Taggart 2018). While populism is a more broad concept referring to the contest between the "pure" people and the "corrupt" elite (Mudde and Kaltwasser 2017), and Euroscepticism is a more specific one, referring to negative stances on EU integration, their interrelation may lie on the fact that EU is framed as the "corrupt" elite which opposes and exploits the "pure" people and is acting against their sovereignty. This is evident in the speeches of some populist leaders like Orbán's and Kaczyński's (Csehi and Zgut 2021).

In our study, this is evident by the high contribution of the EU stances categories in the predictions of our model (Figure 2). Also, in Figure 7 below we present the scatterplot between our measure and the CHES Euroscepticism position which is measured on a scale of 1 (Strongly Opposed) to 7 (Strongly in Favor) which we inverted in order for the diagram to be more interpretable. A strong enough correlation of 0.72 is exhibited and also is statistically significant. Also we find strongly populist parties to be as strongly Euroscepticist ones like XA, KKE, FN and PVV. On the other side of the spectrum we observe low populism parties to also be pro-European ones like the Greek Panhellenic Socialist Movement (PASOK), the Flemish Liberals and Democrats openVLD, as well as Greens of Germany (90/Greens) etc.

It seems that, while these two phenomena are not necessarily complementary to each other, they have a strong connection. Also, we see that our populist measure effectively takes into account the pro-anti EU dimension, which is very important given that the common political framework in which all of the European populist parties act, and thus help them reshape their discourses, is given by the EU institutions, as well as common problems on the European level like the economic crisis of 2008, the pandemic and the war in Ukraine.

Figure 7: Correlation of Populism Score with the CHES Euroscepticism dimension



Discussion

For the purposes of this study, we used the most commonly used "ideational approach" definition to populism, which operationalizes the phenomenon as a discursive scheme which contrasts the "pure" people to the "corrupt" elite, while simultaneously expressing a Manichean worldview of politics and considering the people to be homogeneous and having common interests (general will). Using that operationalization we developed a continuous measure for populism through supervised machine learning, based on the Comparative Manifesto Project coding categories. We showcased its efficiency to scale political parties on a continuous populism scale (1-10), which makes it possible to perform cross-national and longitudinal analyses between different manifestos. In addition, the measure can be aggregated at a higher grouping level as well, thus making possible the comparison between parties and between countries on populism.

As for reliability, the measure performs well since is it's written in R code which can be replicated by other researchers and yield the same manifesto populism scores every time. Regarding validity, the measure seems to measure what it truly aims to measure. This is exhibited by the high and statistically significant correlation coefficients between it and the Chapell Hill Expert Survey (CHES) items, the Populism and Political Parties Expert Survey (POPPA) items as well as the V-Party Dataset items. Also, there is a tendency to grasp right-wing populism a little bit more than the other types of populism. However, the difference is not big. A reason behind that difference, is that our sample is slightly more representative of right-wing populist parties since it is consisted of EU parties, and thus this should not be an issue of our model or the CMP categories.

Also, theoretically and conceptually speaking, this tendency of our model may not be necessarily a problem because a recent ML-based study showed that right-wing populism is the most consistently defined and stable type of populism in the relevant literature (Naxera, Kaše, and Stulík 2023), while the other types fail to constitute strong conceptual categories on their own, they are not used consistently and they also many times overlap with right-wing populism.

In addition, the measure correlates with the Euroscepticism element of the CHES, indicating that: first, our measure also grasps efficiently Eurosceptic content in party manifestos and second, that Euroscepticism and Populism, although not the complementary to one another, they are strongly interrelated.

The present study is one of the few studies that deal with populism on a large scale comparative perspective, and so avoiding the usual case study approach which is common in the populism literature. Furthermore, our model was trained on a large scale and cross-national pooled manifesto dataset and so the paper is one of the few large scale comparative projects on populism using computational text analysis methods. The other two are that of Cocco and Monechi (2021) and that of Hawkins and Silva (2018).

Also some of the contributions of this paper are the following: First, by using automated methods we offer a more cost-efficient alternative to measure populism than more resource-intensive methods of human annotation which require plenty of coders, time and

money. Second, we can easily quantify the level of populism across a large number of manifestos, parties and countries very quickly. Third, we develop a continuous measure, which, in oppose to dichotomous measures, it avoids arbitrary classifications and can also be in place to detect a "strategic" kind of populism.

Furthermore, we recognize the work of Cocco and Monechi (2021), as one of the main research papers that mobilized tools from the machine learning field in order to automate and improve the comparative measurement of populism through space and time. As long as we wanted to facilitate the quantitative comparative study of populism as well, our work capitalizes on the effort of (Cocco and Monechi 2021) and takes a step forward. First of all, we include 26 countries instead of 6. Also, although our models are similar, ours are trained on a pooled cross-national dataset, and not on each country separately. So our model draws information from a common pool of data and not from a specific linguistic context only, thus making comparisons between countries more feasible.

Also, our model is more computationally efficient because we use the relative frequencies of the CMP coding categories as an initial text representation. So we are working with summary information of just the 56 CMP coding categories, making the algorithm run faster, and not with a high-dimensional bag of words representation, which could include potentially millions of words (or tokens) and so would be more cumbersome to run. Finally, our model is trained on continuous labeled data and not dichotomous as those of Cocco and Monechi, thus avoiding some drawbacks of the dichotomous operationalization, discussed earlier in the paper.

However, there are some potential problems related to our method which need to be addressed here. First of all, because we examined a relatively short time period in our analysis we cannot easily make within-parties and within-countries comparisons of populism scores through time, as most of the 190 parties in our dataset have at maximum 2 election manifestos. In addition, splitting the dataset for training and testing, also results in the reduction of the manifestos left in the test set (30% of the manifestos). However, as new elections take place and the CMP database is updated with new annotated manifestos, researchers in the future could run our model and obtain updated populism scores of the existing- and/or new- parties, thus having more data points and more scores to conduct temporal comparisons of within-parties/countries populism.

The second problem relates to whether the existing CMP coding categories are enough to measure populism or there should be new categories included in future CMP versions. Should there not be any new categories, then there are proposed ways in the literature that the researchers can discover the "hidden" meaning of the existing ones (Horn et al. 2017), or take into account the uncertainty of the categories due to potential measurement error during the analysis (Benoit, Laver, and Mikhaylov 2009), in order to improve the CMP data quality.

Finally, a third problem would be the lack of expert measurements of party populism before 2018. Given that we opt to measure populism as a continuous dimension that changes over time, we need to have repeated expert measurements over time in order for our dependent variable to be updated and so to run our models repetitively in order to get updated populism scores. Since, there are no such measurements before 2018, we cannot capture populism accurately going many years back. However, new survey rounds in the future,

will help overcome this problem for now and onwards. For that matter, the 2023 round of the POPPA survey, is already fielded and about to be uploaded.

In conclusion, given that over the last decades populist actors and parties are rising forces in the political arena of liberal democracies, and the scientific literature on the phenomenon is flourishing, we argue that there is a need for further quantification and automation of the populism measurement in order for fruitful comparative analyses to take place in the future. We hope that this paper takes a little step towards that research direction.

References

Aslanidis, Paris. 2018. "Measuring Populist Discourse with Semantic Text Analysis: An Application on Grassroots Populist Mobilization." *Quality & Quantity* 52(3): 1241–63.

Benoit, Kenneth, Michael Laver, and Slava Mikhaylov. 2009. "Treating Words as Data with Error: Uncertainty in Text Statements of Policy Positions." *American Journal of Political Science* 53(2): 495–513.

Bonikowski, Bart, and Noam Gidron. 2016. "The Populist Style in American Politics: Presidential Campaign Rhetoric, 1952-1996." *Social Forces* 94.

Canovan, Margaret. 1999. "Trust the People! Populism and the Two Faces of Democracy." *Political Studies* 47(1): 2–16.

Cocco, Jessica Di, and Bernardo Monechi. 2021. "How Populist Are Parties? Measuring Degrees of Populism in Party Manifestos Using Supervised Machine Learning." *Political Analysis*: 1–17.

Csehi, Robert, and Edit Zgut. 2021. "'We Won't Let Brussels Dictate Us': Eurosceptic Populism in Hungary and Poland." *European Politics and Society* 22(1): 53–68.

Gherghina, Sergiu, and Nanuli Silagadze. 2020. "Populists and Referendums in Europe: Dispelling the Myth." *The Political Quarterly* 91(4): 795–805.

Gidron, Noam, and Bart Bonikowski. 2013. "Varieties of Populism: Literature Review and Research Agenda." https://papers.ssrn.com/abstract=2459387 (May 24, 2023).

Harmsen, Robert. 2010. "Concluding Comment: On Understanding the Relationship between Populism and Euroscepticism." *Perspectives on European Politics and Society* 11(3): 333–41.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning*. New York, NY: Springer. http://link.springer.com/10.1007/978-0-387-84858-7 (May 24, 2023).

Hawkins, Kirk A., and Cristóbal Rovira Kaltwasser. 2017. "What the (Ideational) Study of Populism Can Teach Us, and What It Can't." *Swiss Political Science Review* 23(4): 526–42.

Hawkins, Kirk A, and Bruno Castanho Silva. 2018. “A Head-to-Head Comparison of Human-Based and Automated Text Analysis for Measuring Populism in 27 Countries.” : 42.

Horn, Alexander, Anthony Kevins, Carsten Jensen, and Kees Van Kersbergen. 2017. “Peeping at the Corpus – What Is Really Going on behind the Equality and Welfare Items of the Manifesto Project?” *Journal of European Social Policy* 27(5): 403–16.

Hunger, Sophia, and Fred Paxton. 2021. “What’s in a Buzzword? A Systematic Review of the State of Populism Research in Political Science.” *Political Science Research and Methods*: 1–17.

Jacobs, Kristof, Agnes Akkerman, and Andrej Zaslove. 2018. “The Voice of Populist People? Referendum Preferences, Practices and Populist Attitudes.” *Acta Politica* 53(4): 517–41.

Kaltwasser, Cristóbal Rovira, Paul A. Taggart, Paulina Ochoa Espejo, and Pierre Ostiguy. 2017. *The Oxford Handbook of Populism*. Oxford University Press.

Laclau, Ernesto. 2005. *On Populist Reason*. Verso.

Lehmann, Pola / Franzmann, Simon / Burst, Tobias / Lewandowski, Jirka / Matthieß, Theres / Regel, Sven / Riethmüller, Felicia / Zehnter, Lisa (2023): Manifesto Corpus. Version: 2023-a. Berlin: WZB Berlin Social Science Center/Göttingen: Institute for Democracy Research (IfDem).

Meijers, Maurits J., and Andrej Zaslove. 2021. “Measuring Populism in Political Parties: Appraisal of a New Approach.” *Comparative Political Studies* 54(2): 372–407.

Montgomery, Jacob M., and Santiago Olivella. 2018. “Tree-Based Models for Political Science Data.” *American Journal of Political Science* 62(3): 729–44.

Mudde, Cas. 2004. “The Populist Zeitgeist.” *Government and Opposition* 39(4): 541–63.

Mudde, Cas, and Cristóbal Rovira Kaltwasser. 2013. “Exclusionary vs. Inclusionary Populism: Comparing Contemporary Europe and Latin America.” *Government and Opposition* 48(2): 147–74.

Mudde, Cas, and Cristobal Rovira Kaltwasser. 2017. *Populism: A Very Short Introduction*. Oxford University Press.

Mudde, Cas, and Cristóbal Rovira Kaltwasser. 2018. “Studying Populism in Comparative Perspective: Reflections on the Contemporary and Future Research Agenda.” *Comparative Political Studies* 51(13): 1667–93.

Naxera, Vladimír, Vojtěch Kaše, and Ondřej Stulík. 2023. “‘The More Populism Types You Know, the Better Political Scientist You Are?’ Machine-Learning Based

Meta-Analysis of Populism Types in the Political Science Literature." *Journal of Contemporary European Studies* 0(0): 1–18.

Ostiguy, Pierre. 2017. "Populism: A Socio-Cultural Approach." *The Oxford Handbook of Populism*. https://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780198803560.001.0001/oxfordhb-9780198803560-e-3 (May 2, 2022).

Pauwels, Teun. 2011. "Measuring Populism: A Quantitative Text Analysis of Party Literature in Belgium." *Journal of Elections, Public Opinion and Parties* 21(1): 97–119.

Pirro, Andrea LP, and Paul Taggart. 2018. "The Populist Politics of Euroscepticism in Times of Crisis: A Framework for Analysis." *Politics* 38(3): 253–62.

Rooduijn, Matthijs, and Teun Pauwels. 2011. "Measuring Populism: Comparing Two Methods of Content Analysis." *West European Politics* 34(6): 1272–83.

Weyland, Kurt. 2017. "Populism: A Political-Strategic Approach." *The Oxford Handbook of Populism*. https://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780198803560.001.0001/oxfordhb-9780198803560-e-2 (May 2, 2022).

APPENDIX

Data

Table 3: Parties, Countries and Election Dates in the dataset.

Party Name	Country Name	Party Abbreviation	Election Dates
50Plus	Netherlands	50PLUS	2021-03-17, 2017-03-15, 2012-09-12
ANO 2011	Czech Republic	ANO	2017-10-21, 2013-10-26
Alliance of Federation of Young Democrats - Christian Democratic People's Party	Hungary	FiDeSz-KDNP	2018-04-08
Alliance of Federation of Young Democrats - Hungarian Civic Union - Christian Democratic People's Party	Hungary	FiDeSz-MPSz-KDNP	2014-04-06, 2010-04-11

Alliance of Liberals and Democrats	Romania	ALDE	2016-12-11
Alliance‘90/Greens	Germany	90/Greens	2021-09-26, 2017-09-24, 2013-09-22
Alternative for Bulgarian Revival	Bulgaria	ABV	2014-10-05
Alternative for Germany	Germany	AfD	2021-09-26, 2017-09-24, 2013-09-22
Austrian Freedom Party	Austria	FPÖ	2019-09-29, 2017-10-15, 2013-09-29
Austrian People’s Party	Austria	ÖVP	2019-09-29, 2017-10-15, 2013-09-29
Austrian Social Democratic Party	Austria	SPÖ	2019-09-29, 2017-10-15, 2013-09-29
Basque Nationalist Party	Spain	PNV/EAJ	2019-11-10, 2019-04-28, 2016-06-26, 2015-12-20, 2011-11-20
Bridge	Slovakia		2016-03-05, 2012-03-10, 2010-06-12
Bridge of Independent Lists	Croatia	MOST	2020-07-05, 2016-09-11, 2015-11-08
Brothers of Italy	Italy	FDI	2018-03-04
Brothers of Italy - National Centre-right	Italy	FDI-CDN	2013-02-24
Bulgaria without Censorship	Bulgaria	BBZ	2014-10-05
Bulgarian Socialist Party	Bulgaria	BSP	2013-05-12
Canarian Coalition	Spain	CC	2011-11-20
Canarian Coalition and Canarian Nationalist Party	Spain	CC-PNC	2016-06-26, 2015-12-20
Canarian Coalition–New Canaries	Spain	CCa-PNC-NC	2019-11-10, 2019-04-28
Catalan Republican Left	Spain	ERC	2019-11-10, 2019-04-28, 2016-06-26, 2015-12-20, 2011-11-20
Centre Party	Sweden	CP	2018-09-09, 2014-09-14, 2010-09-19
Christian Democratic Appeal	Netherlands	CDA	2021-03-17, 2017-03-15, 2012-09-12, 2010-06-09
Christian Democratic Movement	Slovakia	KDH	2016-03-05, 2012-03-10, 2010-06-12

Christian Democratic People's Party of Switzerland	Switzerland	CVP/PDC	2019-10-20, 2015-10-18, 2011-10-23
Christian Democratic and Flemish	Belgium	CD&V	2019-05-26, 2014-05-25, 2010-06-13
Christian Democrats	Sweden	Kd	2018-09-09, 2014-09-14, 2010-09-19
Christian Democrats in Finland	Finland	KD	2019-04-14, 2015-04-19, 2011-04-17
Christian Union	Netherlands	CU	2021-03-17, 2017-03-15, 2012-09-12, 2010-06-09
Christian and Democratic Union - Czech People's Party	Czech Republic	KDU-ČSL	2017-10-21, 2013-10-26, 2010-05-29
Citizens - Party of the Citizens	Spain	C's	2019-11-10, 2019-04-28, 2016-06-26, 2015-12-20
Citizens for European Development of Bulgaria	Bulgaria	GERB	2017-03-26, 2014-10-05, 2013-05-12
Citizens' Alliance	Cyprus	SYPOL	2016-05-22
Civic Democratic Party	Czech Republic	ODS	2017-10-21, 2013-10-26, 2010-05-29
Civic Platform	Poland	PO	2015-10-25, 2011-10-09
Coalition of the Radical Left	Greece	SYRIZA	2019-07-07, 2015-09-20, 2015-01-25
Coalition of the Radical Left - Unionist Social Front	Greece	SYRIZA-EKN	2012-06-17
Communist Party of Bohemia and Moravia	Czech Republic	KSČM	2017-10-21, 2013-10-26, 2010-05-29
Communist Party of Greece	Greece	KKE	2019-07-07, 2015-09-20, 2015-01-25, 2012-06-17, 2012-05-06
Conservative Democratic Party of Switzerland	Switzerland	BDP/PBD	2019-10-20, 2015-10-18, 2011-10-23
Conservative Party	United Kingdom	Conservatives	2019-12-12, 2017-06-08, 2015-05-07, 2010-05-06
Conservative People's Party of Estonia	Estonia	EKRE	2019-03-03, 2015-03-01

Conservative People's Party	Denmark	KF	2019-06-05, 2015-06-18, 2011-09-15
Croatian Democratic Assembly of Slavonia and Baranja	Croatia	HDSSB	2015-11-08, 2011-12-04
Croatian Democratic Union	Croatia	HDZ	2020-07-05, 2016-09-11, 2011-12-04
Croatian Peasant Party	Croatia	HSS	2011-12-04
Croatian People's Party - Liberal Democrats	Croatia	HNS	2020-07-05
Czech Pirate Party	Czech Republic	Piráti	2017-10-21
Czech Social Democratic Party	Czech Republic	ČSSD	2017-10-21, 2013-10-26, 2010-05-29
DENK	Netherlands	DENK	2021-03-17, 2017-03-15
Danish People's Party	Denmark	DF	2019-06-05, 2015-06-18, 2011-09-15
Danish Social-Liberal Party	Denmark	RV	2019-06-05, 2015-06-18, 2011-09-15
Democratic Coalition	Cyprus	DISY	2016-05-22, 2011-05-22
Democratic Coalition	Hungary	DK	2018-04-08, 2014-04-06
Democratic Convergence of Catalonia	Spain	CDC	2016-06-26
Democratic Left Alliance	Poland	SLD	2011-10-09
Democratic Movement	France	MoDem	2017-06-11, 2012-06-10
Democratic Party	Cyprus	DiKo	2016-05-22, 2011-05-22
Democratic Party	Italy	PD	2018-03-04, 2013-02-24
Democratic Party of Pensioners of Slovenia	Slovenia	DeSUS	2018-06-03, 2014-07-13, 2011-12-04
Democrats'66	Netherlands	D'66	2021-03-17, 2017-03-15, 2012-09-12, 2010-06-09
Direction-Social Democracy	Slovakia	Smer	2016-03-05, 2012-03-10, 2010-06-12
Ecological and Environmental Movement	Cyprus		2011-05-22

Ecologists	Belgium	ECOLO	2010-06-13
Election Action of Lithuania's Poles	Lithuania	LLRA/AWPL	2020-10-11, 2016-10-09, 2012-10-14
Estonian Center Party	Estonia	K	2019-03-03, 2015-03-01, 2011-03-06
Estonian Greens	Estonia	EER	2011-03-06
Estonian Reform Party	Estonia	ER	2019-03-03, 2015-03-01, 2011-03-06
Europe Ecology - The Greens	France	EÉLV	2017-06-11, 2012-06-10
FDP.The Liberals	Switzerland	FDP/PLR	2019-10-20, 2015-10-18, 2011-10-23
Family of the Irish	Ireland		2016-02-26, 2011-02-25
Federal Democratic Union	Switzerland	EDU/UDF	2019-10-20
Finnish Centre	Finland	SK	2019-04-14, 2015-04-19, 2011-04-17
Finnish Social Democrats	Finland	SSDP	2019-04-14, 2015-04-19, 2011-04-17
Five Star Movement	Italy	M5S	2018-03-04, 2013-02-24
Flemish Interest	Belgium	VB	2019-05-26, 2014-05-25, 2010-06-13
Forum for Democracy	Netherlands	FvD	2021-03-17, 2017-03-15
Francophone Democratic Front of Francophones	Belgium	FDF	2014-05-25
Francophone Socialist Party	Belgium	PS	2014-05-25, 2010-06-13
Free Democratic Party	Germany	FDP	2021-09-26, 2017-09-24, 2013-09-22
Free Party	Estonia	EVA	2015-03-01
Freedom and Direct Democracy	Czech Republic	SPD	2017-10-21
Freedom and Solidarity	Slovakia	SaS	2016-03-05, 2012-03-10, 2010-06-12
French Communist Party	France	PCF	2017-06-11
Geneva Citizens' Movement	Switzerland	MCG	2015-10-18, 2011-10-23
Go Italy	Italy	FI	2018-03-04
Golden Dawn	Greece	XA	2019-07-07, 2015-01-25, 2012-06-17, 2012-05-06

Green Ecology Party	Sweden	MP	2018-09-09, 2014-09-14, 2010-09-19
Green Left	Netherlands	GL	2021-03-17, 2017-03-15, 2012-09-12, 2010-06-09
Green Liberal Party	Switzerland	GLP	2019-10-20, 2015-10-18, 2011-10-23
Green Party	Czech Republic	SZ	2013-10-26, 2010-05-29
Green Party	Ireland	Greens	2016-02-26, 2011-02-25
Green Party of England and Wales	United Kingdom	GPEW	2019-12-12, 2017-06-08, 2015-05-07
Green Party of Switzerland	Switzerland	GPS/PES	2019-10-20, 2015-10-18, 2011-10-23
Green Union	Finland	VL	2019-04-14, 2015-04-19, 2011-04-17
Green!	Belgium	groen!	2019-05-26, 2014-05-25, 2010-06-13
Homeland Union - Lithuanian Christian Democrats	Lithuania	TS-LKD	2020-10-11, 2016-10-09, 2012-10-14
Human Shield	Croatia		2015-11-08
Hungarian Democratic Alliance of Romania	Romania	UDMR/RMDSz	2016-12-11, 2012-12-09
Hungarian Socialist Party	Hungary	MSzP	2018-04-08, 2014-04-06, 2010-04-11
Independent Alliance	Ireland		2016-02-26
Independent Greeks	Greece	ANEL	2015-01-25, 2012-06-17, 2012-05-06
Indomitable France	France		2017-06-11
Kotleba – People's Party Our Slovakia	Slovakia	ĽSNS	2016-03-05
Labour Party	Ireland	Labour	2016-02-26, 2011-02-25
Labour Party	Lithuania	DP	2020-10-11, 2016-10-09, 2012-10-14
Labour Party	Netherlands	PvdA	2021-03-17, 2017-03-15, 2012-09-12, 2010-06-09
Labour Party	United Kingdom	Labour	2019-12-12, 2017-06-08, 2015-05-07, 2010-05-06
Law and Justice	Poland	PiS	2019-10-13, 2015-10-25, 2011-10-09
League	Italy	L	2018-03-04

Left Bloc	Portugal	BE	2019-10-06, 2015-10-04, 2011-06-05
Left Party	Sweden	V	2018-09-09, 2014-09-14, 2010-09-19
Left Wing Alliance	Finland	VAS	2019-04-14, 2015-04-19, 2011-04-17
Liberal Alliance	Denmark		2019-06-05, 2015-06-18, 2011-09-15
Liberal Democrats	United Kingdom	LibDems	2019-12-12, 2017-06-08, 2015-05-07, 2010-05-06
Liberal Movement	Lithuania	LRLS	2020-10-11, 2016-10-09, 2012-10-14
Liberal People's Party	Sweden	FP	2014-09-14, 2010-09-19
Liberals	Denmark	V	2019-06-05, 2015-06-18, 2011-09-15
Liberals	Sweden	L	2018-09-09
Lithuanian Peasant and Green Union	Lithuania	LVŽS	2020-10-11, 2016-10-09, 2012-10-14
Lithuanian Social Democratic Party	Lithuania	LSDP	2020-10-11, 2016-10-09, 2012-10-14
Mayors and Independents	Czech Republic	STAN	2017-10-21
Milan Bandić 365 - The Party of Labour and Solidarity	Croatia		2015-11-08
Moderate Coalition Party	Sweden	MSP	2018-09-09, 2014-09-14, 2010-09-19
Movement for Rights and Freedoms	Bulgaria	DPS	2017-03-26, 2013-05-12
Movement for a Better Hungary	Hungary	Jobbik	2018-04-08, 2014-04-06, 2010-04-11
Movement of Ecologists - Citizens' Cooperation	Cyprus		2016-05-22
National Coalition	Finland	KK	2019-04-14, 2015-04-19, 2011-04-17
National Front	France	FN	2017-06-11, 2012-06-10
National Liberal Party	Romania	PNL	2016-12-11
National Union Attack	Bulgaria	ATAKA	2014-10-05, 2013-05-12
New Democracy	Greece	ND	2019-07-07, 2015-09-20, 2015-01-25, 2012-06-17, 2012-05-06

New Flemish Alliance	Belgium	N-VA	2019-05-26, 2014-05-25, 2010-06-13
New Slovenian Christian People's Party	Slovenia	Nsi	2018-06-03, 2014-07-13, 2011-12-04
Northern League	Italy	LN	2013-02-24
Open Flemish Liberals and Democrats	Belgium	openVLD	2019-05-26, 2014-05-25, 2010-06-13
Order and Justice	Lithuania	PTT	2016-10-09, 2012-10-14
Ordinary People and Independent Personalities	Slovakia	OľaNO	2016-03-05, 2012-03-10
Panhellenic Socialist Movement	Greece	PASOK	2015-09-20, 2015-01-25, 2012-06-17, 2012-05-06
Party for the Animals	Netherlands	PvdD	2021-03-17, 2017-03-15, 2012-09-12, 2010-06-09
Party of Freedom	Netherlands	PVV	2021-03-17, 2017-03-15, 2012-09-12, 2010-06-09
People Before Profit	Ireland	PBP	2016-02-26
People's Movement Party	Romania	PMP	2016-12-11
People's Party	Spain	PP	2019-11-10, 2019-04-28, 2016-06-26, 2015-12-20, 2011-11-20
People's Party for Freedom and Democracy	Netherlands	VVD	2021-03-17, 2017-03-15, 2012-09-12, 2010-06-09
Peter Pilz List	Austria	PILZ	2017-10-15
Polish Peasants' Party	Poland	PSL	2015-10-25, 2011-10-09
Politics Can Be Different	Hungary	LMP	2018-04-08, 2014-04-06, 2010-04-11
Popular Unity	Greece	LAE	2015-09-20
Pro Patria	Estonia		2019-03-03
Pro Patria and Res Publica Union	Estonia	IRL	2015-03-01, 2011-03-06
Progressive Party of the Working People	Cyprus	AKEL	2016-05-22, 2011-05-22
Red-Green Unity List	Denmark	EL	2019-06-05, 2015-06-18, 2011-09-15

Reform Movement	Belgium	MR	2014-05-25, 2010-06-13
Reformed Political Party	Netherlands	SGP	2021-03-17, 2017-03-15, 2012-09-12, 2010-06-09
Republic Onwards!	France	LREM	2017-06-11
Save Romania Union	Romania	USR	2016-12-11
Scottish National Party	United Kingdom	SNP	2019-12-12, 2017-06-08, 2015-05-07
Slovak National Party	Slovakia	SNS	2016-03-05, 2012-03-10, 2010-06-12
Slovenian Democratic Party	Slovenia	SDS	2018-06-03, 2014-07-13, 2011-12-04
Social Democratic Center-Popular Party	Portugal	CDS-PP	2019-10-06, 2011-06-05
Social Democratic Labour Party	Sweden	SAP	2018-09-09, 2014-09-14, 2010-09-19
Social Democratic Party	Denmark	SD	2019-06-05, 2015-06-18, 2011-09-15
Social Democratic Party	Estonia	SDE	2019-03-03, 2015-03-01, 2011-03-06
Social Democratic Party	Portugal	PSD	2019-10-06, 2011-06-05
Social Democratic Party	Slovenia	SD	2018-06-03, 2014-07-13, 2011-12-04
Social Democratic Party of Germany	Germany	SPD	2021-09-26, 2017-09-24, 2013-09-22
Social Democratic Party of Switzerland	Switzerland	SPS/PSS	2019-10-20, 2015-10-18, 2011-10-23
Social Democrats' Movement	Cyprus	EDEK	2016-05-22, 2011-05-22
Socialist Party	France	PS	2017-06-11, 2012-06-10
Socialist Party	Ireland		2011-02-25
Socialist Party	Netherlands	SP	2021-03-17, 2017-03-15, 2012-09-12, 2010-06-09
Socialist Party	Portugal	PS	2019-10-06, 2015-10-04, 2011-06-05
Socialist Party Different	Belgium	sp.a	2019-05-26, 2014-05-25, 2010-06-13
Socialist People's Party	Denmark	SF	2019-06-05, 2015-06-18, 2011-09-15
Soldiers of Destiny	Ireland		2016-02-26, 2011-02-25
Spanish Socialist Workers' Party	Spain	PSOE	2019-11-10, 2019-04-28, 2016-06-26,

			2015-12-20, 2011-11-20
Sweden Democrats	Sweden	SD	2018-09-09, 2014-09-14, 2010-09-19
Swedish People's Party	Finland	RKP/SFP	2019-04-14, 2015-04-19, 2011-04-17
Swiss Labour Party	Switzerland	PdAS/PdTS	2019-10-20, 2015-10-18
Swiss People's Party	Switzerland	SVP/UDC	2019-10-20, 2015-10-18, 2011-10-23
The Greens	Austria	GRÜNE	2019-09-29, 2017-10-15, 2013-09-29
The Left	Germany	LINKE	2021-09-26, 2017-09-24, 2013-09-22
The New Austria	Austria	NEOS	2013-09-29
The New Austria and Liberal Forum	Austria	NEOS	2019-09-29, 2017-10-15
The Party of Wales	United Kingdom	PC	2019-12-12, 2017-06-08, 2015-05-07
The Republicans	France		2017-06-11
The River	Greece		2015-01-25
Ticino League	Switzerland	LdT	2019-10-20
Tradition, Responsibility, Prosperity 09	Czech Republic	TOP09	2017-10-21, 2013-10-26, 2010-05-29
True Finns	Finland	PS	2019-04-14, 2015-04-19, 2011-04-17
Union for a Popular Movement	France	UMP	2012-06-10
Union of Centrists	Greece	EK	2015-09-20
Union of Democrats and Independents	France	UDI	2017-06-11
United Kingdom Independence Party	United Kingdom	UKIP	2017-06-08, 2015-05-07
United Left	Spain	IU	2019-11-10, 2019-04-28, 2011-11-20
We Are Family	Slovakia		2016-03-05
We Ourselves	Ireland	SF	2016-02-26, 2011-02-25
We can	Spain		2019-11-10, 2019-04-28, 2016-06-26, 2015-12-20
Will	Bulgaria		2017-03-26

Table 4: Comparative Manifesto Project Main Coding Categories

Code Number	Code Description
per101	Foreign Special Relationships: Positive
per102	Foreign Special Relationships: Negative
per103	Anti-Imperialism
per104	Military: Positive
per105	Military: Negative
per106	Peace
per107	Internationalism: Positive
per108	European Community/Union: Positive
per109	Internationalism: Negative
per110	European Community/Union: Negative
per201	Freedom and Human Rights
per202	Democracy
per203	Constitutionalism: Positive
per204	Constitutionalism: Negative
per301	Decentralization
per302	Centralisation
per303	Governmental and Administrative Efficiency
per304	Political Corruption
per305	Political Authority
per401	Free Market Economy
per402	Incentives: Positive
per403	Market Regulation
per404	Economic Planning
per405	Corporatism/Mixed Economy
per406	Protectionism: Positive
per407	Protectionism: Negative
per408	Economic Goals
per409	Keynesian Demand Management
per410	Economic Growth: Positive
per411	Technology and Infrastructure: Positive
per412	Controlled Economy
per413	Nationalisation
per414	Economic Orthodoxy
per415	Marxist Analysis
per416	Anti-Growth Economy: Positive
per501	Environmental Protection
per502	Culture: Positive
per503	Equality: Positive
per504	Welfare State Expansion
per505	Welfare State Limitation
per506	Education Expansion
per507	Education Limitation
per601	National Way of Life: Positive
per602	National Way of Life: Negative
per603	Traditional Morality: Positive

per604	Traditional Morality: Negative
per605	Law and Order: Positive
per606	Civic Mindedness: Positive
per607	Multiculturalism: Positive
per608	Multiculturalism: Negative
per701	Labour Groups: Positive
per702	Labour Groups: Negative
per703	Agriculture and Farmers: Positive
per704	Middle Class and Professional Groups
per705	Underprivileged Minority Groups
per706	Non-economic Demographic Groups

Table 5: Dimensions of populism on the POPPA expert survey: descriptions and scale.

Dimensions	**Scale**
manichean = Some parties see politics as a moral struggle between good and bad. This is often described as a Manichean worldview. Please tick the box that best describes the degree to which each party holds a Manichean worldview.	0 = Not Manichean at all : 10 = Extremely Manichean
indivisible = Some parties consider the ordinary people to be indivisible (i.e. the people are seen as homogenous). Please tick the box that best describes each party in this respect.	0 = Not at all : 10 = Very much
generalwill = Some parties consider the ordinary people's interests to be singular (i.e. one can speak of a 'general will')? Please tick the box that best describes each party in this respect.	0 = Not at all : 10 = Very much
peoplecentrism = Some parties believe that sovereignty should lie exclusively with the ordinary people (i.e. the ordinary people, not the elites, should have the final say in politics). Please tick the box that best describes the extent to which each party considers the ordinary people to be sovereign.	0 = Not at all : 10 = Very much
antielitism = Some parties can be characterized by their anti-elitism. Please tick the box that best describes the extent to which each party can be considered to be anti-elitist.	0 = Not at all anti-elitist : 10 = Very anti-elitist
populism = Variable based on the factor regression scores of the following items:	0 = Not at all populist : 10 = Very populist

‘manichean’, ‘indivisible’, ‘generalwill’, ‘peoplecentrism’, and ‘antielitism’.	

Methods

The RMSE is calculated with the following formula, which is pretty similar to the standard deviation formula, and the perfect model would have a value of 0:

$$\sqrt{\sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{n}} \quad (1)$$

Table 6: Supervised machine learning models and their RMSE

Model	RMSE
GBM	0.70
Ridge Regression	0.77
SVM (Polynomial Kernel)	0.78
SVM (Radial Kernel)	0.73
Random Forest	0.72
SVM Linear	0.85
Tree	0.89

Tree models have proved to be very useful in political science research (see Montgomery and Olivella 2018 for more detail). That is because they can grasp a lot of interactions between the variables and thus increase their predictive accuracy without the need to rely on cumbersome model specifications with high order polynomials, term multiplications etc.

The intuition behind tree methods is to split the data according to some splitting criterion creating decision trees with leafs and nodes. Data points all the way down to the terminal nodes are forming groups which are used in order to make predictions to unseen data. In the regression setting, the prediction of a new data point is made using the average of the data points in its region. In the classification setting, the prediction is made by using the most commonly occurring class in each region.

For the regression tasks, like ours, the splitting rule for growing the tree is minimizing a loss function which is almost identical to that of the sum of squared residuals,

$$\sum_{m=1}^{M} \sum_{x_i \in R_m} \left(y_i - \hat{y}_{R_m}\right)^2 + \mathrm{a}|\mathrm{M}|$$

(2)

where "M" is the number of terminal regions of the tree and "R_m" one specific region. $\hat{y}_{R_m}$ is the mean of the response in a specific region, and so the residual is calculated as the squared difference between the observed value and the mean of the response in the region. And this is summed across all of the regions. On the other hand, "a" is a constraint parameter which controls the complexity of the model. As "a" gets larger the tree is pruned and thus becomes smaller, resulting in less overfitting in the training data. The "a" value is user-specified and is estimated through K-fold cross validation.

However, a single tree in most of the cases (such as in our case as shown in table 6) results in worse prediction accuracy than many trees. To that extent, boosting methods can achieve higher accuracy by combining many trees.

Boosting is a special case of tree methods which relies on adding several trees in order to make predictions. Gradient Boosting specifically (our final model), fits respective trees to the negative gradient of the loss function, because the negative gradient is a vector pointing out to the steepest decrease in the loss function. In mathematic notation:

$$-\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)} \qquad \text{for i=1,…,n}$$

(3)

,where $L(y_i, F(x_i))$ is just the loss function, in our case the well-known sum of square residuals, which we differentiate with respect to the predicted values. The negative sign in front marks that we are taking the negative gradient. Also, the subscript $F(x) = F_{m-1}(x)$, means that in each step of the algorithm we update the predictions.

However, it turns out that the negative gradient of the loss function is just a residual (not even a squared one). The original loss function for regression when using Gradient Boost is:

$$\frac{1}{2}(y - \hat{y})^2$$

(4)

If we differentiate it with respect to the predicted value, using the chain rule, it becomes:

$$\frac{\partial}{\partial \hat{y}} \frac{1}{2}(y - \hat{y})^2 = \frac{2}{2}(y - \hat{y})(-1) = -(y - \hat{y})$$

(5)

And when also putting the negative sign in front, according to equation 3, then it simply becomes:

$$(y - \hat{y})$$

(6)

That means that fitting regression trees to the negative gradient is identical to fitting them in just the residuals and the same thing also holds for the classification setting where the loss function used is the negative log likelihood function of the logistic regression procedure. That being said the GBM algorithm proceeds as follows (Hastie, Tibshirani, and Friedman 2009). First it initiates the model with a constant value. Then in step 2 it computes the residuals (r) by differenciating the loss function as showed above. Then, it fits a regression tree $f^m(x)$ to the residuals utilizing the criterion in equation 2. Then, it updates the previous model by adding a new tree and making new predictions like $F_{m-1}(x) \leftarrow F_0(x) + \lambda f^m(x)$. The λ parameter is called the learning rate and it has the aim to avoid overfitting the model at the training data. It is user-specified and it is estimated through cross-validation. After the new predictions, the residuals are recalculated and updated and the second step is repeated over and over. Finally, in the third step there is the output of the finalized additive model. With mathematical notation:

Step 1: $F_0(x) = \text{argmin}_\gamma \sum_{i=1}^{n} L(yi, \gamma)$

Step 2: for m=1,2,…,M, repeat:

Compute $r = -[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}]_{F(x)=F_{m-1}(x)}$, for i=1,…,n

a) Fit a regression tree $f^m(x)$ to the residuals
b) $F_{m-1}(x) \leftarrow F_0(x) + \sum_{i=1}^{m} \lambda f^m(x)$

Step 3: Output the final model,

$F(x) = \sum_{m=1}^{M} F_m(x)$

There are four tuning parameters that there need to be specified by the user in the GBM model. The learning rate λ, the total number of trees to be used in the model, the interaction depth and the number of points in the terminal nodes of the trees. The interaction depth is the number of splits in each tree and thus controls the amount of interaction. Those values are determined after K-fold cross validation. Our tuning parameters for the GBM after using 10-fold cross validation on the training set were 0.1 for the learning rate, 150 for the number of trees, 2 for the interaction depth and 10 for the number of points in the terminal nodes of the trees.

In Table 7, we present the tuning parameters for all of our trained models (including GBM), as well as final selected values. The values were selected using the RMSE criterion.

Table 7: Tuning parameters of the models and final selected values.

Model	Tuning Parameter	Selected Value
GBM	λ number of trees interaction depth number of points in the terminal nodes of the trees	0.1 150 2 10
Ridge Regression	λ	0.83
SVM (Polynomial Kernel)	d C scale	2 0.25 0.01
SVM (Radial Kernel)	γ C	0.01 1
Random Forest	Minimum number of splits per tree	29
SVM Linear	C	1
Tree	alpha	11.15

Table 8: CHES dimensions: descriptions and scale

Dimensions	Description	Scale
Antielite_Salience	Salience of anti-establishment and anti-elite rhetoric	0 = Not important at all : 10 = Extremely important
EU_Position	Overall orientation of the party leadership towards European integration in YEAR.	1 = Strongly opposed 2 = Opposed 3 = Somewhat opposed 4 = Neutral 5 = Somewhat in favor 6 = In favor 7 = Strongly in favor
LRGEN	Position of the party in YEAR in terms of its overall ideological stance.	0 = Extreme left : 5 = Center : 10 = Extreme right

Table 9: V-Party dimensions: descriptions and scale

Dimensions	Description	Scale

Anti-Elitism	How important is anti-elite rhetoric for this party?	0: Not at all important. The leadership of this party never makes statements against the elite. 1: Not important. The leadership of this party rarely makes statements against the elite. 2: Somewhat important. The leadership of this party sometimes makes statements against the elite. 3: Important. The leadership of this party often makes statements against the elite. 4: Very important. The leadership of this party makes statements against the elite whenever possible.
People-Centrism	Do leaders of this party glorify the ordinary people and identify themselves as part of them?	0: Never. The party leadership never glorifies and identifies with the ordinary people. 1: Usually not. The party leadership generally does not glorify and identify with the ordinary people. 2: About half of the time. The party leadership sometimes glorifies and identifies with the ordinary people. 3: Usually. The party leadership generally glorifies and identifies with the ordinary people, which they claim to represent. 4: Always. The party leadership always glorifies and identifies with the ordinary people, which

		they claim to represent.
Populism	To what extent do representatives of the party use populist rhetoric (narrowly defined)? Aggregation: The index is computed as the harmonic mean of rescaled anti-elitism and people-centrism items' posterior distributions.	Interval, from low to high (0-1)

Figure 8: Correlation of the Populism Score with the CHES: Anti-Elite salience item, by party ideology. (Populism Score built using the V-Party dataset).