UNIVERSITÀ DI TRENTO

Dipartimento di Sociologia e Ricerca Sociale

Corso di Laurea triennale in Studi Internazionali

Predictive analysis on populism and liberalism using Random Forest Algorithm in Norway, Hungary and Germany

Prof. Stefano Benati

Filippo Nardi

anno accademico 2020/2021

Index:

The scripts and the databases that I used (as well as this thesis) can be found on my GitHub profile @filipponardi17

Introduction:

In this thesis I will examine a prediction method called Random Forest. With this method I will try to predict which variables defines the liberal and populist individuals in 3 different countries: Norway, Germany and Hungary. Random Forest is a very interesting technique of research, while is known as "a Swiss Army Knife of machine learning algorithms" only recently social researchers have started implementing it. I used the data from the World Value Survey and analyzed them with R using the Package 'randomForest'. What I found was a series of variables that in some cases were the same between the countries and in others underline the difference between these countries, although there are not the signs of perfect patterns between countries, we can appreciate some similarities and some considerations can be drawn.

The idea for this thesis is drawn by the work of Professor Stefano Benati, Università Degli Studi di Trento and Carlo Ruzza, Università Degli Studi di Trento, who used the Random Forest algorithm to predict the populist individuals in France and Poland, arriving at very interesting results similar to this analysis. In this work I will provide the same technique of prediction called Random Forest, it is a statistical learning method introduced by Leo Breiman in 2001. It is extensively used in many fields of application, due for their excellent predictive performance (Poggi, 2020). Additionally, they allow to consider qualitative and quantitative explanatory variables together without preprocessing.

Populism, according to the minimal concept is defined as a thin-centered ideology, which is based not only on the Manichean distinction between 'the pure people' and 'the corrupt elite', but also on the defense of popular sovereignty at any cost (Mudde & Kaltwasser, 2017). On the contrary, Liberalism is defined as a political or social philosophy advocating the freedom of the individual, parliamentary systems of the government, nonviolent modification or political, social, or economic institutions to assure unrestricted development in all spheres of human endeavor, and governmental guarantees of individual rights and civil liberties (Collins English dictionary 2020).

Random forests are a set of decision trees that have been generated using a random subset of data. In fact, the name "Random Forest" comes from combining the randomness that is used to choose the subset of data with having a group of decision trees, thus a forest (Hartshorn, 2016). The data that I used belongs to the World Value Survey wave 7, it comes from interviews that were conducted between 2017 and 2021, with only a dozen of countries that have taken the surveys during the pandemic. All countries employed random probability representative samples of the adult population.

The vast majority of surveys were conducted using face-to-face interview (PAPI/CAPI) as the data collection mode.

The countries that I analyzed are Norway, Hungary and Germany. I specifically chose these countries as all three are in Europe, although Norway is not the in the EU, its two borders are with Sweden and Denmark, both members of the EU. Additionally, Norway was a founding member of the European Free Trade Association, which was the predecessor of the European Economic Community; therefore, Norway holds a tight position in Europe. While Norway is not a Member of the European Union, both Germany and Hungary are part of it. Economically these countries are very different from each other, subsequent data will be expressed in US dollars and come from Data World Bank in 2020. Norway is one of the highest countries in GDP per capita with 67,329.68 dollars per capita and a GDP of 362 billions of dollars. Germany meanwhile has a GDP per capita of 46,208.43 dollars, while having a GDP of 3,806 trillions, as it is the fourth highest GDP in the world. Vice versa Hungary has a GDP per capita of 15,980.74 and a GDP of 155 billions. We can see here that both Norway and Germany have a strong economy for their population, while Hungary have a weaker economy.

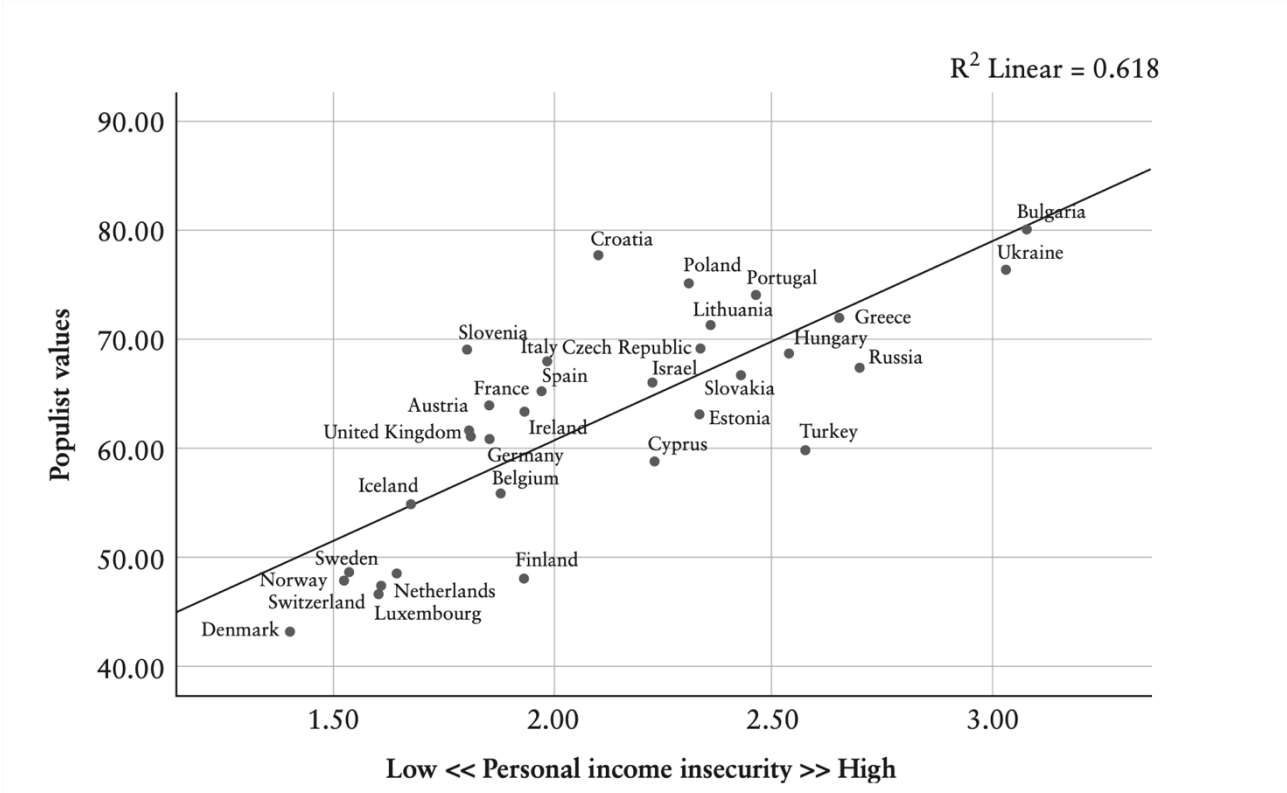

Figure 1

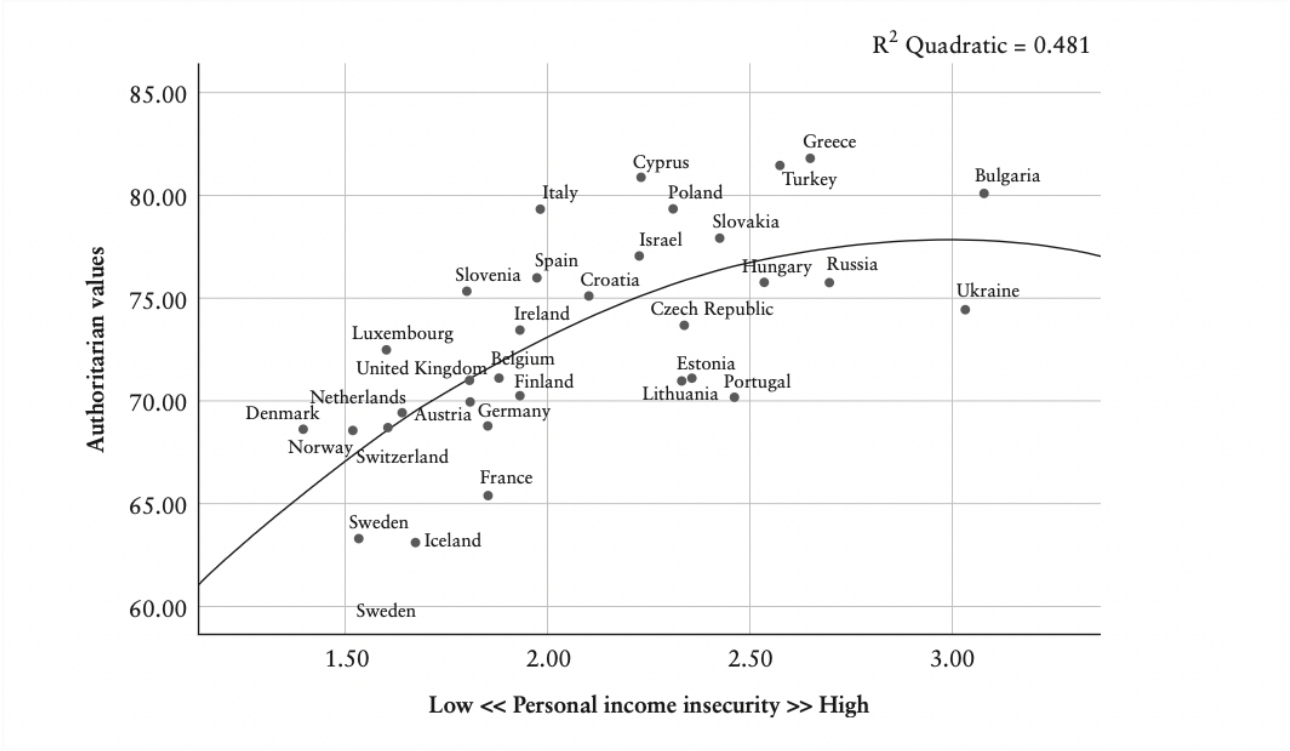

Figure 2

These graphs, provided by Norris and Inglehart book "Cultural Backlash Trump, Brexit, and Authoritarian Populism", shows how higher income insecurity its generally linked to a higher probability of having a populist or authoritarianism ideology (Authoritarian values are expressed in the Schwartz scale, populist values are expressed in Mistrust of parliament, parties and politician. Data is standardized and come from The European Social Survey, Cumulative File Rounds 1–7). Cultural values are indeed found among those people most likely to feel a sense of economic insecurity and grievances. (Norris & Inglehart, 2020). While authoritarian values are substantially predicted by generational cohorts more than by any of the economic indicators, both economic and cultural factors are significant for predicting populist attitudes (mistrust of politicians, parliaments, and parties) particularly the subjective economic characteristics (feelings of economic insecurity and especially approval of the state of the national economy) are more important than birth cohort. The other factor I took in consideration for choosing the country to analyze, is the current government and the degree of progressiveness in each country. Norway is in fact the most progressive country in the world according to the 2021 social progress index, it is ranked first with a score of 92 out of a 100. Vice versa Hungary is ranked as 42nd in the world with a score of 80 out of a 100, one of the lowest scores in the European Union. Germany is placed well at the 10th position in the world with a score of 90 out of 100. Interestingly the political situation in these countries only partially represent the

situation reported by the social progress index. This because although Norway has a very progressive parliament and government, Hungary with its leader Viktor Orbán is known for pursuing a populist ideology, exemplary it is his anti-EU political battle. Meanwhile Germany, at the time of data collection, was in the last years of the Angela Merkel government. She began her mandate as prime minister the country in 2005 and only recently, December 2021, she completed her mandate. Merkel was the leader of the political party Christlich Demokratische Union Deutschlands (CDU shortened), which is a Christian – liberal – conservative government. In fact, during her mandate, we can remember an initial opposition to same sex marriages in Germany (Thomson, 2017).

In the first paragraph of the thesis, I will explain the analytical framework of this thesis, such as the R scripts I used and how the data was reduced. While afterwards I will spend a paragraph for each country analyzed, here I will comment relevant results of the random forests and the data. In the conclusion I will comment the whole research and the most valuable results.

Chapter 1: Analytical Framework

As stated previously, this analysis is conducted using the World Value Survey, Wave 7, then the data is analyzed using the software R, specifically through the use of the package 'randomForest'. Firstly, I retained all the surveys regarding opinions and values and the variables regarding demographic data about the interviewed. Lately, I discarded variables deemed as irrelevant, such as the job of the ancestors and a few others to remain with a total 119 survey questions. After being remained with the variables regarded as more important, I cleaned the data to make it understandable for R and easier to deal with. For example, the variable called "E069_11", later renamed as c_gvr, is the code for the question on confidence of the Government of the country. The question of the variable is "Please look at this card and tell me, for each item listed, how much confidence you have in them, is it a great deal, quite a lot, not very much or none at all?". The positive values (greater than one) are on a scale from 1 (A great deal) to 4 (None at all), but the database also give us negative values from -1 to -5. These are the codes for: -5 Missing: Other, -4 Not asked in survey -3 Not applicable, -2 No answer, 1 Don´t know. I erased the negative values from the database as I only took in consideration positive values for this analysis, which is people who responded to the question with a clear value. Subsequently in R using the function "as.numeric" I converted the values into numeric values, in order to be able to create means and use the values to do other calculations. As mentioned before, in this phase I also took the time to rename some variables in order to make them more humanly understandable; for example, the variable named "E069_11" in the wave 7 codebook was renamed as "c_gvr", which is shortened for "Confidence in the government". The most important variable for this analysis is "E181_EVS5", that responds to the question "Which political party appeals to you the most?". This variable will be the so called "leaf" of the random forest; the term "leaf" indicates what is trying to be predicted in a decision tree and as stated before, a random forests are many decision trees randomly generated. In this case, since the goal is to predict the liberal individual, I will evaluate a liberal individual according to the vote decision. With this I am taking as an assumption that a liberal individual will vote for a liberal political party. I only took in consideration parties that have gotten at least 2% of preferences in the most recent elections (apart from the 3rd of April 2022 Hungary elections, in that case I considered the 2018 results). The ideology of the political parties is divided as following; the classification is based mostly on Inglehart and Norris book "Cultural Backlash Trump, Brexit, and Authoritarian Populism" and other sources that will be cited following.

The acronym POP stands for populist party, LIB stands for liberal party and NA stands for Non Available as either that party no longer exist or it hasn't gotten at least 2% preferences in the last

election. The code at the left of stands for the classification code used by the World Value Survey, Wave 7 database.

The party classification for Norway is as it follows:

57801 NO: Labour Party	**LIB**
57802 NO: Democrats in Norway	**NA**
57803 NO: Progress Party	**POP**
57804 NO: Conservative Party	**LIB**
57805 NO: Christian Democratic Party	**LIB**
57806 NO: Coastal Party	**NA**
57807 NO: Green Party	**LIB**
57808 NO: The Christians	**NA**
57809 NO: Pensioners' Party	**NA**
57810 NO: Red Party	**POP**
57811 NO: Centre Party	**POP**
57812 NO: Socialist Left Party	**LIB**
57813 NO: Liberal Party	**LIB**

The party classification for Hungary is as it follows:

34801 HU: Hungarian Socialist Party (MSZP)	**LIB**

34802 HU: Fidesz	**POP**
34803 HU: Christian Democratic People's Party (KDNP)	**POP**
34804 HU: Movement for a Better Hungary (Jobbik)	**POP**
34805 HU: Politics Can Be Different (LMP)	**LIB**
34806 HU: Democratic Coalition (DK)	**LIB**
34807 HU: Together – Party for a New Era (Együtt)	**NA**
34808 HU: Dialogue for Hungary (PM)	**LIB**
34809 HU: Momentum Movement (Momentum)	**LIB**
34810 HU: Hungarian Liberal Party (Liberálisok)	**NA**
34811 HU: Hungarian Two-tailed Dog Party (MKKP)	**NA**
34812 HU: Hungarian Workers' Party (Munkáspárt)	**NA**

The party classification for Hungary is as it follows:

27601 DE: Christian Democratic Party/Christian Social Union	**LIB**
27602 DE: German Social-Democratic Party	**LIB**
27603 DE: German Liberal Party	**LIB**
27604 DE: The Green Party	**LIB**
27605 DE: The Left	**POP**
27606 DE: Alternative for Germany	**POP**
27607 DE: Other, please specify (WRITE IN)	**NA**

27666 DE: No [no other] party appeals to me (spontaneous) **NA**

Other sources for the party classification were http://popu-list.org, updated by the University of Amsterdam or https://www.globalpartysurvey.org or otherwise from my personal interpretation.

Subsequently after having coded all the political parties I started working on R. For each country I created a specifical database regarding only the specific country. In fact, there will be 3 specific database that I will be working with: W7_no, W7_hu and W7_ge. These databases will be once again reduced considering only who declare his vote and thus not considering non available votes, these new databases will be called WnoR, WhuR and WgeR.

I implemented the random forest as follows: at first, I performed the so called "training" for the random forest trying to predict populist and liberal voter using all the 118 variables. Then, according to the Mean Decrease Index and the Mean decrease Gini index I used the most occurrent variables in the three random forests, one for each country, coupled with another factor that I wanted to investigate.

Mean decrease accuracy is the percent of observations classified correctly. The decrease in accuracy is a measure of how much greater the predictive accuracy of the row variable in the variable importance table is compared to prediction with a random (but proportional) variable. This coefficient may be taken as a measure of the importance of the row variable as a predictor (as a parent node). Note that this criterion compares the importance of the variable with a proportional random variable, not with other variables in the model (Garson, 2021). Mean Decrease in Gini is the average (mean) of a variable's total decrease in node impurity, weighted by the proportion of samples reaching that node in each individual decision tree in the random forest. This is effectively a measure of how important a variable is for estimating the value of the target variable across all of the trees that make up the forest. A higher Mean Decrease in Gini indicates higher variable importance (Louppe et al, 2013).

Using the feature importance data of the Random Forests I gathered the first nine important variables, according to Mean decrease accuracy, plus the variable income level, which I wanted to investigate, and constructed a random forest classifier for each country.

Chapter 2: Norway prediction

Norway elects its legislature on a national level, the parliament has 169 members elected for a four-year term by a form of proportional representation (Cirone et al, 2021). This is fundamental for this analysis as a proportional system usually means that there are more different parties. In fact, the world value survey lists 13 parties, out of the 13 parties I listed 4 of them as non-available as they have not reached the 2% in the last elections. Using the command table(W7_no$pop_vote, useNA = "always") we see the following results:

lib pop <NA>

773 232 117

This means that there are 773 individuals who are classified as liberal voters, 223 classified as populist voters and 117 non available. Performing the training of the random forest using all 118 variables the 9 top variables in the training according to mean decrease accuracy index are:

1. Group 1. Anti-Cosmopolitanism, this is group is composed by:

    - **c_eu1**: How much confidence you have in the European Union? (1 = A great deal, 4 = None at all)
    - **city_siz**: Size of town where interview was conducted (5 categories, 1= under 5000, 5= 500000 and more)

2. Group 2. Political Affinity. This group is composed by:

    - **Dmc_cnt**: How democratically is this country being governed today? (1=not democratically, 10 = fully democratically).
    - **C_gvr**: Confidence in the government (1 = A great deal, 4 = None at all)
    - **pol_rad**: Political radicalization. It measures the distance between voter's political position and the centre of the political space. It is measured as $(p_i - 5)^2$, where p_i is the variable pol sp.

3. Group 3. Religious Values:

- **Js_div**: How much justifiable is divorce? (1 = Never, 10 = Always)
- **cq_obd**: Is child obedience an important quality that should be taught at home= (1= mentioned 0= not mentioned)

4. Group 4. Political vindications:

    - **imp_imm**: Evaluate the impact of immigrants on the development of Norway (1 = very bad, 5 = very good).
    - **job_ntn**: Employers should give priority to Norwegian people over immigrants? (1= Strongly agree, 5=Strongly disagree)

To these groups I am going to add a 5th group, which will be:

5. Income level:
    - **Incm_lvl**: What group your household is according to incomes? (1= 1st decile, 10= 10th decile)

In the training the OOB estimate error rate is 22.39%. OOB is the acronym for Out Of Bag (error rate), the reason it is important is that it gives an estimate of how good the random forest is without using any additional data, and without using a set aside data set. This because Random Forests inherently set aside some of the data. When each tree is generated from the dataset it uses, on average, only 63.2% of the dataset. So, the remaining 36.8% is implicitly set aside. The data that is not being used for any given tree is known as "Out of Bag" for that tree (Hartshorn, 2016). Thus, the out of bag data is ran in the Random Forest to evaluate the predictions, doing this ends up in an error rate of 22.39% in this case. Subsequently confusion matrix regarding the training random forest:

Confusion matrix:

	lib	pop	Class error
lib	773	0	0.53947368
pop	225	7	0.9698276

This matrix gives us the misclassified individuals. Out of the 773 liberals none is being misclassified with an error of 0%. Regarding populist voters out of the 232 populist voters 225 are being misclassified as liberals with an error rate of 97%. This prediction model tends to classify nearly every voter as liberal and is therefore inconsistent.

Below I will post the graphs of the training.

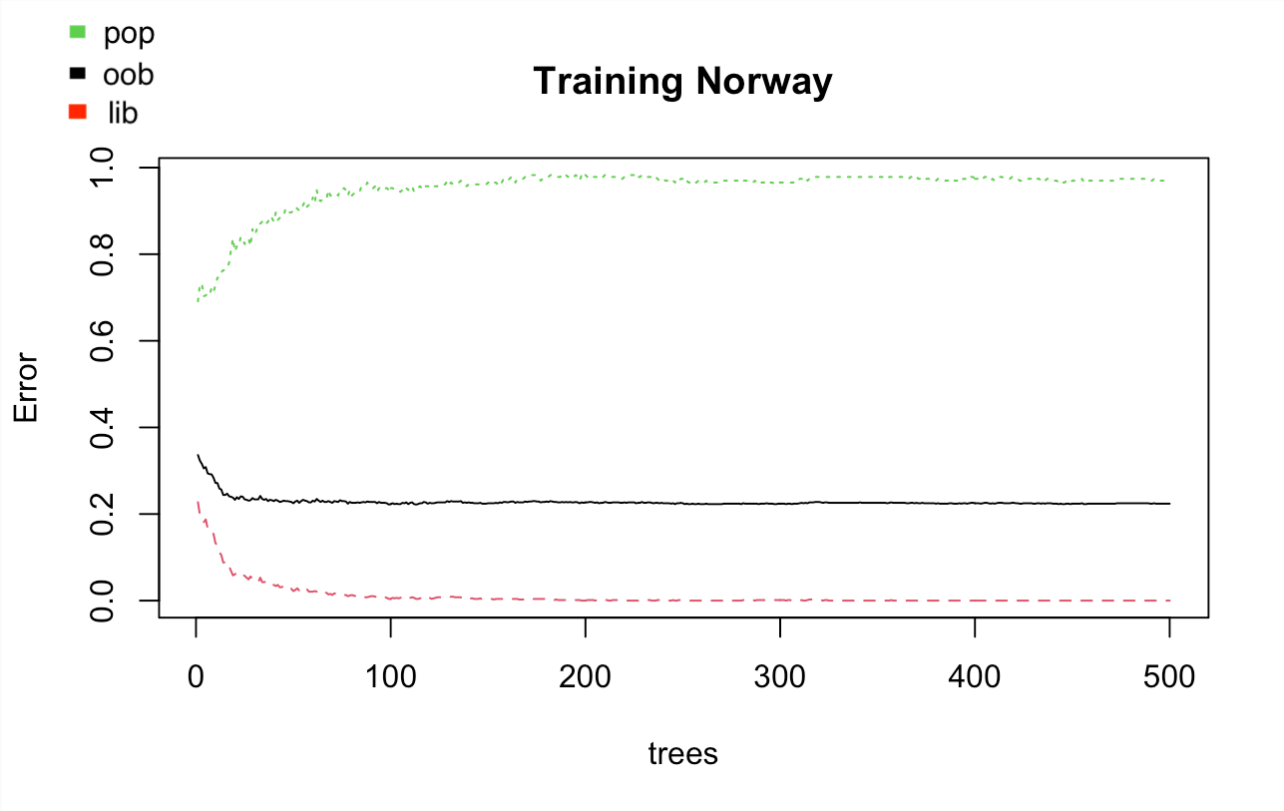

Figure 3

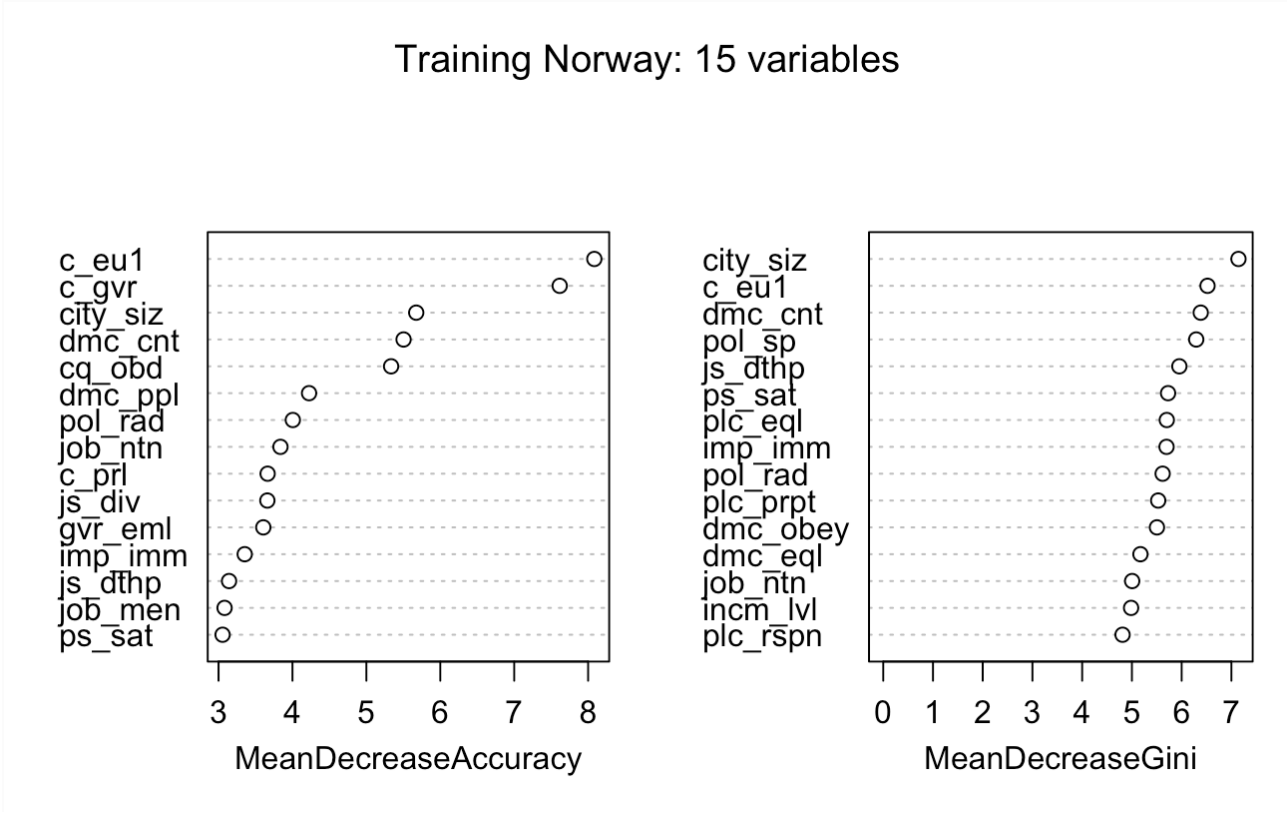

Figure 4

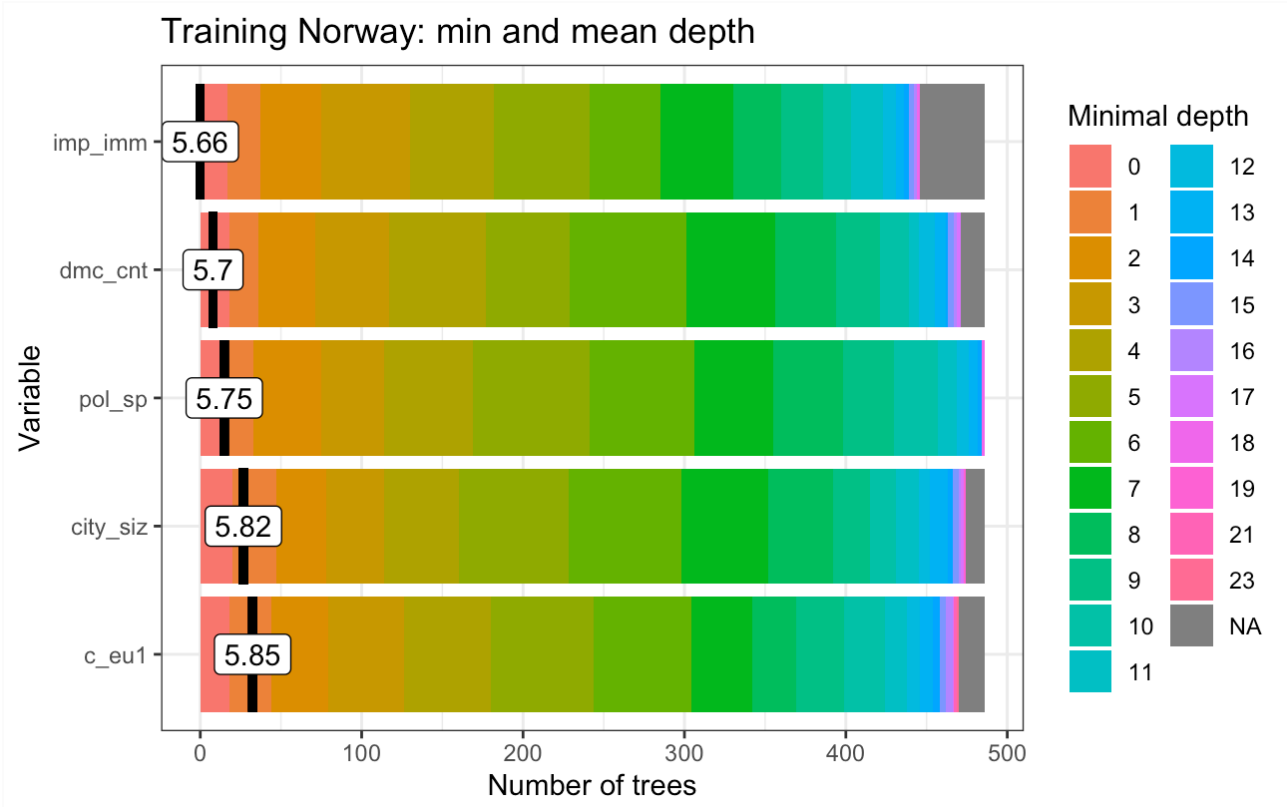

Figure 5

Doing the actual random forest, I will use, as aforementioned, only 10 variables, as opposed of the 118 of the training. The results are:

OOB estimate of error rate: 23.48%

Confusion matrix:

	lib	pop	Class error
lib	724	49	0.06338939
pop	187	45	0.80603448

As we can see the results are not changed heavily. The out of bag error is only 1% higher compared to the training while the confusion matrix is a bit more different. The liberal class now has 6% error rate, compared to the 0% of the training, while the populist class has a 16% upgrade, going from a 97% class error to a 81% class error. Below we can see figures of the error rates, OOB and confusion matrix, the variable importance plots and minimal and mean plot distribution.

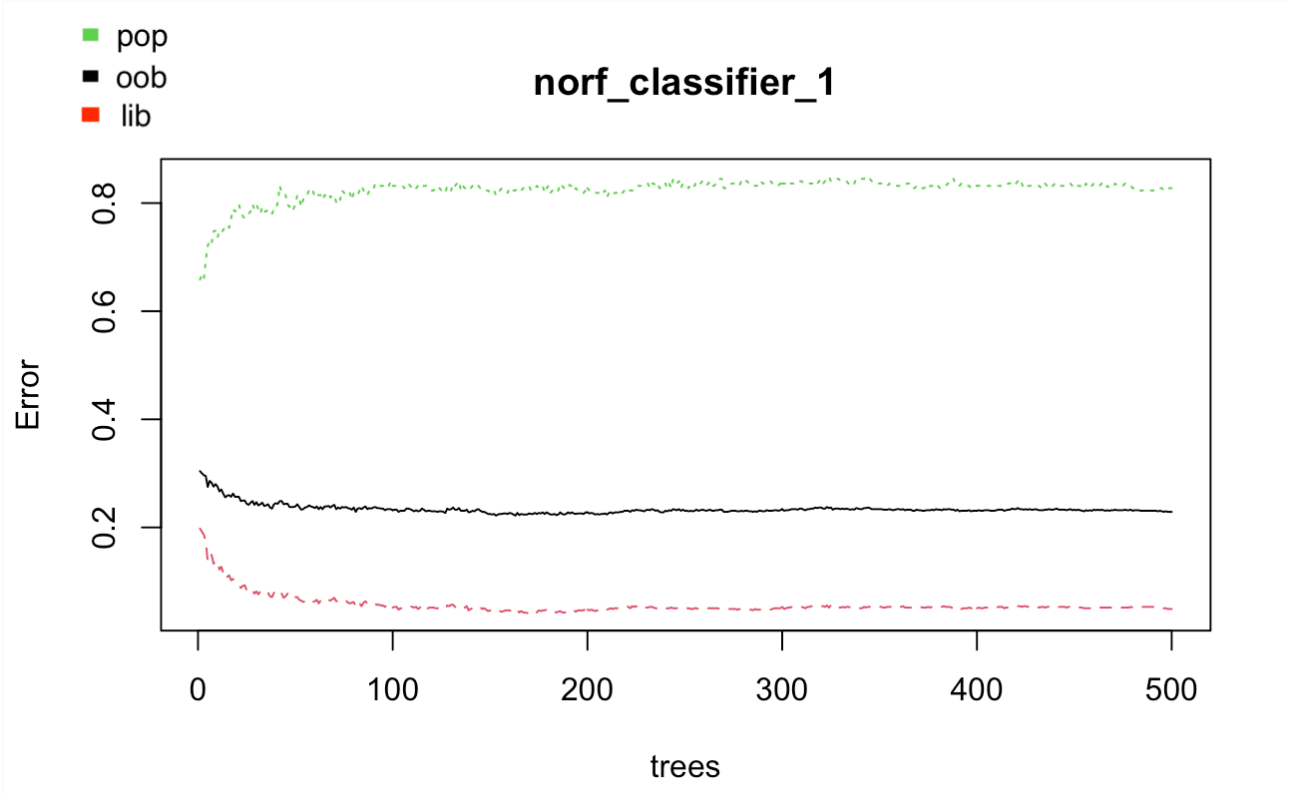

Figure 6

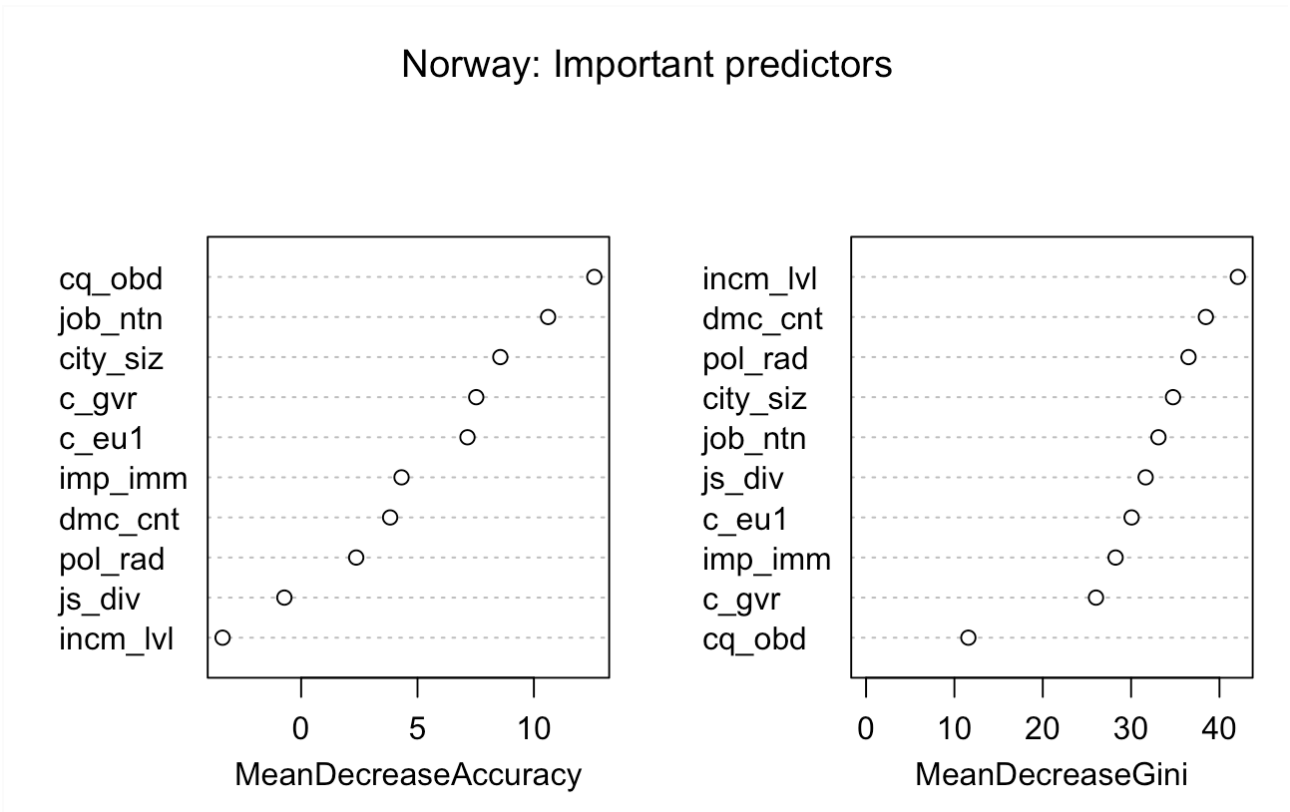

Figure 7

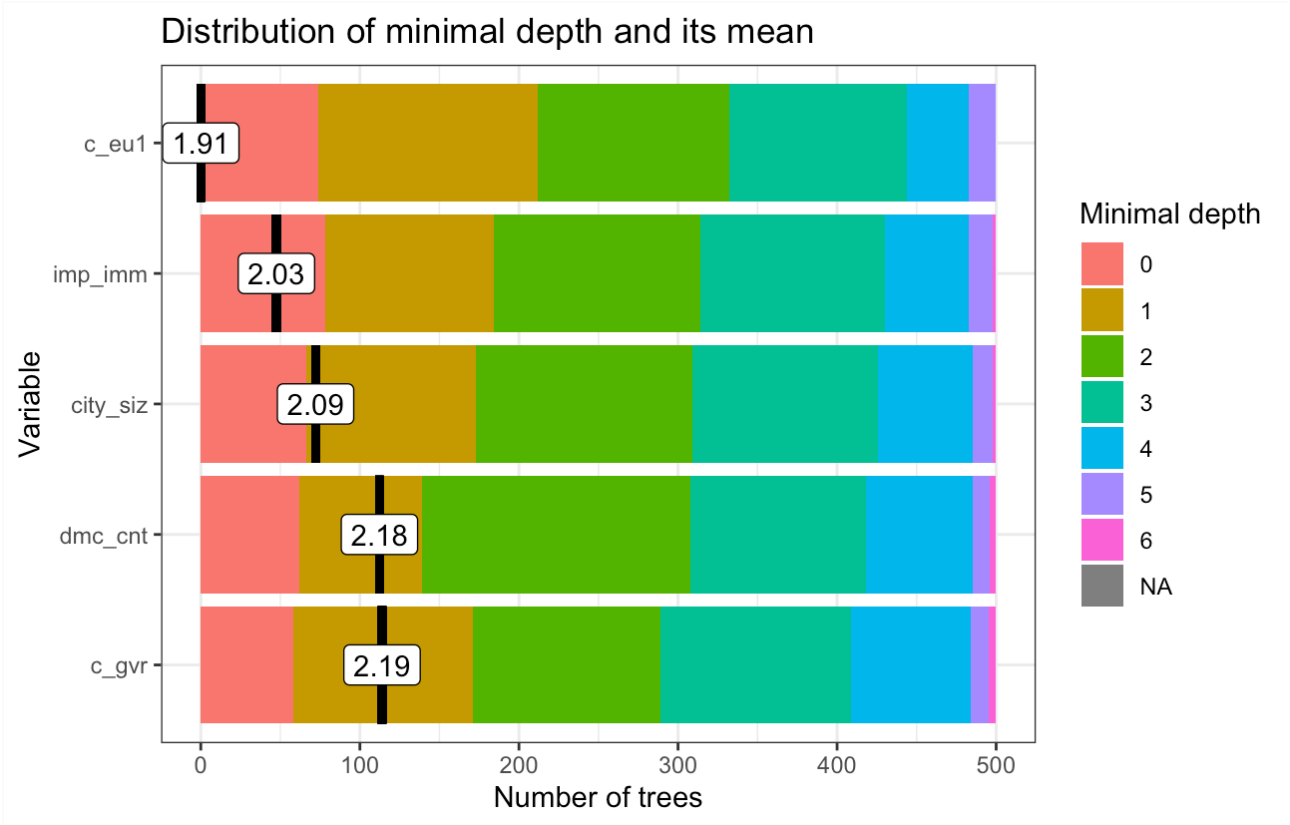

Figure 8

Minimal depth is a feature that indicates how close to the root of the decision tree a variable was. This plot provides another way to evaluate variable importance. “Depth” for a variable is how many nodes down in the tree are located, starting numbering at 0 for the root node. “Minimal depth” is the lowest depth value (hence highest in the tree) for the first instance of a given splitting variable. “Mean minimal depth” is minimal depth for a variable averaged across all the trees in the forest. Moreover, mean minimal depth is an indicator of variable importance since variables with lower mean minimal depth are higher in the tree, and thus serve to sort more observations (Garson, 2021).

Using the function “tapply” in R we can see the mean of the variables for both populists and liberals. From these figures we cans see how child obedience is the principal variable in terms of mean decrease accuracy, while income level is the principal variable regarding of mean decrease Gini, the variable which stay closer to the root, thus being the highest in minimal depth distribution, is confidence in the European union.

Regarding group 1:

Confidence in the European union has a μˆlib= 2.49 and a μˆpop= 2.90, a difference in of 0,50 out of 4 but we can see that the mean of both liberals and populist is still fairly high, this could be one of the reasons that Norway never joined the European Union. This is also consistent with all previous analysis of populism, where national identity implies a rejection of supranational identity, in this case the European Union and this finding supports the argument that populism is correlated with a form of euro-skepticism (Harmsen, 2010). Consistently, city size shows that liberals tends to live in more big cities, in contrast to smaller cities where populists lives, μˆlib= 3.03 and μˆpop= 2.46. This is congruous with the fact that liberals tends to live in large cities as it is connected with cosmopolitanism (Maxwell, 2019), where populists indeed tends to live in smaller cities as a form or rejects of cosmopolitanism.

Regarding group 2:

The second variable in terms of Mean Decrease Gini is dmc_cnt, the μˆlib= 8.18 and the μˆpop = 7.43. As the current Norway's government is very liberal, we can see how there is discontent from the populists in the country. Confidence in the government has a μˆlib=2.29 and the μˆpop = 2.51, a rather small difference but it shows the same phenomena as the previous variable where populists tend to have more discontent than liberals. Political radicalization has a μˆlib=4.44 and a μˆpop = 6.44, this also showing that liberals tends to have a lower political radicalization than populists voter.

Regarding group 3:

In the justifiability of divorce lib have a μˆlib=8.26 and a μˆpop = 7.88, for the child obedience the μˆpop = 0.22 while for liberals μˆlib= 0.12. This shows that liberal are less likely to have religious values, while populists have stronger opinion in particular towards the justifiability of marriage.

Regarding group 4:

The second variable according to Mean Decrease Accuracy, is the favoring national jobs over immigrants; here we can see how a μˆlib= 3.69 while the μˆpop = 3.11, here populists are more likely to agree with the affirmation. This is also another common theme among populists, where they

advocate for national identity with emphasis on ethnicity while liberals, in contrast, tend more towards an evolution of national identity and open borders (Charvet, 2018). We see the same thing in the impact of immigrants on the development of the country, here we have a μˆlib= 3.59 while a μˆpop = 3.17 showing that immigrants are less of a threat to liberals than populists.

Regarding group 5:

Regarding income level the μˆpop=5.12 while the μˆlib=4.67, therefore we can see a substantial difference in income level between populists and liberals. This is actually an antithesis to what is theorized in literature, as several researchers found that populist are linked to having a lower income. (O'Connor, 2017).

Finally, a graph multidimensional scaling graph of the proximities of the random forest.

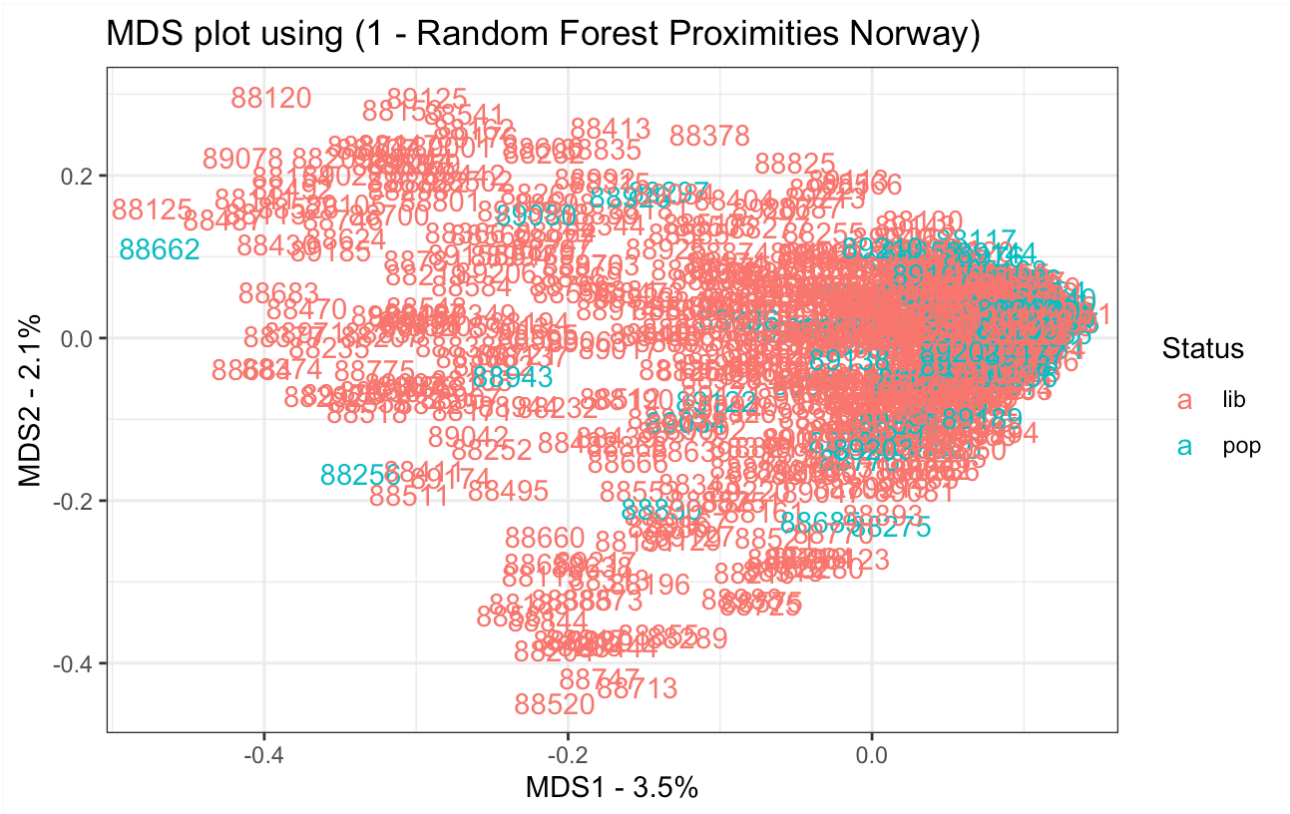

Figure 9

From this figure we can see that there is not a perfect pattern discerning liberals and populist in a visual way. The vast majority of the respondents are concentrated in the middle right part, while we see outliers' values in the respondents such as respondent 88662 in the middle left part of the graph. Seeing those values, we should think if it's a real value of it could be manipulated in any sense (maybe the respondent chose the party political randomly) as it is quite literally surrounded by libera votes.

Chapter 3: Hungary Prediction

The 199 members of the Hungarian National Assembly are elected using a mixed-member majoritarian voting system, with voters having two votes. 106 members are elected by First Past the Post in single-member constituencies and the remaining 93 seats are decided by a closed list PR system in a single national constituency (Kovács & Vida, 2015). This fact should technically mean that we should see less parties, but that's not the case as the WVS indicates 12 parties, out of these 4 are listed as non-available as either they have not reached 2% preferences in the 2018 elections or as they do not exist anymore. At the time of data collection the government in charge in Hungary was populist, with Viktor Orbán as prime minister and with his party Fidesz–KDNP having 44% of the preferences.

Before performing training Hungary had a particularity, oppositely as Norway we can see here that we have 228 liberals and 707 populists. This clearly shows that populist will be in the majority, the complete opposite of Norway. The other particularity is that we have 579 non available, only about a 100 come from NA parties, the others probably responded to the questionnaire with a negative values such as -1 don't know or -2 no answer.

Performing the training I used all the 118 variables, we found that the Hungary training ends up with a respectable OOB error rate of 15.19%, 6% lower than the Norwegian one. The confusion matrix is as follows:

Confusion matrix:

	lib	pop	Class error
lib	105	123	0.53947368
pop	19	688	0.02687412

Additionally, the confusion matrix seems to be somewhat balanced, oppositely but same as the Norwegian confusion matrix it tends to predict most of the people as populist; more than half of the liberals are predicted as populists and nearly every populist voter is predicted as a populist. The class error for liberals is in fact 53% while for populists is 2%.

The first 9 variables in the training according to the mean decrease accuracy index are:

1. Group 1: Political Affinity. This group is composed by:

   - **Dmc_cnt**: How democratically is this country being governed today? (1=not democratically, 10 = fully democratically).
   - **Ps_sat**: how satisfied are you with how the political system functioning in your country today? (1 = not satisfied, 10 = completely satisfied).
   - **Pol_sp**: Self-positioning in the political space (1=left, 10 = right)
   - **C_gvr**: Confidence in the government (1 = A great deal, 4 = None at all)
   - **C_prl**: how much Confidence you have in the Parliament (1= A great deal, 4= None at all)

2. Group 2: Anti-Cosmopolitanism:

   - **C_eu1**: How much confidence you have in the European Union? (1 = A great deal, 4 = None at all)
   - **C_onu**: How much confidence you have in the united nations (1 = A great deal, 4 = None at all)

3. Group 3: Political Vindications:

   - **Imp_imm**: Evaluate the impact of immigrants on the development of Hungary (1 = very bad, 5 = very good).
   - **post_mat**: Post-materialism Index (1=materialist, 3=post materialist)

to these variables I am going to add a tenth variable which is income level:

4. Group 4: Income level:

- **Incm_lvl**: What group your household is according to incomes? (1= 1st decile, 10= 10th decile)

Below I will post the graphs of the training.

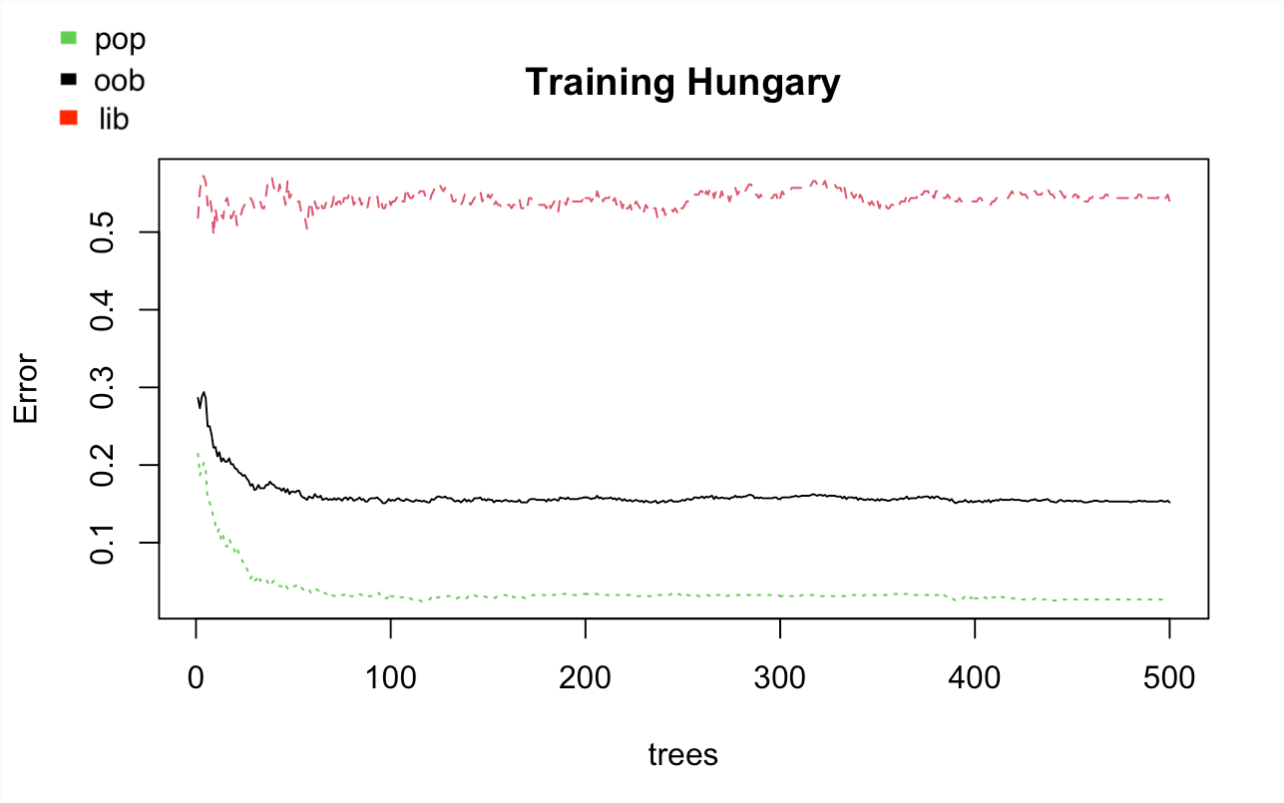

Figure 10

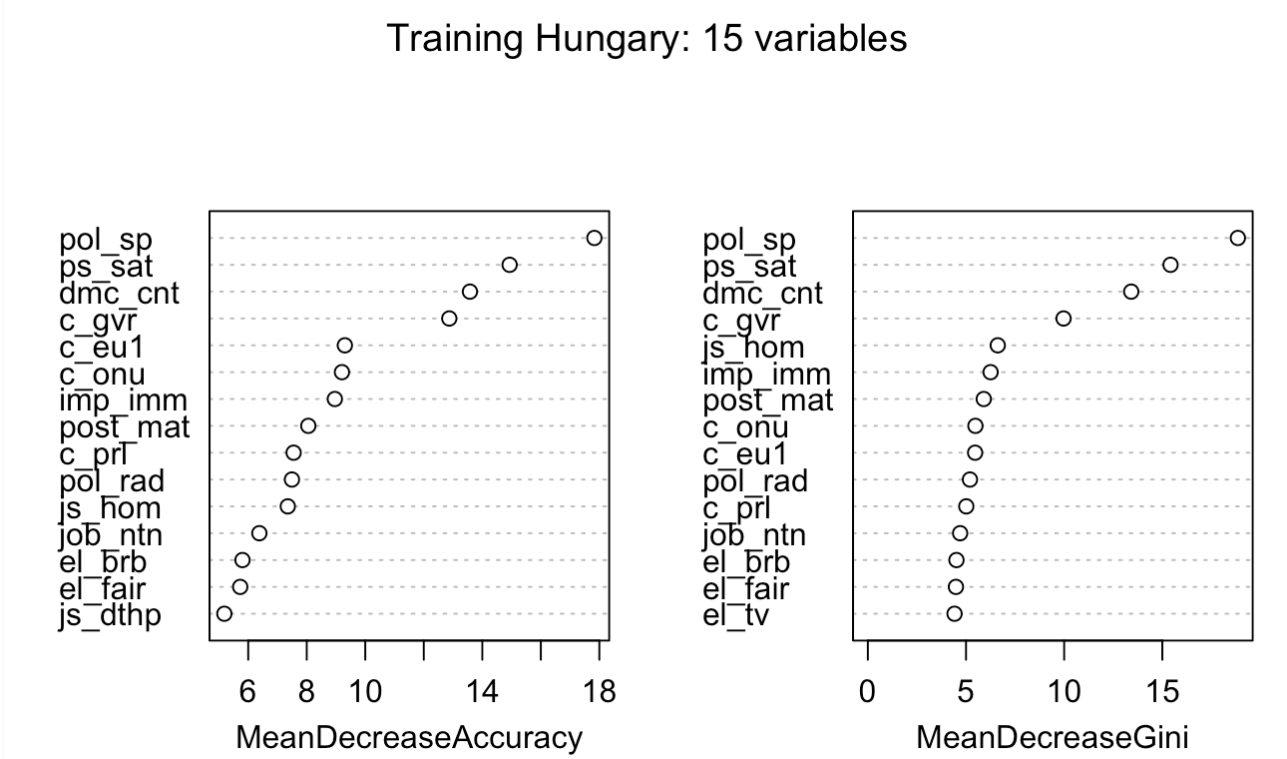
Training Hungary: 15 variables
pol_sp
ps_sat
dmc_cnt
c_gvr
c_eu1
c_onu
imp_imm
post_mat
c_prl
pol_rad
js_hom
job_ntn
el_brb
el_fair
js_dthp
6
8
10
14
18
MeanDecreaseAccuracy
pol_sp
ps_sat
dmc_cnt
c_gvr
js_hom
imp_imm
post_mat
c_onu
c_eu1
pol_rad
c_prl
job_ntn
el_brb
el_fair
el_tv
0
5
10
15
MeanDecreaseGini

Figure 11

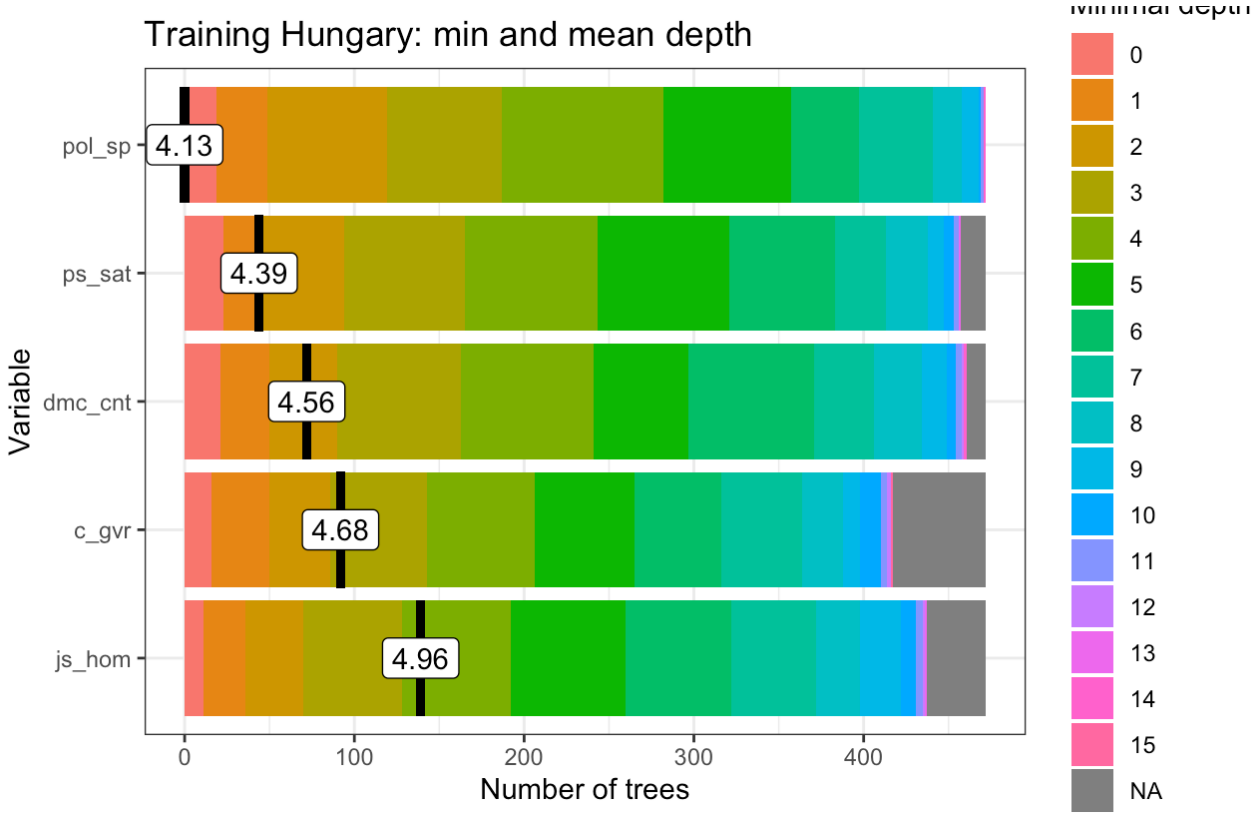

Figure 12

Doing the actual random forest, I will use, as aforementioned, only 10 variables, as opposed of the 118 of the training, the random forest will be called “hurf_classifier_1”. We can see nearly a 3% upgrade in the OOB error rate , going from a 15.19% error to a 12.83% making it more reliable in that sense. The confusion matrix is as follows:

	lib	pop	Class error
lib	147	81	0.35526316
pop	39	668	0.05516266

In the confusion matrix we can see a massive improve in the class error of liberals going from more than 50% to a 36%. While for the populists we see a negligible 2% downgrade going from a 3% error rate to a 5%. Overall, it seem like this Random Forest should be more accurate than the training one. I will post the Hungary classifier graphs here below:

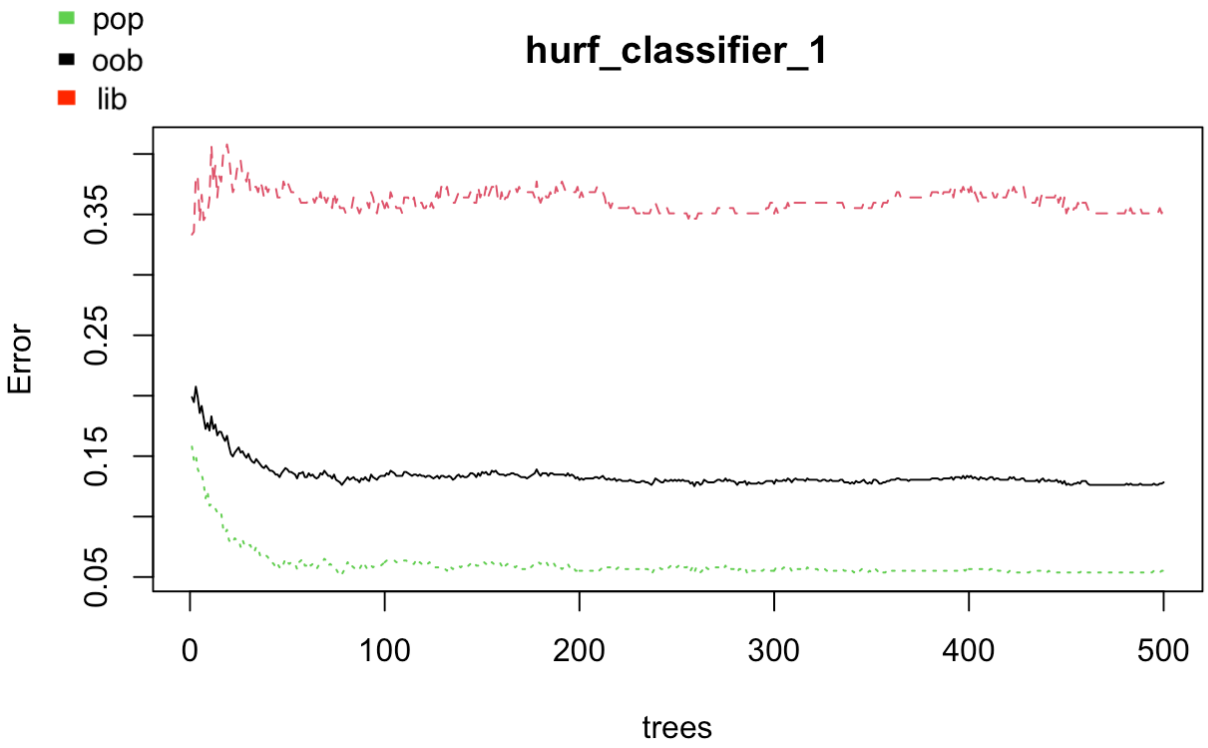

Figure 13

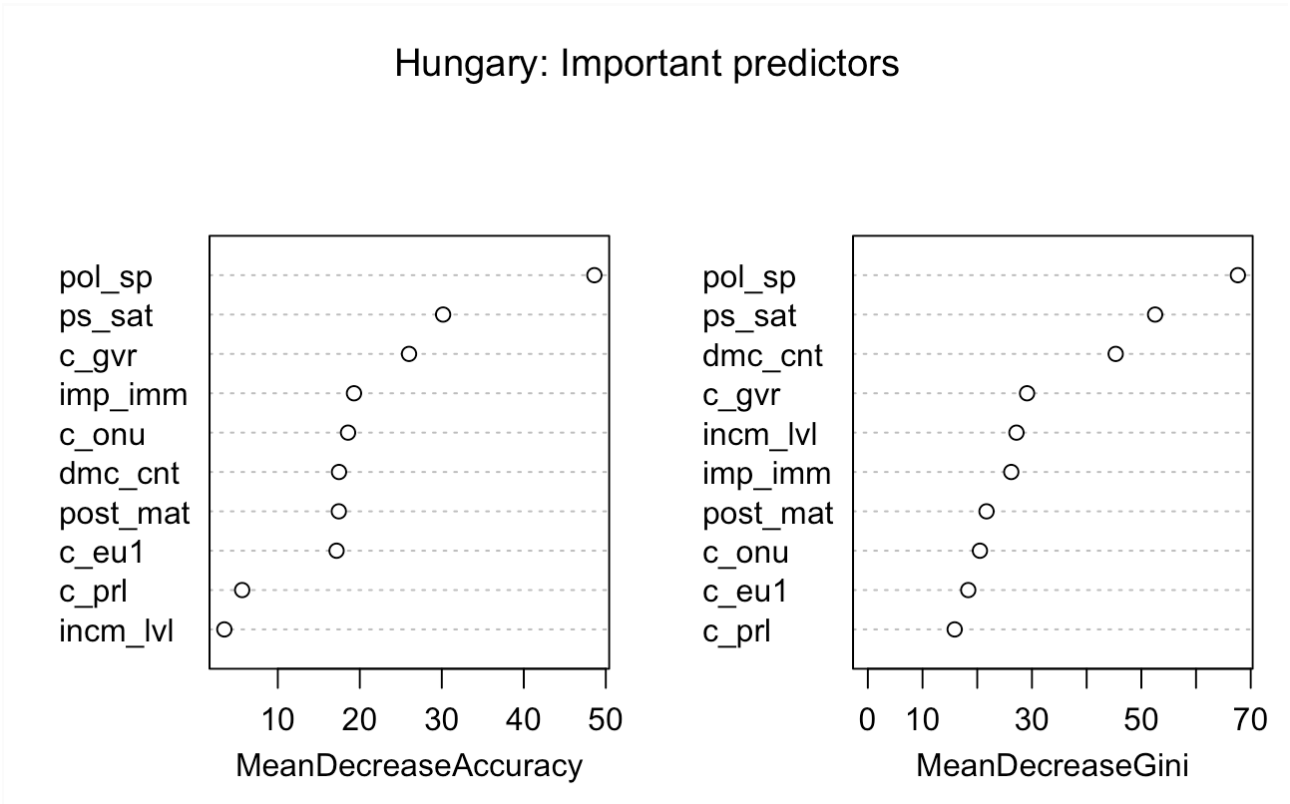

Figure 14

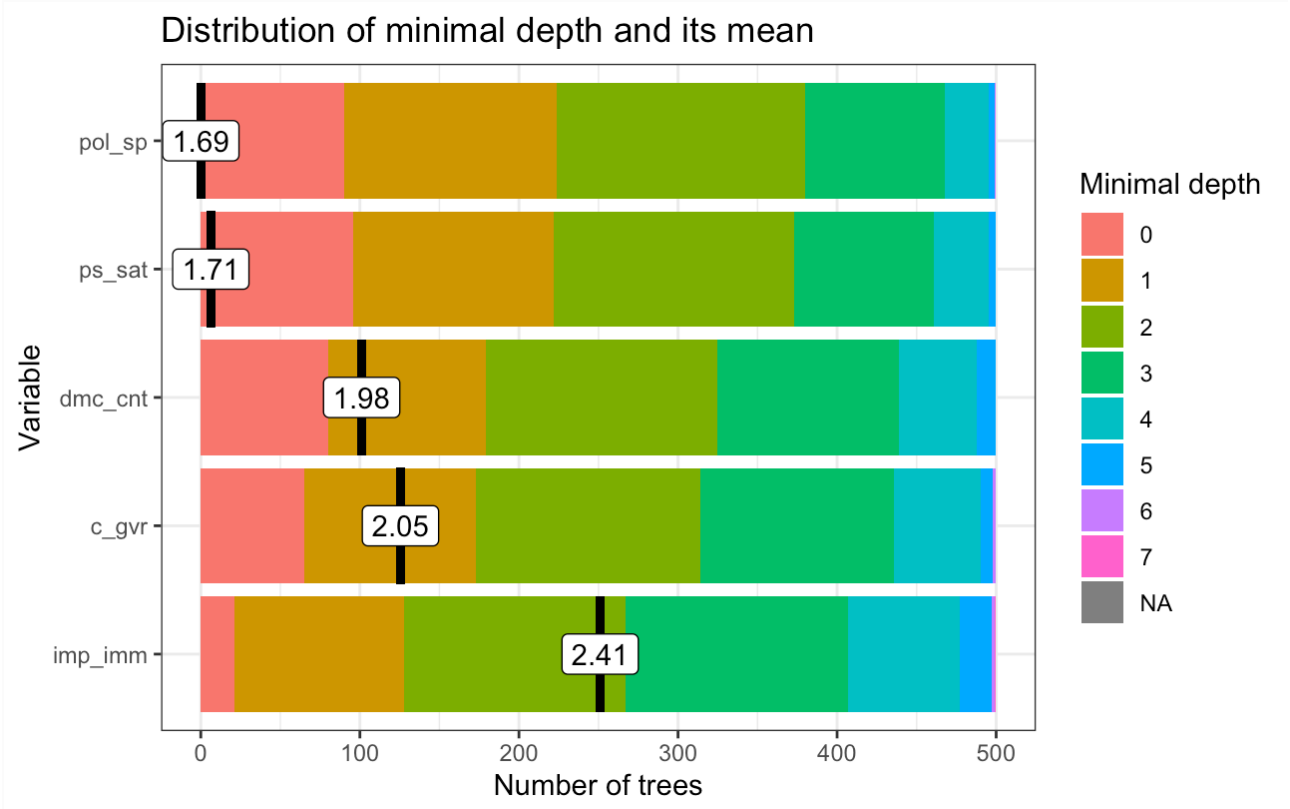

Figure 15

As we can see from the graphs the most important variable is by far pol_sp, which is equal to Self-positioning in the political space, it results as first in Mean decrease Accuracy, in Mean decrease Gini and also in Minimal plot distribution. The same thing also applies for ps_sat, satisfaction in the political system, which is second in all three of the variables.

Regarding group 1: The variable how democratically is your country has a μˆlib= 3.72 and the μˆpop = 6.59, according to liberals the country is not run in a democratic way, the opposite for populists. Since the government is run by populist parties, we can how that is reflected with strong opinions in the electoral body. The satisfaction on the political system has a μˆlib= 2.81 and the μˆpop = 6.08, similar values as before for the reasons. These values show us an interesting finding: populism is often associated with rejecting representative democracy in favor of direct democracy, such as referendums and similar appeals to the will of the people. The findings that populist voters have confidence in government and are satisfied with democracy when in power suggest that public opinion ideology is helpful and conditional on political debate. In other words, the will of the people is betrayed precisely when the populist party fails to win the election. The political space of the electoral body in Hungary is as follows, μˆlib= 4.27 and μˆpop = 7.04, this shows as a right-wing

tendency for the populist voters and a center-left wing for liberal electors. The confidence in the government shows similar results with the other variables, showing liberals very dissatisfied with the current government, while the opposite applies for populists. Finally, confidence in the parliament shows again similar results as previous variables as μˆlib= 3.12 and μˆpop = 2.45, once again populist feel more confidence towards a populist parliament.

Regarding group 2: The variable confidence in the European Union displays a μˆlib= 2.14 and μˆpop = 2.69, showing more confidence towards the European by the liberals than the populists. Liberals are in fact associated with the embracing of cosmopolitanism, this variable being an example of that. Confidence in the United Nations has a μˆlib= 2.18 and μˆpop = 2.74, showing a mistrust towards United Nations. This is a typical populist feature as they mistrust in supranational entities favoring national interests (Norris & Inglehart, 2019)

Regarding group 3: The evaluation of the impact of immigrants on the development of Hungary has μˆlib= 2.76 and μˆpop = 2.10, once again liberals are less concerned with immigrants while the populists have strong opinions towards immigrations, usually affiliated with xenophobia (Inglehart & Norris, 2019). Regarding the post-materialism values μˆlib= 2.24 and μˆpop = 1.75 populists, on the post materialism scale liberals are more likely to have post materialist values while the opposite applies for populists. Literature shows that in fact this, that post-materialism is a modest predictor for the liberal vote (Middendorp, 1992).

Regarding group 4: Income level has a μˆlib= 5.40 and μˆpop = 4.96, this show that liberals are slightly wealthier than the populists and actually slightly more than the national mean. This is also consistent with literature where the populists are associated with a poorer income. This is also in countertrend to the Norway income level where the opposite was shown, going then against literature.

Finally, a graph multidimensional scaling graph of the proximities of the random forest.

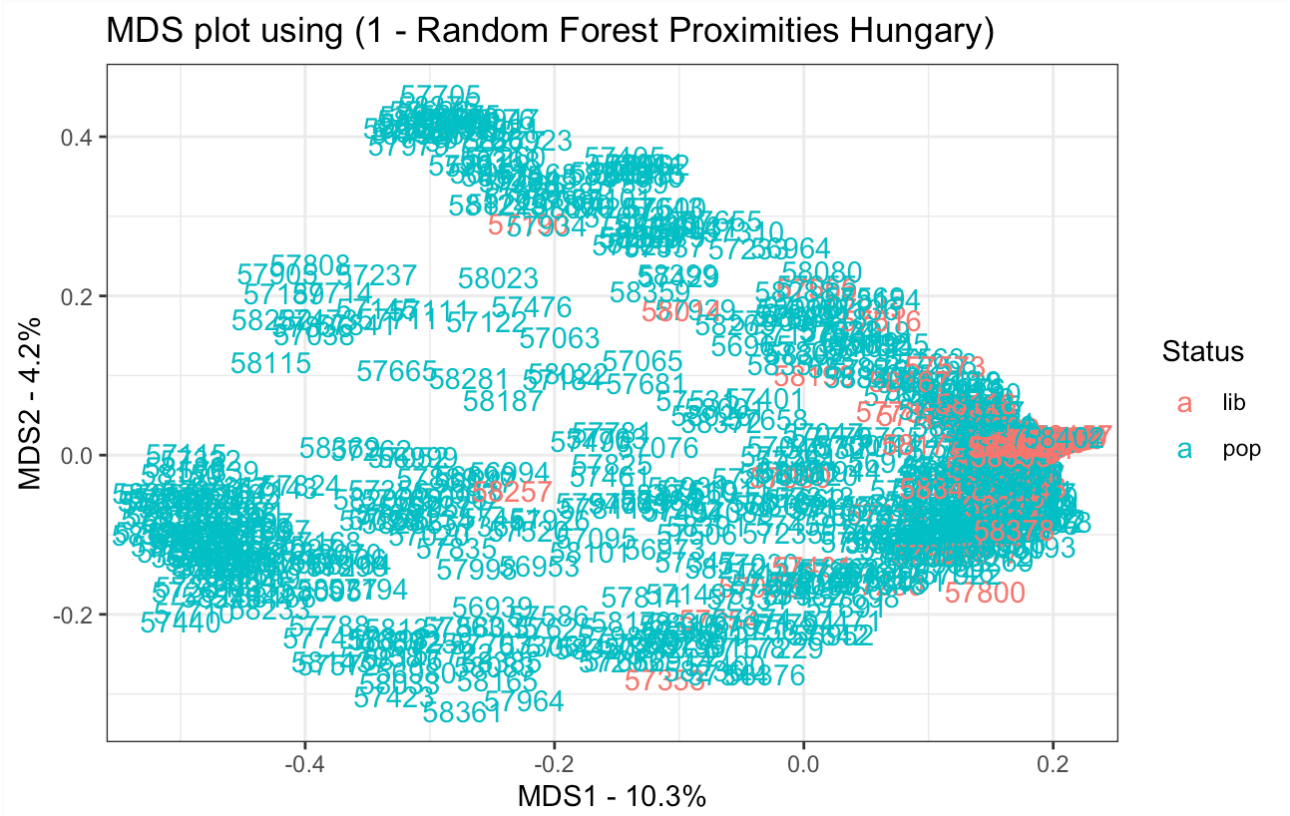

Figure 16

As we can see the graph look very different from the Norway one, the major voters clusters are in the middle right part where the vote is very mixed, there are probably concentrated the more moderate values, while at the opposite many populist voters are concentrated in the bottom left part of the graph or in middle top part. We saw how much the values in Hungary can correctly work as predictors for the populist voters, rather than the liberal voters, because of this I suspect that the bottom left, and middle-left respondents are the classical stereotype of the populist voters and thus are being clustered together.

Chapter 4: Germany prediction

At the time of data collection Germany was led by the coalition CDU/CSU, a center-right Christian-democratic political alliance headed by now former prime minister Angela Merkel. The German electoral system in summary, is a customized proportional electoral system (with correction

mechanisms) this should technically means more party but the WSV only accounts for 6 parties, none of these 6 has reached under 2% so they will all be in the calculations.

Before performing training, we can see that the liberals in Germany will be 1434, the populists will be 255 and the NA (see Hungary for reasons) are 2009. We can see how the liberals are in the vast majority in this case.

Performing the training, as usual, I used all the 118 variables, we found that the German training results in a respectable OOB error rate of 13.14%. The confusion matrix is as follows:

	lib	pop	Class error
lib	1424	10	0.006973501
pop	212	43	0.831372549

As we can we ends up with a similar situation to the Norway analysis, the liberal voters are classified nearly perfectly with a 1% error rate, while the populist voters are more commonly misclassified with an 83% error rate. As with the first analysis this random forest tends to classify most voters as liberals.

The first 9 variables in the training according to the mean decrease accuracy index are:

1. Group 1: Political Affinity. This group is composed by:

    - **Pol_rad**: Political radicalization. It measures the distance between voter's political position and the centre of the political space. It is measured as (pi − 5)2, where pi is the variable pol sp.
    - **C_gvr**: Confidence in the government (1 = A great deal, 4 = None at all)
    - **Dmc_cnt**: How democratically is this country being governed today? (1=not democratically, 10 = fully democratically).
    - **Pol_sp**: Self-positioning in the political space (1=left, 10 = right)
    - **C_prl**: how much Confidence you have in the Parliament (1= A great deal, 4= None at all)
    - **Ps_sat**: how satisfied are you with how the political system is func- tioning in your country today? (1 = not satisfied, 10 = completely satisfied).

2. Group 2: Anti-Cosmopolitanism:

    - **C_eu1**: How much confidence you have in the European Union? (1 = A great deal, 4 = None at all)

3. Group 3: Political vindications:

    - **Imp_imm**: Evaluate the impact of immigrants on the development of Germany (1 = very bad, 5 = very good).
    - **Job_ntn**: Employers should give priority to German people over immigrants? (1= Strongly agree, 5=Strongly disagree)

to these variables I am going to add a tenth variable which is income level:

4. Group 4: Income level:

    - **Incm_lvl**: What group your household is according to incomes? (1= 1st decile, 10= 10th decile)

Below I will post the graphs of the training.

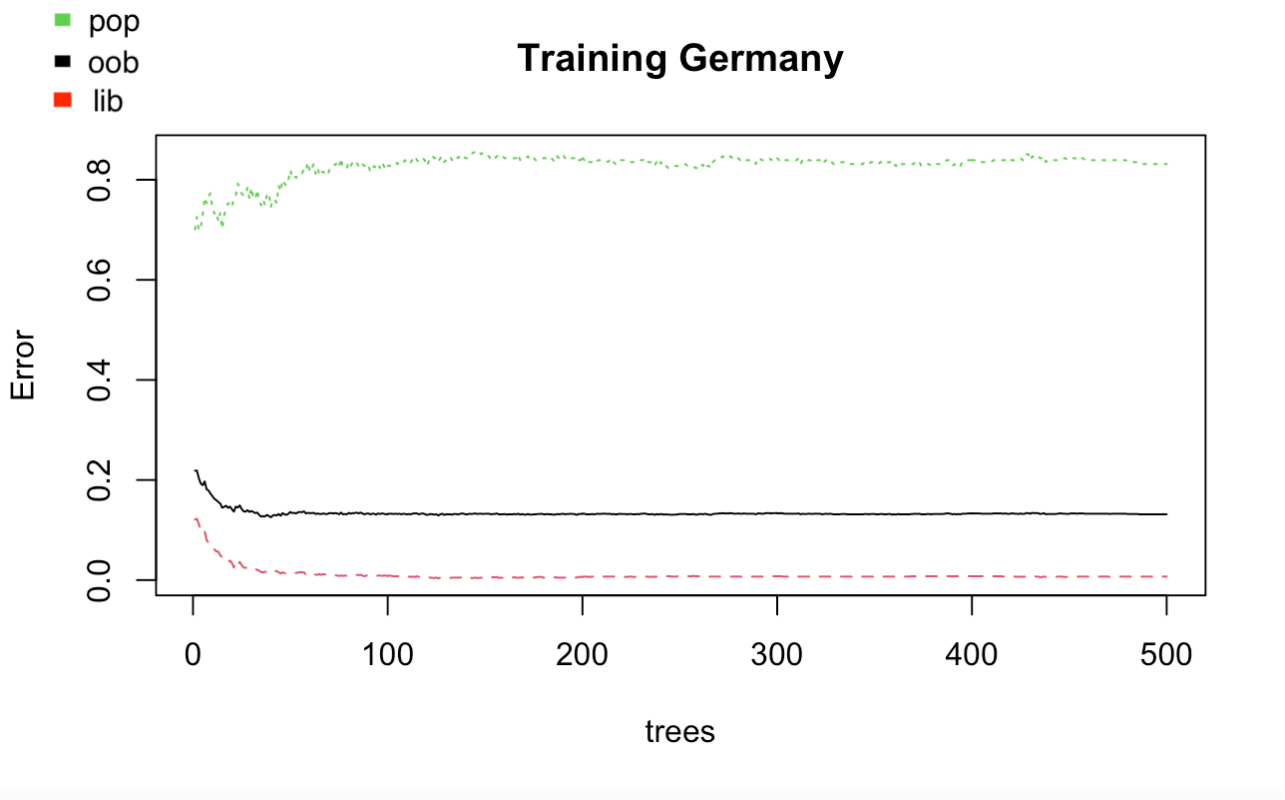

Figure 17

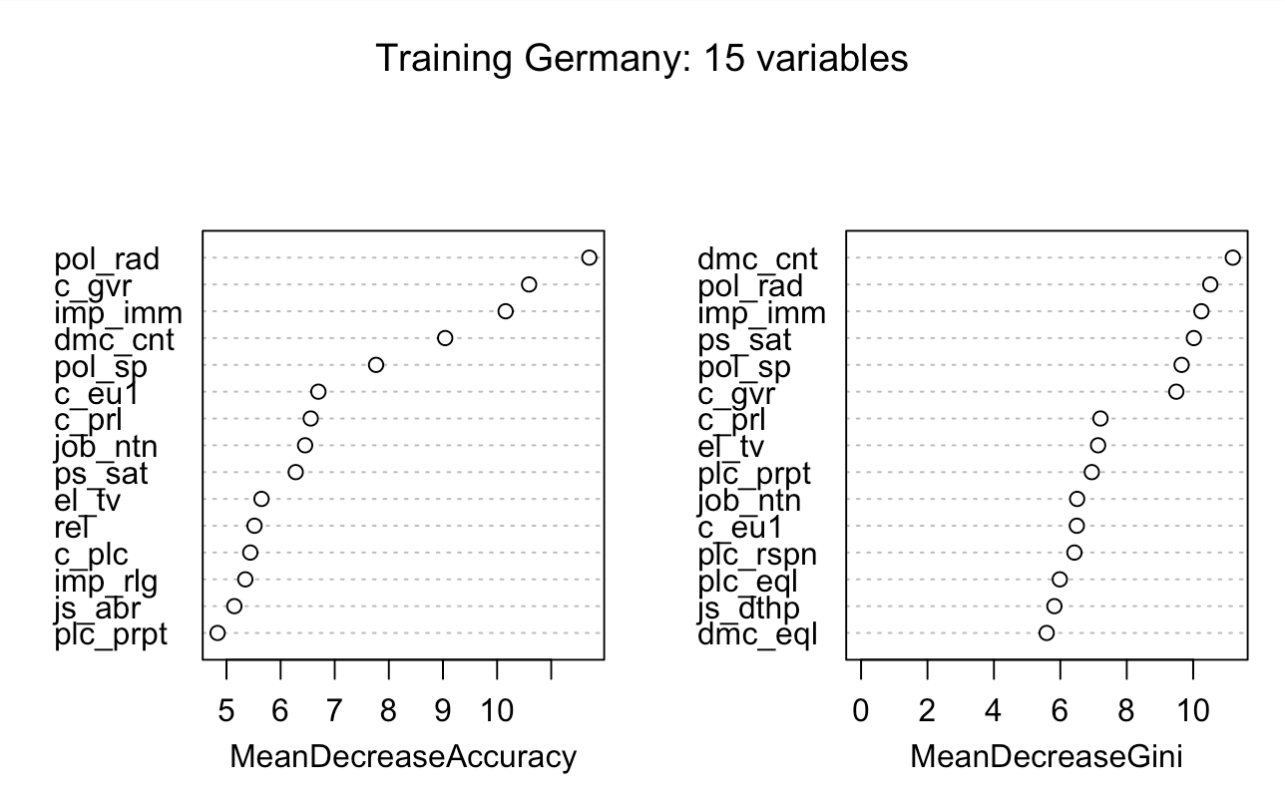

Figure 18

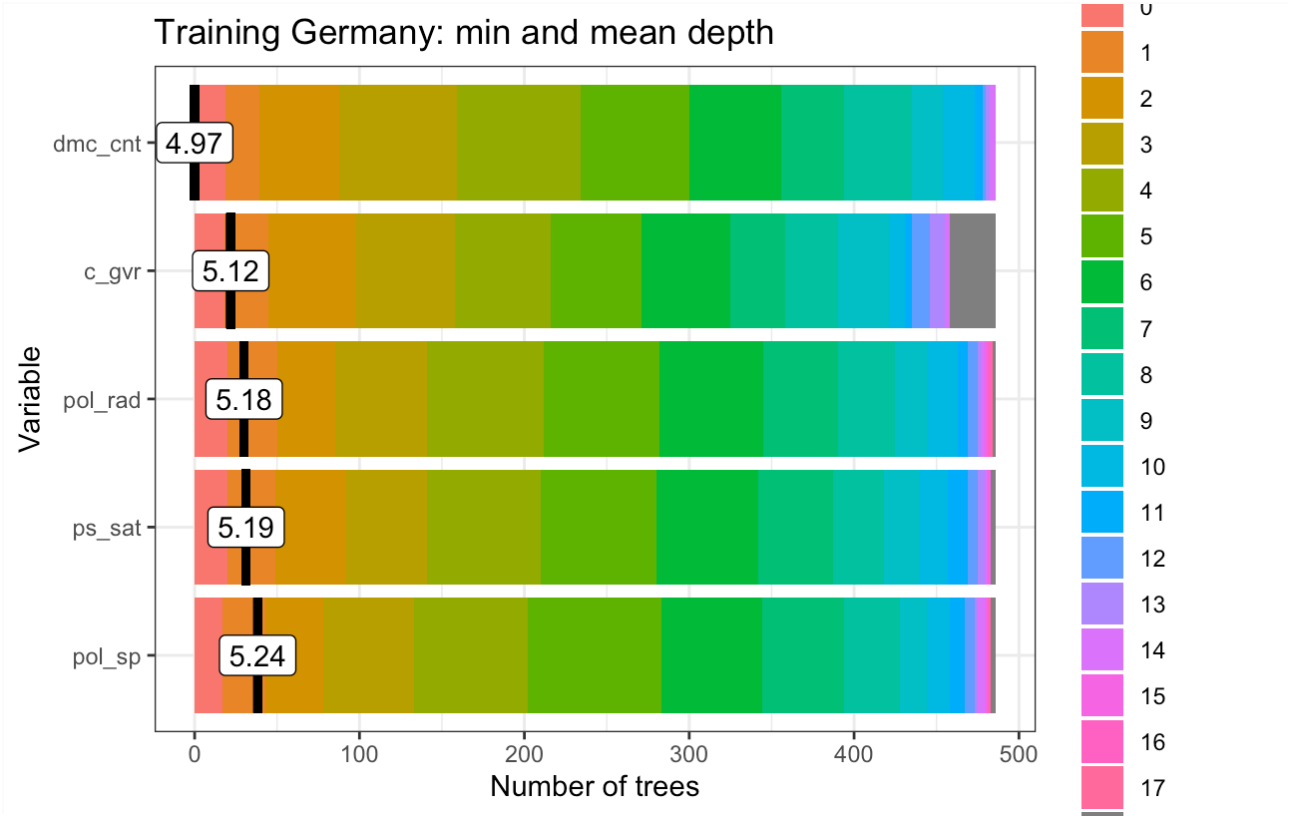

Figure 19

Doing the actual random forest, I will use once again, only 10 variables, as opposed of the 118 of the training, the random forest will be called “gerf_classifier_1”. We can see basically the same OOB error rate going from 13.14% to 13.2%. While the confusion matrix is as follows:

	lib	pop	Class error
lib	1388	46	0.0320781
pop	177	78	0.6941176

We can see how the lib class error is still close to the training one, going from 1% to a 3%, while pop class error saw a massive improvement going from 83% class error to a 69%. Overall, between the 3 random forests classifier, this is has more similarity with the Norwegian case rather than the Hungarian one.The one that does a more accurate job in predicting liberals and populist among all

three random forests should be the Hungarian one. I will post below the graphs of this random forest.

Figure 20

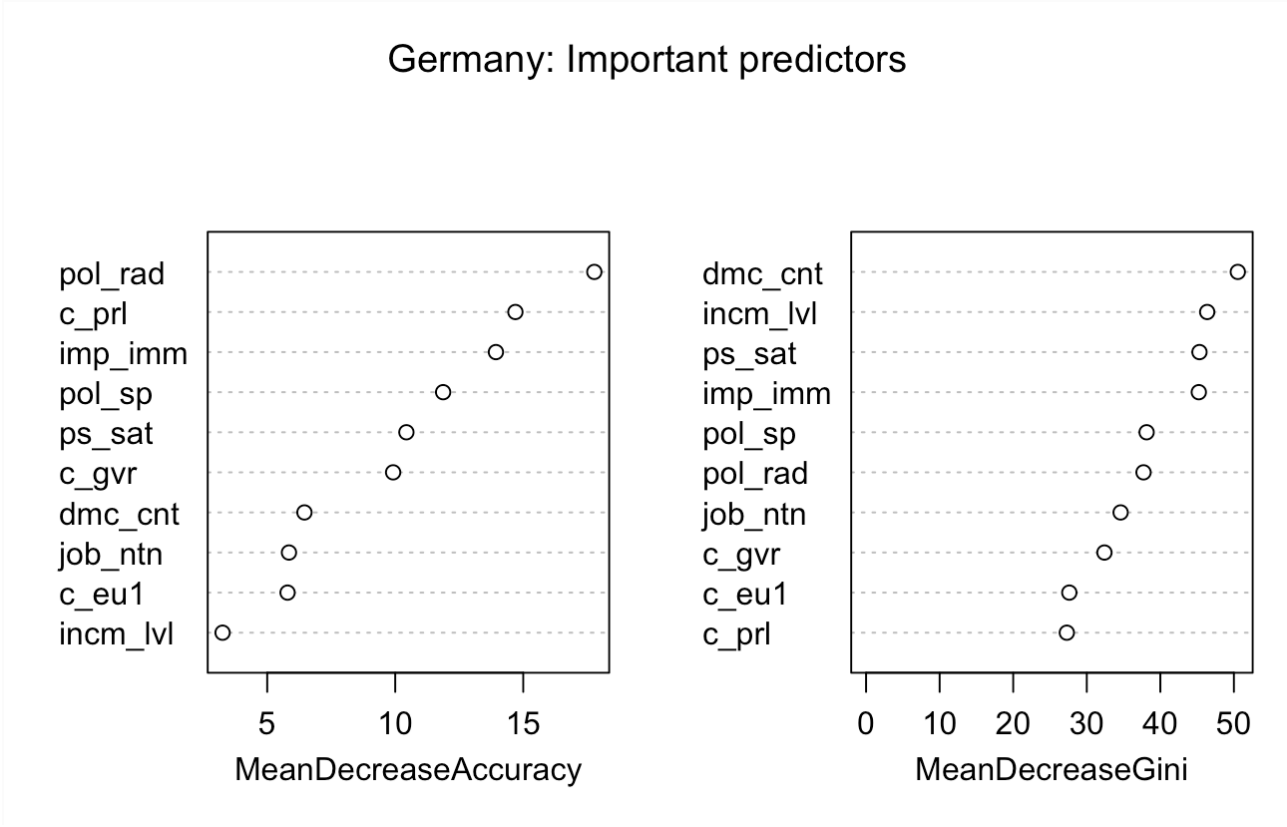

Figure 21

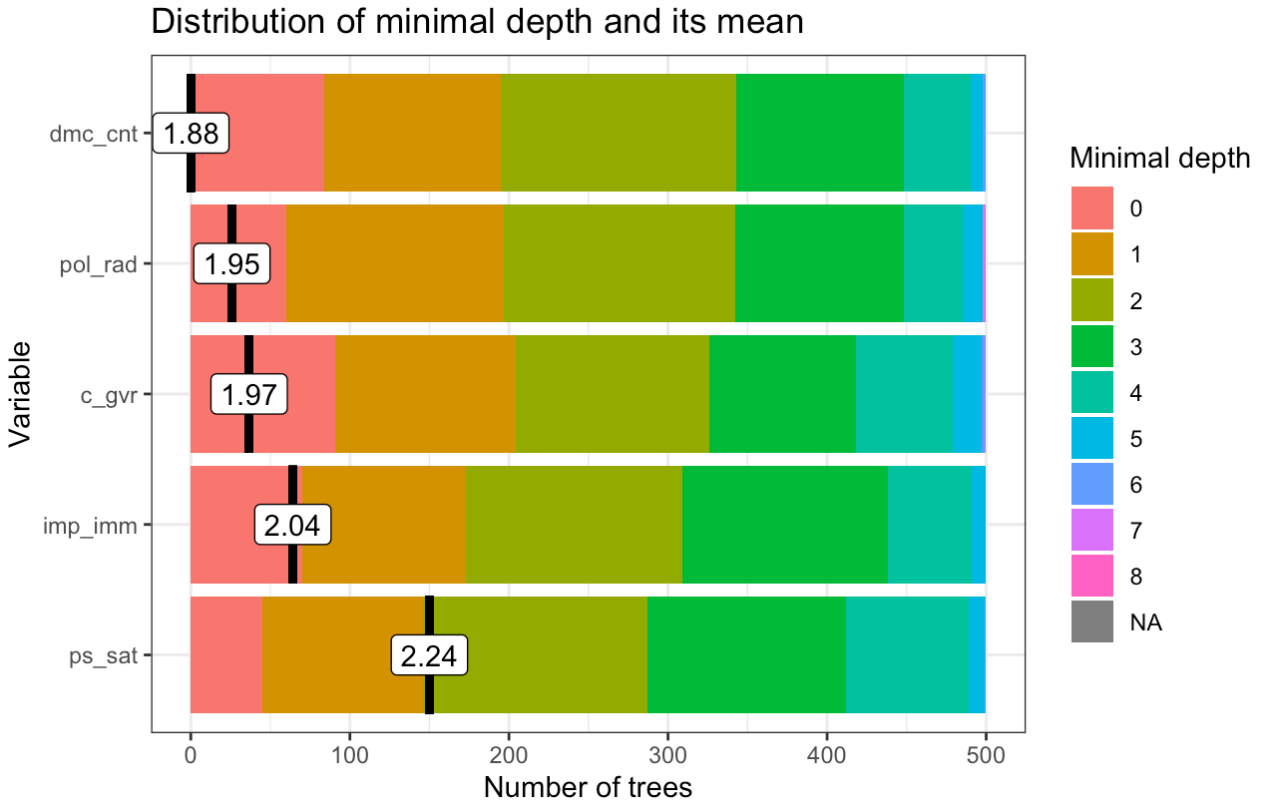

Figure 22

As we can see the first variable according to mean decrease accuracy is political radicalization, which is also the second variable according to the minimal (and mean) depth distribution. The first variable in mean decrease Gini is in fact dmc_cnt which is also the first in depth distribution. This should give us the idea of what are the most important variables in this specific random forest. Interestingly income level has a similar collocation to the Norwegian random forest as it ranks really high in the Gini decrease but at the same time it ranks as last in mean decrease accuracy.

Regarding group 1: Political radicalization has a μˆlib= 2.71 and the μˆpop = 6.04, this is in fact a common feature among both classes, populists are likely to have radical tendencies, whereas the opposite applies for liberals (Marne et al, 2021). Confidence in the government has a μˆlib= 2.60 and the μˆpop = 3.27 showing again how populist are much more dissatisfied than liberals, as liberals are currently running the government not populists. How democratically is your country has a μˆlib= 7.56 and the μˆpop = 5.73, showing once again as in the Norway case that since a liberal govern is currently running the State, only populists are not satisfied with the degree of democracy in the country. Political space interestingly shows a μˆlib= 4.88 and the μˆpop = 4.68 where both classes are fairly close to each other. This because interestingly there is a populist party in the far left space, whereas in Hungary there was not. Confidence in the parliament has μˆlib= 2.52 and the μˆpop = 3.19, reiterating the same values as before. Satisfaction with the current political system has a μˆlib= 6.09 and the μˆpop = 4.08 showing the same as before, with only populists dissatisfied. Political space interestingly shows a μˆlib= 4.88 and the μˆpop = 4.68 where both classes are fairly close to each other. Seemingly a common theme among populists in this group, is the dissatisfaction and or mistrust in the political sphere in a country where populists are not governing.

Regarding group 2: Confidence in the European Union has a μˆlib= 2.54 and the μˆpop = 3.07, showing that as found in research that populists have anti-cosmopolitism tendencies as they are more in favor of a national ruling and despise supranational entities such as the European Union and ONU or NATO. The opposite applies for liberals, as this class advocates for cosmopolitism and open borders (Arditi, 2007).

Regarding group 3: The evaluation of the impact of immigrants on the development of Germany has a μˆlib= 3.07 and the μˆpop = 2.57, as expected populists evaluate the effect of immigrants worst than the liberals. This because populists advocate for national identity with emphasis on ethnicity whereas liberals, in contrast, tends more towards an evolution of national identity and open borders. Prioritizing national job over immigrants has a μˆlib= 3.34 and the μˆpop = 2.86, the same applies here as populists have more tendencies to prioritize national jobs rather than immigrants.

Regarding group 4: The income level here has a μˆlib= 5.78 and the μˆpop = 5.18, this is consistent with research as it shows a lower income level of populists, whereas liberals are slightly wealthier.

Finally, a graph multidimensional scaling graph of the proximities of the random forest.

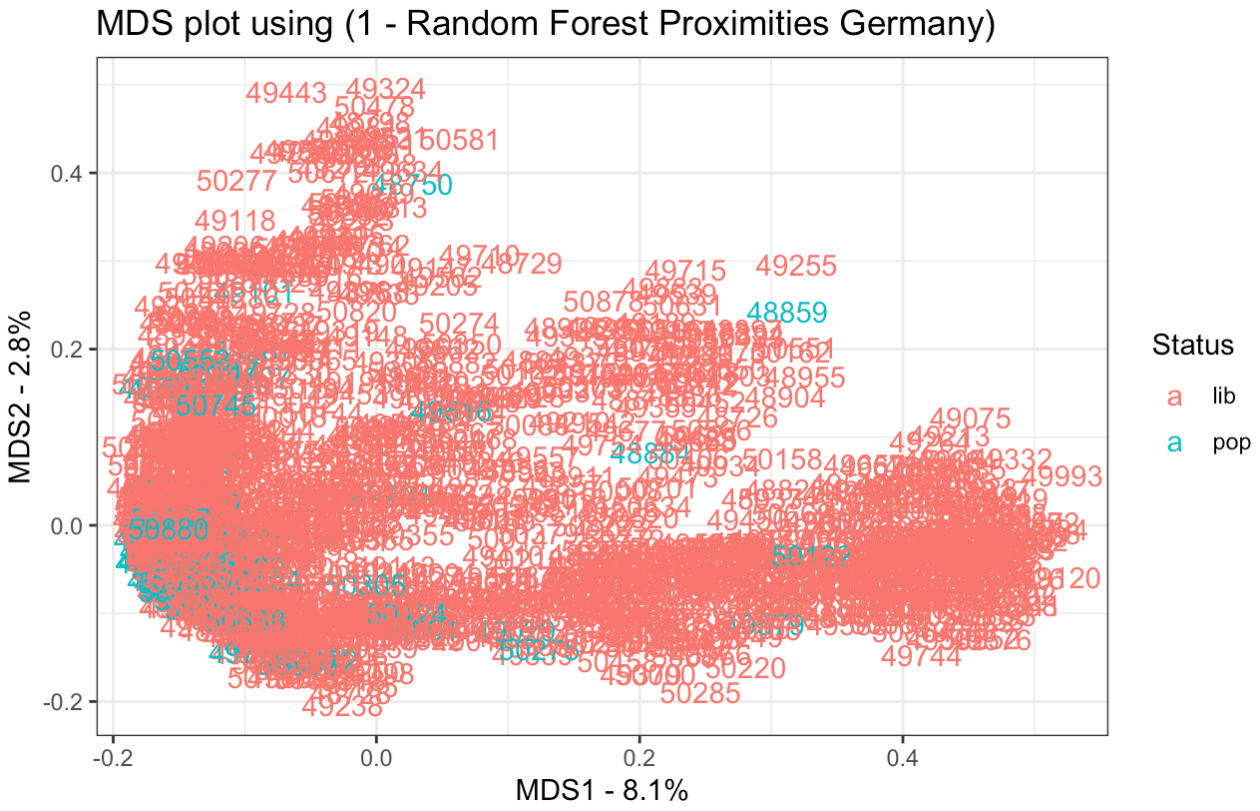

Figure 23

Same as with the Norway multidimensional scaling graph, not much can be appreciated with this graph. We can see how much of the entries are concentrated in the bottom / center left and in the bottom right. As well as this we can appreciate a few outlier values. This graph is a demonstration of the vast heterogeneity of the country that is being analyzed, whereas in country like Hungary we can see how the strong opinions forms more clearly labeled groups.

Chapter 5: Discussion and conclusion

After having analyzed all the three countries and the three Random Forests we can appreciate the differences. Starting from the analytical point of view, the finest random forest in term of classifying, thus meaning the one with the lowest error rates, is undoubtedly the Hungarian one (hurf_classifier_1), as it has respectable error rates in both class error and OOB error. Regarding the Norwegian and German analysis, the matter is similar; given the heterogeneity of both countries and the mainly moderate point of view of both classes in the countries, it is rather difficult to predict what an individual preference of vote is going to be. A clear visual indicator of this phenomena can be seen by the MDS plots, as in both Germany and Norway there is no clear indication of where the main populist and liberal clusters are, while for the Hungarian case we can appreciate a few populist and liberal clusters. The main discovery of this predictive analysis lays in the fact that populists, that have in their values the rejection of parliamentarism, the rejection of elitism and a call for the direct democracy are only dissatisfied with these forms as long as the government running the country is of an opposing faction, this can be seen in the Hungarian Group 1 of variables as this was the common theme among every single variable, while for both other countries the response of the populist voters among the political affinities were the opposite, showing a lack of confidence towards a government of an opposing faction. Apart from income level, which was artificially added, we saw 4 variables that were recurrent in all 3 random forests; these were: confidence in the European Union, the impact of immigrants on development of the country, the confidence in the government and the grade of democracy in the country. They belong to 3 different group: Political Affiliation, Anti-Cosmopolitanism and Political Vindication. We can see how comparing the confidence in the European Union for all 3 countries gives us similar values as liberals are generally closer to the value of 2 (more confident), while populists are closer to the value of 3 (less confident), showing a clear tendency towards anti-cosmopolitanism from populists while at the same time liberals embrace this

supranational entity. Other 2 variables belong to political affiliation and the values are more different from each other, indeed confidence in the government shows a phenomena where Norway and Germany have similar values, liberals more confident than populists, due to the current government being of a liberal faction, while in Hungary where there is a populist govern we can appreciate the exact opposite phenomena, with populists more confident than liberals towards the government. The same happens when looking at the other variable which is grade of democracy in the country. This raises questions on the populist ideology, the call for a form of direct democracy and anti-elitism are still key characteristics of the ideology or has it evolved in something else? As for their voters we can see how, they are not concerned with this argument, since their faction is currently guiding the country and they seem to be satisfied with it. The other interesting discovery in the field would rely on the on the economical aspect of populism, indeed according to researchers, populism is affiliated with economic inequality and in fact to a loss of prosperity (O'connor, 2017). Thus, we should expect populist to lay in a lower bracket of income level compared to liberals, this is in fact the case for Germany and Hungary, while the opposite applies for Norway where interestingly, although not by much, populists are wealthier than liberals. Once again the question, is populism a watertight ideology or does it evolve with time? Another interesting fact about this analysis is the lack of religious values in the study, as I took only the first 9 variables (plus the income level variable) I expected to see some religious values as researchers have linked the rise of populism with religious values (Zúquete, 2017). I especially expected to see some in the latter two analysis as the Hungarian population has strong religious values and in Germany as for 20 years, they had a prime minister which was the head of a Christian party. We only saw 2 religious values in Norway, being child obedience and the justifiability of the divorce, a country where religion is not particularly rooted in the society. Another interesting finding of this analysis is the constant presence of some sort of political vindications, in fact as previously mentioned, the impact of immigrants on the development of the country was a constant variable in all 3 random forests, particularly the concerning towards immigrants is one of the main points where liberals and populists differentiate themselves. We can then assess how populism and liberalism are different from country to country, mainly from the political affiliation point of view, both political vindication and anti-cosmopolitism were at least comparable. All three countries are in my opinion facing a crisis of identity and the decay of values typical of western countries in the 21st century. This is represented by the growth of populism and thus the populist point of view towards political vindication, political affinity and anti-cosmopolitanism. Although there are similarities, the values and attitudes towards these sentiments, as we said, are different from country to country and comparing each other may results in fallacious conclusions. Each country is facing different crisis and different situations, this is the main answer to why these values are different from each other.

Overall, in this analysis we can draw 2 main conclusions; the first being how random forest can be a useful tool in social research, as we saw in the Hungarian chapter it is possible to predict to a certain degree certain types of voters, although being in specific condition. The other conclusion we can draw is how in my opinion liberalism and populism are two ideologies that are not fixated in time, but rather both keep evolving to respond to new issues given by the political and social trends of times.

Bibliography:

Arditi, B. (2007). *Politics on the edges of liberalism: Difference, populism, revolution, agitation*. Edinburgh University Press.

Charvet, J. (2018). *Liberalism: the basics*. Routledge.

Charvet, J. (2019). *Liberalism: the basics*. Routledge.

Cirone, A., Cox, G. W., & Fiva, J. H. (2021). Seniority-based nominations and political careers. *American Political Science Review*, *115*(1), 234-251.

Collins English dictionary (2020) Available from: http://www.collinsdictionary.com/english/creative

G. David Garson - Data Analytics for the Social Sciences_ Applications in R-Routledge (2021)

Genuer, R., & Poggi, J. M. (2020). *Random Forests with R*. Springer Cham.

Harmsen, R. (2010). Concluding comment: On understanding the relationship between populism and Euroscepticism. *Perspectives on European Politics and Society*, *11*(3), 333-341.

Hartshorn, S. (2016). Machine learning with random forests and decision trees: A Visual guide for beginners.

Kovács, Z., & Vida, G. (2015). Geography of the new electoral system and changing voting patterns in Hungary. *Acta Geobalcanica*, *1*(2), 55-64.

Louppe, G., Wehenkel, L., Sutera, A., & Geurts, P. (2013). Understanding variable importances in forests of randomized trees. *Advances in neural information processing systems*, *26*.

Marcos-Marne, H., Llamazares, I., & Shikano, S. (2021). Left-Right radicalism and Populist attitudes in France and Spain. *Journal of Contemporary European Studies*, 1-15.

Maxwell, R. (2019). Cosmopolitan immigration attitudes in large European cities: Contextual or compositional effects?. *American Political Science Review*, *113*(2), 456-474.

Middendorp, C. P. (1992). Left-Right Self-identification and (Post) materialism in the Ideological Space; their Effect on the Vote in the Netherlands. *Electoral Studies*, *11*(3), 249-260.

Mudde, C., & Kaltwasser, C. R. (2017). *Populism: A very short introduction*. Oxford University Press.

Norris, P., & Inglehart, R. (2019). *Cultural backlash: Trump, Brexit, and authoritarian populism*. Cambridge University Press.

Social Progress Imperative: 2021 Social Progress Index. Social Progress Imperative. Washington, DC. Available at: www.socialprogress.org

Three Connections between Rising Economic Inequality and the Rise of Populism Author(s): Nat O'Connor Source: Irish Studies in International Affairs , Vol. 28 (2017), pp. 29-43

Thomson Reuters. (2017, June 30). *Germany's Merkel voted against Gay Marriage*. Reuters. Retrieved May 3, 2022, from https://www.reuters.com/article/us-germany-gay-marriage-merkel-idUSKBN19L0SU

Zúquete, J. P. (2017). Populism and religion (pp. 445-466). Oxford: Oxford University Press.