How can one predict if a person has boycotted a product or service in the last 12 months?

Samuel Dornan

Introduction

The European Social Survey is a pan-European survey designed to measure the attitudes and beliefs of European nationals surrounding a wide variety of factors ranging from political beliefs to social issues. One of these data points measured is the question of whether or not the respondent has boycotted any products within the last 12 months. In this paper, I will attempt to determine if it is possible to predict whether or not a person has boycotted a product in the last twelve months using a number of factors such as whether they have posted political content, whether they feel it is important to conform with social norms, the amount of time they spend online, and so forth.

Motivation and aims

Though the ESS Round 10 data was collected between 2020 and 2022 (European Social Survey, 2021) and so wouldn't necessarily provide insights which could be directly related to ongoing social phenomena such as the Boycott, Divest & Sanction movement it may still provide some insights into factors which may be related to a person's decision to partake in a boycott for any reason and in aid of any particular cause.

Preliminary research shows that a person's decision to engage or not engage in a boycott may be influenced by a combination of internal and external factors, ranging from one's ability to gain social status or prestige by publicly engaging in a boycott and their ability to replace or otherwise cope without the good in question, to external factors such as whether the people they know are engaged in a boycott.

Boycotts tend to have a negative economic impact on a target company in the short-term, and a more negligible effect in the medium to longer term (Heilmann, 2016), often due to the inherently short-term nature of the act in question and one of the participants' primary aims being to influence the actions of the company in question (Braunsberger and Buckler, 2011).

According to existing literature however, attempts to influence the company's management and profit margins were far from the only motivating factors behind a person's decision to participate in the boycott; some would engage for more selfish reasons, such as in an effort to gain social status or prestige (Klein et al., 2004), while others would engage because their peers were doing the same (Delistavrou et al., 2020).

It is interesting to note that there would appear to be a group of individuals who would often not participate in boycotts, which comprises at least two different subsets of people; the first subset does not participate in the boycott as they would not be as easily able to adapt to the loss of the item in question (Klein et al., 2004), while the second group of individuals would not participate and would also spend a larger amount of time online during the week, and would post a greater amount of political content online than their counterparts who would participate in boycotts.

Description of the data and methods used

The specific dataset studied was the subset of survey responses which originated in Ireland. The data in question was first analysed with a simple summary of the responses in order to assure that all values were as described in the codebook; that is, binary values were binary, categorical values were within the described rating scales, and so forth. Dependent variables which did not match their expected values as per the codebook were removed before any analysis took place.

Specifically, the predictor variables were as follows;

1. Whether a respondent has posted or shared anything about politics online last 12 months (pstplonl),
2. How often a respondent uses the Internet (netusoft),
3. Whether respondents thought that most of the time people were helpful or mostly looking out for themselves (pplhlp),
4. Whether respondents thought that most people try to take advantage of you, or try to be fair (pplfair),
5. Whether respondents thought that it was important to help people and care for others well-being (iphlppl),
6. Whether respondents thought it was important to show their abilities and be admired (ipshabt),
7. Whether respondents thought it was important to be successful and that people recognise achievements (ipsuces),
8. Whether respondents thought it was important to be loyal to friends and devoted to people close (iplylfr), and
9. Whether respondents thought it was important to behave properly (ipbhprp)

While the response variable was a binary variable asking whether or not a person had boycotted any products in the last 12 months (bctprd). It is worth noting that the response variable was recoded, with any responses denoting “No” being recoded from 2s to 0s. This was required in order to run the regression algorithm.

Broadly speaking, predictor variables 1 and 2 were selected as proxies to determine a respondent's online activities, predictor variables 3, 4, 5, 6, and 7 were selected as proxies to determine whether a respondent believed others would engage in self-enhancing behaviours or not, and given this whether they would be encouraged to do the same thanks to social learning (Over and Carpenter, 2012) and predictor variables 8 and 9 were selected as proxies to determine whether a respondent would be influenced by collective actions.

It is worth noting that for every respondent who did admit to engaging or participating in a boycott within the last twelve months, there were approximately six more who did not (Figure 1), creating a heavily skewed distribution of the results within the dataset. A significant number of binary classification models require a binomial distribution for their data as a skewed distribution often leads to predictions being heavily biased towards the most commonly occurring results, potentially leading to false negatives or false positives.

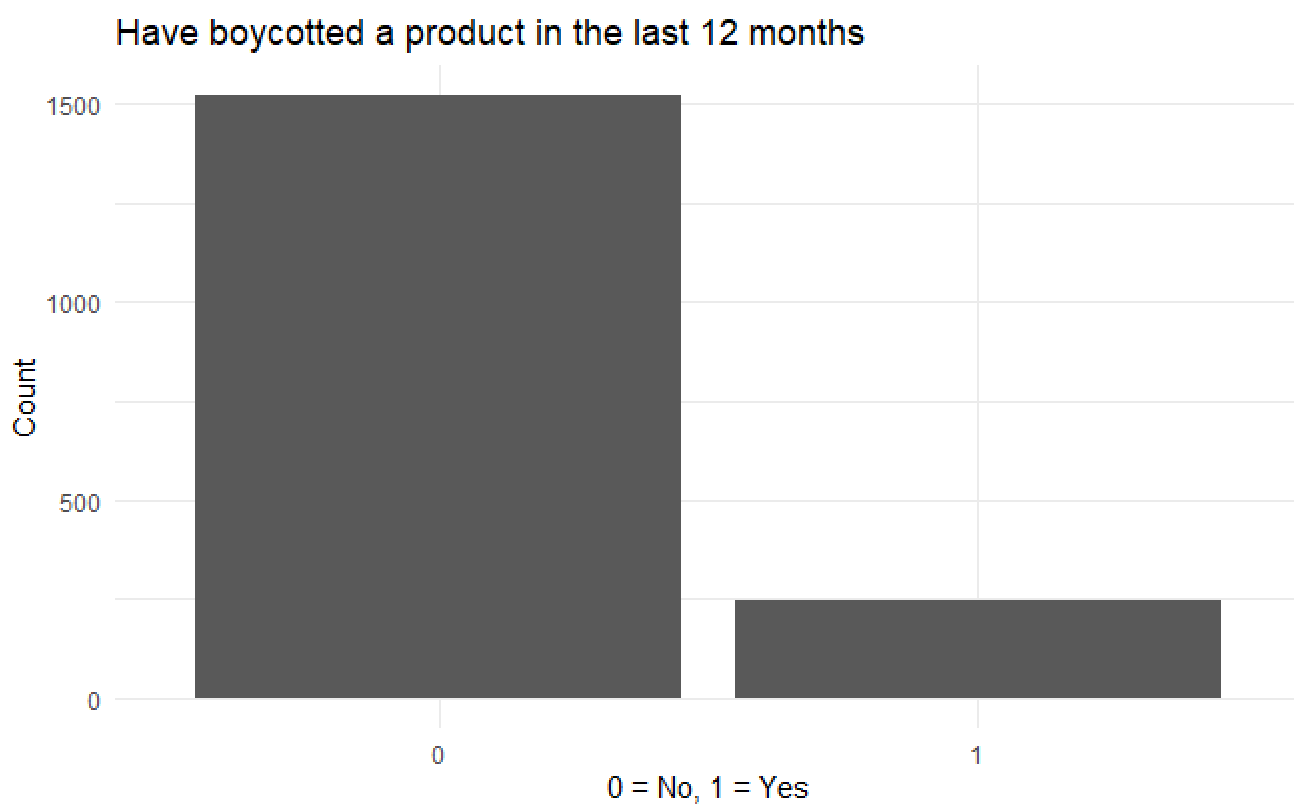


Figure 1: Distribution of participants who have and have not boycotted a product within the last 12 months.

In order to account for this a pair of logarithmic regression analyses were carried out; the first with original data, and the second with data resampled to a more equal number of respondents who did and did not engage in boycotts respectively. The first logarithmic regression carried out with original data was used to determine the similarity between the data provided and the expected results derived from the preliminary research, while the second logarithmic regression was carried out in order to assure that resampling the dependent variable into a binomial distribution would not materially influence the values of the selected variables; that is, strong negative correlations found in the first regression would remain strong negative correlations in the second regression, variables that were not correlated with one's history of engaging in boycotts in the first regression would not suddenly become correlated in the second regression, and so on.

Following this, the data was split with a stratified sample into training and test datasets, with 70% of the values in the training set and 30% in the test set. This is a common split allowing the models to be thoroughly tested and validated with unseen data. Stratified sampling was used as it ensured a constant ratio of yes and no responses for the response variable within both datasets, ensuring that the binomial distribution is maintained and making it easier to carry out classification going forwards.

The data was classified using a Random Forest algorithm due to a combination of factors such as convenience and ease of use, reliability, and its ability to be fine-tuned using a subset of variables rather than all. It is worth noting that Random Forest classifiers do not always generalise well to seen and unseen data.

Analysis and interpretation

A logistic regression analysis found a strong negative correlation between posting political content online and engaging in boycotts, while the amount of time they spent online was found to be weakly but positively correlated with a respondent engaging in a boycott (Table 1). The selected variables designed to determine whether respondents would engage in self-enhancing activities were found to be very weakly correlated with respondents participating in boycotts, while the selected variables designed to show whether or not a respondent would be influenced by collective actions showed a weak negative correlation between whether it was important to help others and engagement in boycotts, and a weak positive correlation between behaving properly and engaging in boycotts.

term	estimate	std.error	statistic	p.value
(Intercept)	0.508617	0.585187	0.869153	0.384763
pstplonl	-2.00897	0.174223	-11.531	0
netusoft	0.215712	0.070868	3.043851	0.002336
pplhlp	0.028735	0.017028	1.687514	0.091505
pplfair	-0.04709	0.027691	-1.70038	0.08906
iphlppl	-0.15042	0.089356	-1.68343	0.092293
ipshabt	0.046305	0.06141	0.754027	0.450833
ipsuces	0.080916	0.059173	1.367439	0.171488
iplylfr	-0.13191	0.083016	-1.58902	0.112057
ipbhprp	0.212577	0.05383	3.949063	7.85E-05

Table 1.

Running a similar logistic regression on reweighted data produced similar results (Table 2).

term	estimate	std.error	statistic	p.value
(Intercept)	0.508617	0.585187	0.869153	0.384763
pstplonl	-2.00897	0.174223	-11.531	0
netusoft	0.215712	0.070868	3.043851	0.002336
pplhlp	0.028735	0.017028	1.687514	0.091505
pplfair	-0.04709	0.027691	-1.70038	0.08906
iphlppl	-0.15042	0.089356	-1.68343	0.092293
ipshabt	0.046305	0.06141	0.754027	0.450833
ipsuces	0.080916	0.059173	1.367439	0.171488
iplylfr	-0.13191	0.083016	-1.58902	0.112057
ipbhprp	0.212577	0.05383	3.949063	7.85E-05

Table 2.

A random forest model (Figure 3) trained on the resampled training data was found to have a mean of squared residuals of 0.0932 and explained approximately 62.71% of the variance within the training dataset. That is, it appeared to be accurate approximately 63% of the time when testing against the training data.

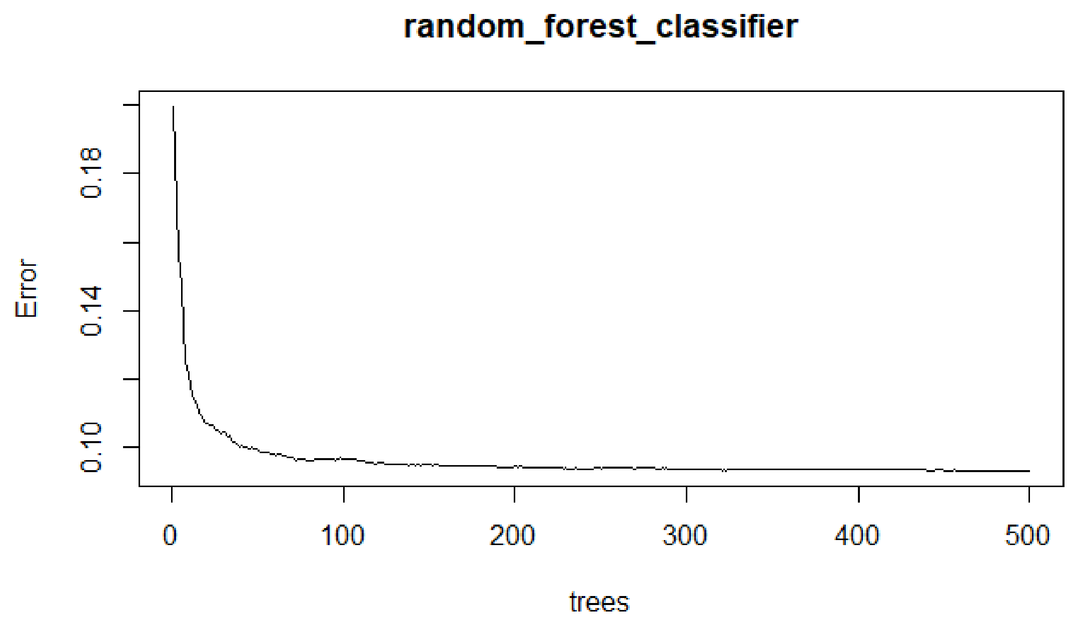


Figure 2: Random Forest Classifier trained on 3 variables (default)

Fine-tuning the classifier showed that the lowest out-of-bag error would be found by increasing the number of variables tried at each split from 3 to 6, and this increased the percentage of the variance explained by the model from 62.71% to 65.55%, while the mean of squared residuals decreased to 0.0086. (Figure 3)

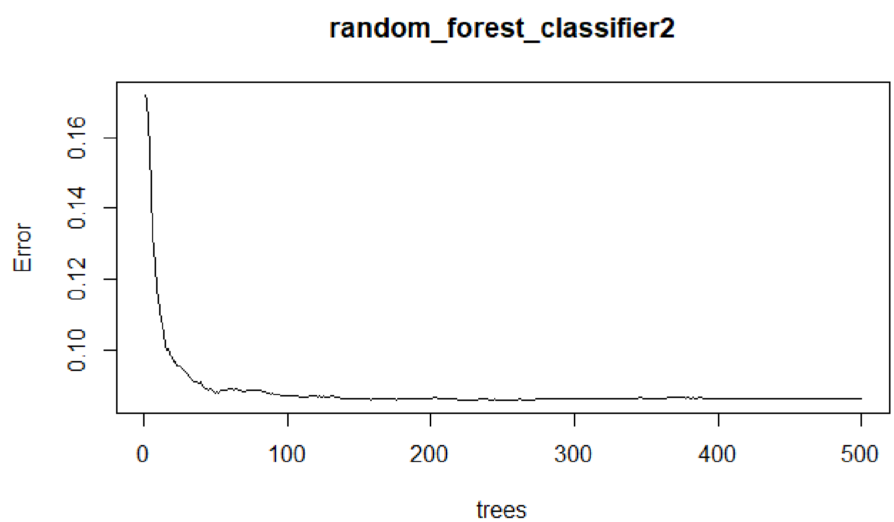


Figure 3: Random Forest trained on 6 variables

Testing the model on unseen data showed it to have lacklustre predictive performance however, with a reported accuracy of between 0 and 0.00691 for both regression models (Table 3, Table 4). Indeed, the accuracy obtained by predicting the most frequently occurring value was found to be 0.513, roughly equal to tossing a coin, which makes a certain degree of sense given that the sample dataset was reweighted to have an even number of people who did and did not participate in a boycott within the last 12 months, but suggests that the predictive power of the model as trained is negligible at best. Given that the original dataset had approximately six times as many people who did not participate in a boycott within the last twelve months as did participate however, this is not entirely surprising and may in practice be a somewhat effective means of predicting whether a person has engaged in a boycott.

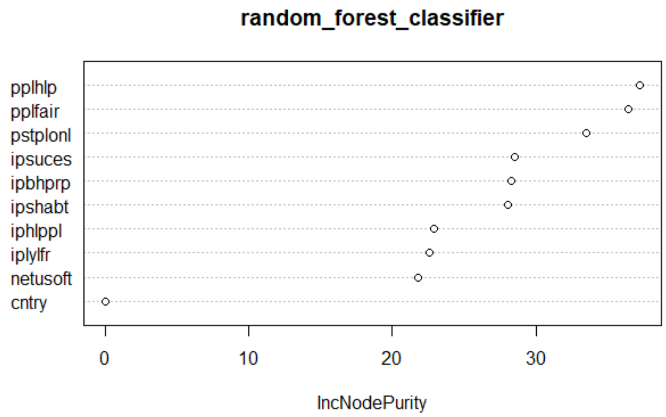
Accuracy	0
Kappa	0
AccuracyLower	0
AccuracyUpper	0.00691
AccuracyNull	0.513158
AccuracyPValue	1
McnemarPValue	NaN

Table 3.

Accuracy	0
Kappa	0
AccuracyLower	0
AccuracyUpper	0.00691
AccuracyNull	0.513158
AccuracyPValue	1
McnemarPValue	NaN

Table 4.

It is interesting to note that the most important variables to the random forest classifier, whether respondents thought that most of the time people were helpful or mostly looking out for themselves and whether or not they thought people were fair, were also two of the variables most weakly correlated with whether people had engaged in a boycott in the last year. Whether people had posted political content online was the third most important variable, and how often a respondent uses the Internet a respondent uses the internet was the second-least important variable, after their country of origin. The country of origin was restricted in such a way that all survey responses originated from Ireland, so the fact that this is the least important variable in the random forest is not surprising.



In spite of the poor predictive power of these results, it is somewhat comforting to note that these results match those found in previous studies; that is, given the correlations between the variables determined in the logistic regression analysis it can be inferred from the data that, among other things, people who post political content online are significantly less likely to engage in boycotts than those who spend less time online. Interestingly however, there was no correlation between whether a person would engage in self-enhancing behaviours and whether they partook in a boycott within the previous 12 months of completing the survey, however it should be noted that previous research found that a person's propensity to engage in a boycott would be limited not only by their ability to engage in self-enhancing behaviours, but also by the severity of the loss resulting from the boycott, which was not an available variable within the dataset.

Conclusion

Few conclusions can be drawn from this analysis which cannot be drawn from other similar literature; the vast majority of respondents surveyed did not engage in boycotts within the last 12 months, and in general within the dataset those respondents who posted political content online were much less likely to engage in boycotts than those who did not post political content. Individuals who felt it was important to be well behaved were slightly more likely to engage in boycotts than those who did not. There was effectively no correlation between whether a respondent might engage in self-enhancing behaviours and whether or not they had engaged in a boycott.

Use of AI Tools

Tool name	Purpose of use	Prompt
GitHub Copilot	Assistance writing R code	@workspace /fix Error in file(con, "w") : cannot open the connection Calls: .main -> execute -> -> writeLines -> file In addition: Warning message: In

chat in VSCode		file(con, "w") : cannot open file 'POL30660_Spring2024_lab07_class08.qmd': No such file or directory Execution halted
GitHub Copilot chat in VSCode	Assistance writing R code	@workspace /fix Error: 'file' has no extension
GitHub Copilot chat in VSCode	Assistance writing R code	@workspace /fix Error in import(ess_data) : could not find function "import"
GitHub Copilot chat in VSCode	Assistance writing R code	How do I create an error handling case to skip an installation if a package already exists in R
GitHub Copilot chat in VSCode	Assistance writing R code	How do I remove values from a dataframe in R
GitHub Copilot chat in VSCode	Assistance writing R code	How do I account for the fact that bctprd has about 6 times as many 2s in it as it has 1s?
GitHub Copilot chat in VSCode	Assistance writing R code	How can I visualize the class imbalance in my dataset?
GitHub Copilot chat in VSCode	Assistance writing R code	How do I account for the fact that bctprd has about 6 times as many 2s in it as it has 1s?
GitHub Copilot chat in VSCode	Assistance writing R code	Error in eval(family$initialize) : y values must be 0 <= y <= 1
GitHub Copilot chat in VSCode	Assistance writing R code	@workspace /fix Error in selected_data.head() : could not find function "selected_data.head"
GitHub Copilot chat in VSCode	Assistance writing R code	Error: object 'pplfair' not found

GitHub Copilot chat in VSCode	Assistance writing R code	Warning message: package 'ggplot' is not available for this version of R A version of this package for your version of R might be available elsewhere, see the ideas at https://cran.r-project.org/doc/manuals/r-patched/R-admin.html#Installing-packages
GitHub Copilot chat in VSCode	Assistance writing R code	How can I calculate the accuracy of my classification model in R?
GitHub Copilot chat in VSCode	Assistance writing R code	Error: `data` and `reference` should be factors with the same levels.
GitHub Copilot chat in VSCode	Assistance writing R code	cm <- caret::confusionMatrix(resampled_test, resampled_test$bctprd) Error: data and reference should be factors with the same levels.
GitHub Copilot chat in VSCode	Assistance writing R code	predictions1 <- factor(predictions1, levels = levels) Error in match(levels, exclude) : 'match' requires vector arguments actual_values <- factor(actual_values, levels = levels) Error in match(levels, exclude) : 'match' requires vector arguments cm <- caret::confusionMatrix(resampled_test, resampled_test$bctprd) Error: data and reference should be factors with the same levels. cm function (x) 2.54 * x <bytecode: 0x000002c716573a78> <environment: namespace:grDevices>
GitHub Copilot chat in VSCode	Assistance writing R code	Mean of squared residuals: 0.1026103 % Var explained: 58.95
GitHub Copilot chat in VSCode	Assistance writing R code	Why am I getting an accuracy of 0
GitHub Copilot chat in VSCode	Assistance writing R code	How do I make sure the distributions between the training and test data are the same
GitHub Copilot	Assistance writing R code	How do I make the distributions between the training and test data the same

chat in VSCode		
GitHub Copilot chat in VSCode	Assistance writing R code	What is a ROC curve
GitHub Copilot chat in VSCode	Assistance writing R code	How do I plot this in R

References

BRAUNSBERGER, K. & BUCKLER, B. 2011. What motivates consumers to participate in boycotts: Lessons from the ongoing Canadian seafood boycott. *Journal of Business Research,* 64**,** 96-102.

DELISTAVROU, A., KRYSTALLIS, A. & TILIKIDOU, I. 2020. Consumers' decision to boycott "unethical" products: the role of materialism/post materialism. *International Journal of Retail & Distribution Management,* 48**,** 1121-1138.

EUROPEAN SOCIAL SURVEY. 2021. *Round 10 fieldwork update* [Online]. Available: https://www.europeansocialsurvey.org/news/article/round-10-fieldwork-update [Accessed 2nd April 2024].

HEILMANN, K. 2016. Does political conflict hurt trade? Evidence from consumer boycotts. *Journal of International Economics,* 99**,** 179-191.

KLEIN, J. G., SMITH, N. C. & JOHN, A. 2004. Why We Boycott: Consumer Motivations for Boycott Participation. *Journal of Marketing,* 68**,** 92-109.

OVER, H. & CARPENTER, M. 2012. Putting the social into social learning: explaining both selectivity and fidelity in children's copying behavior. *Journal of Comparative Psychology,* 126**,** 182.