

UNIVERSITY OF CHICAGO
MS ANALYTICS
DATA MINING FINAL PROJECT – WINTER-2017

THE TRUMP EFFECT AND ITS IMPLICATIONS IN EUROPE

H.A.
Orest Alickolli



BACKGROUND

Data Analytics is becoming an integral part of the electoral process in recent years.



Donald J. Trump ✓
@realDonaldTrump

I call my own shots, largely based on an accumulation of data, and everyone knows it. Some FAKE NEWS media, in order to marginalize, lies!



Le Pen is on course to be France's next president, fur manager says from AI analysis

Marine Le Pen is on course to be the next president of France, according to

CNBC.COM | BY CNBC



BACKGROUND

- The election of Donald Trump has taken the world by surprise and is fuelling populist movements in Europe, e.g. in Italy, Austria and France. Understanding populism and assessing the impact of the “Trump effect” on Europe is a **tremendous challenge**, which this presentation attempts to tackle.
- The goal is to find out where the next wave of populism could hit in Europe by **comparing and contrasting US and EU voter profiles**, opinions of Trump vs. non-Trump voters, Brexit vs. Bremain voters, and future expectations.
- Questionnaire drawn by multiple sources which include the 'angry voter' profile laid out by *Douglas Rivers*, the influence of political and moral attitudes pointed out by *Jonathan Haidt* and the profile of "America's forgotten working class" by *J.D. Vance*.

ANALYSIS PLAN

- **Goals**

- Evaluate the most important predictors which are likely to lead a voter towards a conservative populist choice.
- Models trained based on response on vote given in past **US presidential elections** and in the **British referendum** to leave the European Union.
- Attempt to predict the mood and future tendencies of five major European countries.

- **Sampling Methods**

- Data obtained from Dalia Research, a German tech company specializing in understanding global audiences based on a survey conducted on December 2016.
- Expanding their quarterly “*EuroPulse*” omnibus survey to the USA, Dalia has conducted a representative survey with respondents across all 28 EU member countries and respondents from the United States of America.
- Users are anonymously profiled across key demographic attributes and selected according to census statistics.

ANALYSIS PLAN

• Data Description

• Survey Response

- All predictors are categorical with the exception of age.
- The initial set contained 68 predictors with an aggregate of 377 levels.
- Two datasets:
 - US – sample size 1,052
 - EU – sample size 11,283
 - UK – sample size 1,358
 - France, Germany, Spain, Italy, Poland
 - Rest of EU

• Codebook

- Each level of the categorical responses is associated with a response provided in an attached codebook.

A	D	E	F	G	H	I	J	K
[meta] uuid	[dem] age	[dem] gender	[dem] education_	[dem] degree_of	[dem] residency	[dem] house	[dem] immig	[dem] origin
4c7889c0-9d27-	26	1	5	1	1	2	4	
8442eca0-9d27-	65	1	4	2	1	3	3	1
fb251930-9d2a-	39	2	3	1	1	3	4	
a9cced10-9d30-	19	2	5	1	3		3	13
23406820-9d31-	32	2	4	2	1	2	4	
9c0ed560-9d32-	40	2	3	2	1	2	4	
72e4b580-9d33-	17	1	3	1	1	4	4	
a6e9fb40-9d33-	14	1	2	2	1	4	2	9
fb70d20-9d33-	33	2	4	1	1	1	2	5
3c0a55d0-9d34-	17	1	2	2	1	4	5	
d2516540-9d37-	26	1	5	1	1	3	5	
dfeece40-9d39-	46	2	4	1	1	2	4	
397cc480-9d3a-	51	1	4	1	1	3	4	
18c9bce0-9d3d-	27	2	3	1	1	4	5	
5b1cfde0-9d3d-	28	1	4	1	2	3	5	
8a15f870-9d3d-	36	2	4	1	1	1	1	5
99de5b30-9d3d-	54	2	4	1	1	2	3	1
4ee54520-9d3e-	17	2	2	1	1	5	4	
1d3b6a00-9d3f-	35	2	4	2	1	4	3 2 1 5	



Dalia

Dalia Research GmbH
Skalitzer Str. 100
10997 Berlin | Germany
Phone: +49(0) 30 695 67 373
www.daliaresearch.com

5 Strongly disagree

[question]

government_controlled_
by_elite

Do you agree or disagree with the following statement:
"Government is controlled by the rich elite."

1 Strongly agree

2 Somewhat agree

3 Neither agree nor disagree

4 Somewhat disagree

5 Strongly disagree

[question]

trust_in_own_judgment

Do you agree or disagree with the following statement:
"I'd rather trust my own judgement than the opinion of experts."

1 Strongly agree

2 Somewhat agree

3 Neither agree nor disagree

4 Somewhat disagree

5 Strongly disagree

ANALYSIS PLAN

- **Data Cleaning**

- Coerce to one single answer questions with multiple responses (ex. Social Networks regularly used)

	gender	social_networks_regularly_used
	<int>	<fctr>
40	1	1 12 8 10 9 11 7 5 2 6 4 14
41	2	4 13 5 2 3 1
42	2	1 6
43	1	1 4 3
44	1	4 13
45	2	4 6 13 1
6 rows		

- This type of data recording creates 529 different levels based on multiple combinations for the social network variable alone.
- The US and EU data is subset based on observations from residents (assuming that only residents can vote).

ANALYSIS PLAN

- **Data Cleaning**

- Transform levels from numerical code to the actual textual response.

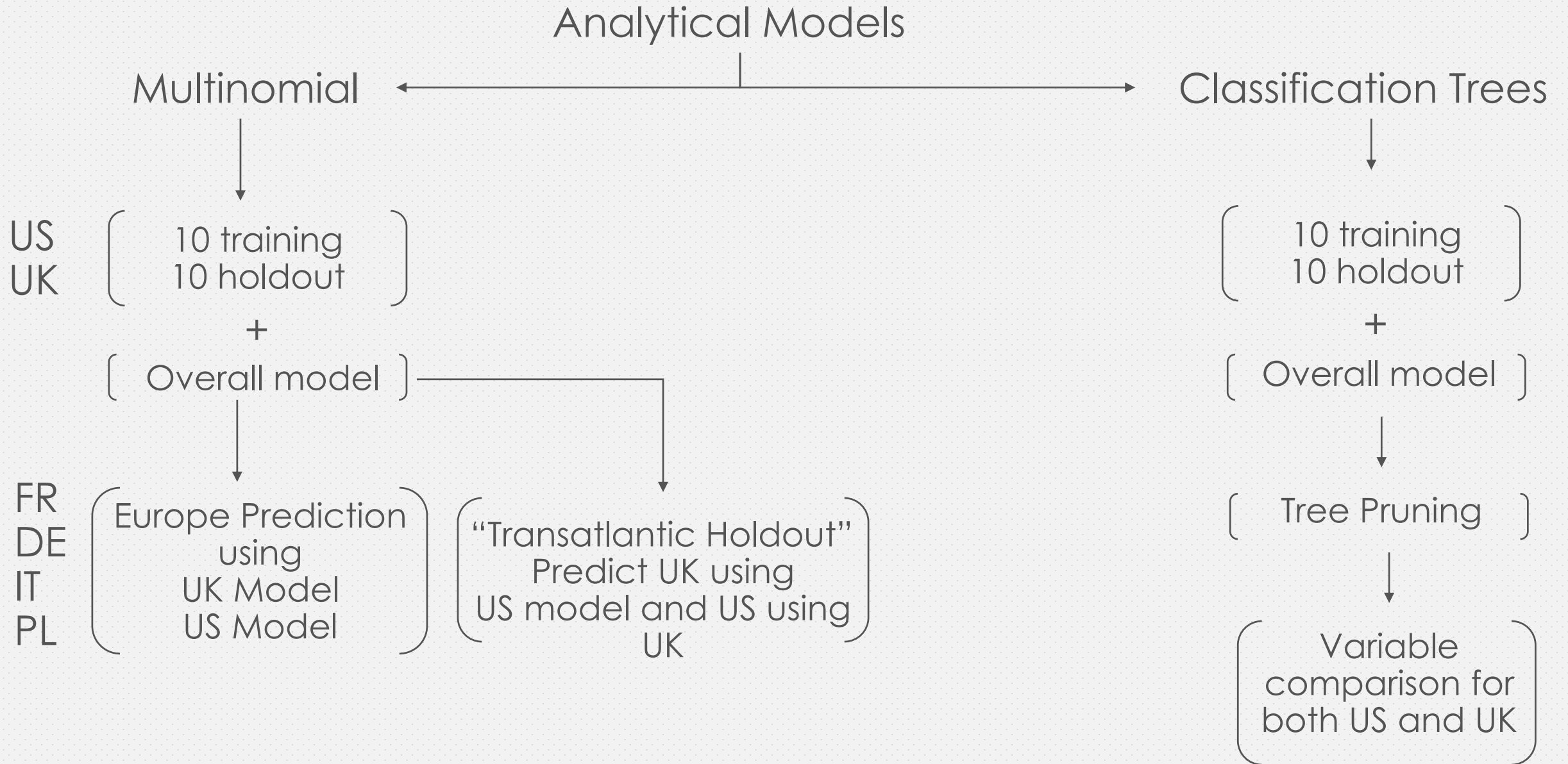
	gender <int>	education_level <int>	degree_of_urbanisation <int>
1	1	5	1
2	1	4	2
3	2	3	1
4	2	4	2
5	2	3	2
6	1	3	1

This facilitates interpretation, especially with classification trees.

	gender <fctr>	education_level <fctr>	degree_of_urbanisation <fctr>
1	male	Other/I'd rather not answer this question	City
2	male	I have completed a university or equivalent degree	Rural area
3	female	completed high school or obtained an equivalent diploma	City
4	female	I have completed a university or equivalent degree	Rural area
5	female	completed high school or obtained an equivalent diploma	Rural area
6	male	completed high school or obtained an equivalent diploma	City

- Reduced number of predictors from 68 to 47 by removing:
 - Predictors with too many missing values
 - Predictors deemed redundant.
 - Incompatible predictors between the US and EU set.

ANALYSIS PLAN



ANALYTICAL MODELS – MULTINOMIAL LOGIT

- US Model
 - Response - Vote on US Election
 - Donald Trump / Populist Conservative
 - Did Not Vote / Swing
 - Other Candidate / Liberal
 - Predictors – All 47 original predictors
- Best Model Selection
 - Forward Stepwise Model Selection based on minimizing AIC.
 - Number of predictors reduced to 20.

	aic <dbl>	deviance <dbl>
US_null_model	2213.888	2209.8880
US_full_model	1705.054	853.0542
US_step_model	1595.901	1163.9007

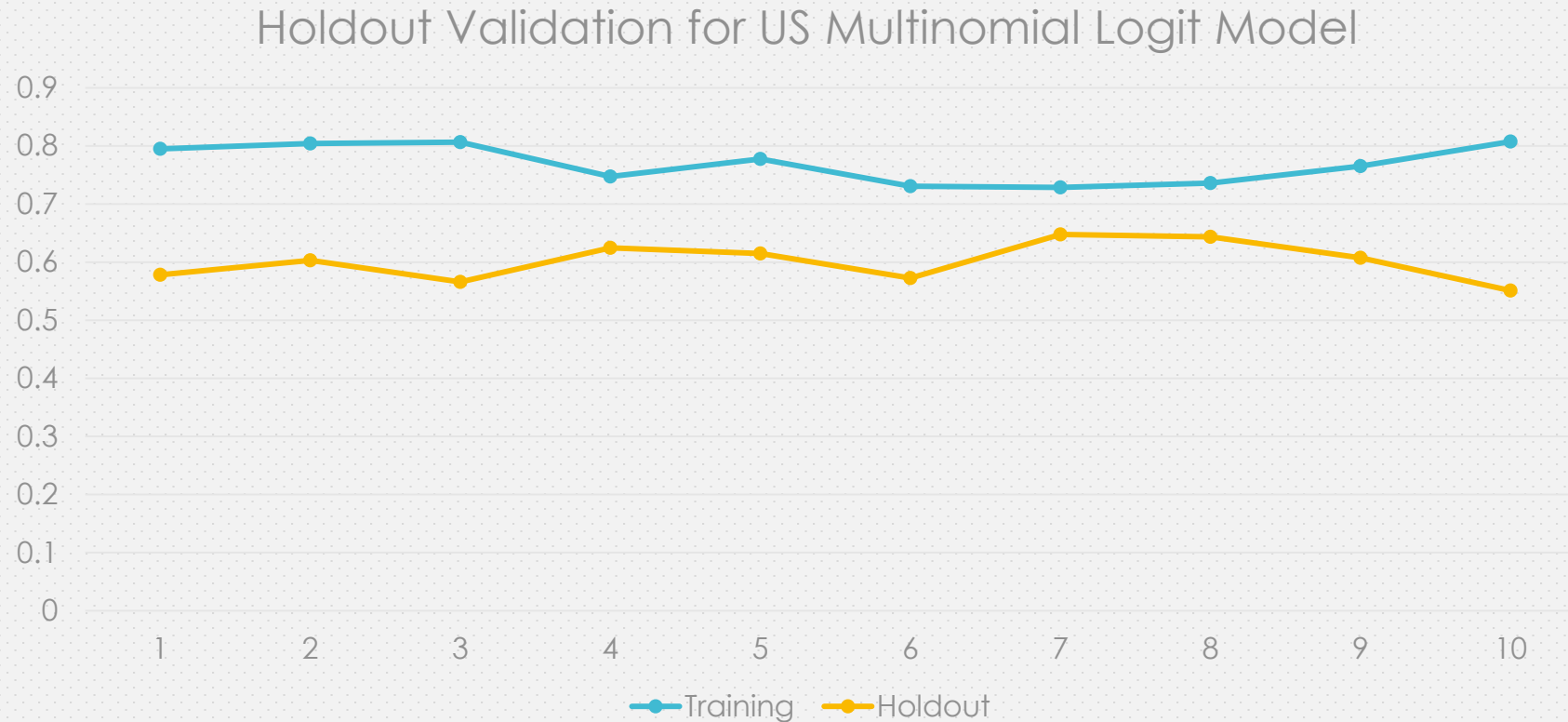
3 rows

Final Predictors – US Vote

economy_country_next12months
worldview
education_level
important_issues_when_voting
country_direction_past5years
employment_status
next_generation_opportunities
likelihood_to_demonstrate
media_tv_hours
opinion_government
change_economy_country_past12months
country_comes_first
family_friends_highereducation
preferred_type_of_political_leader
religion
lgbtq
currentplace_change_past5years
conspiracy
household_finances_past12months
government_controlled_by_elite

ANALYTICAL MODELS – MULTINOMIAL LOGIT

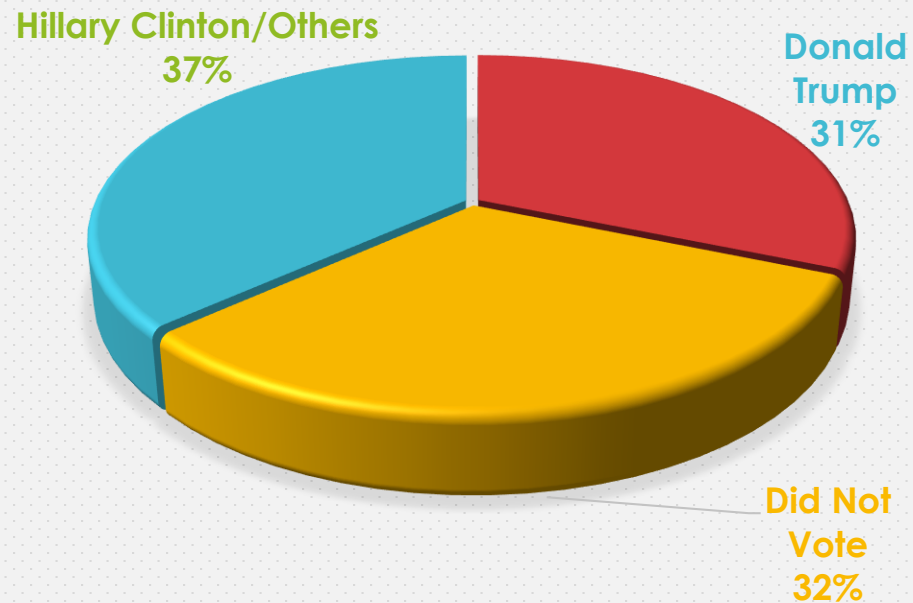
- US Model Holdout Validation
 - 80:20 split
 - 10 random samplings



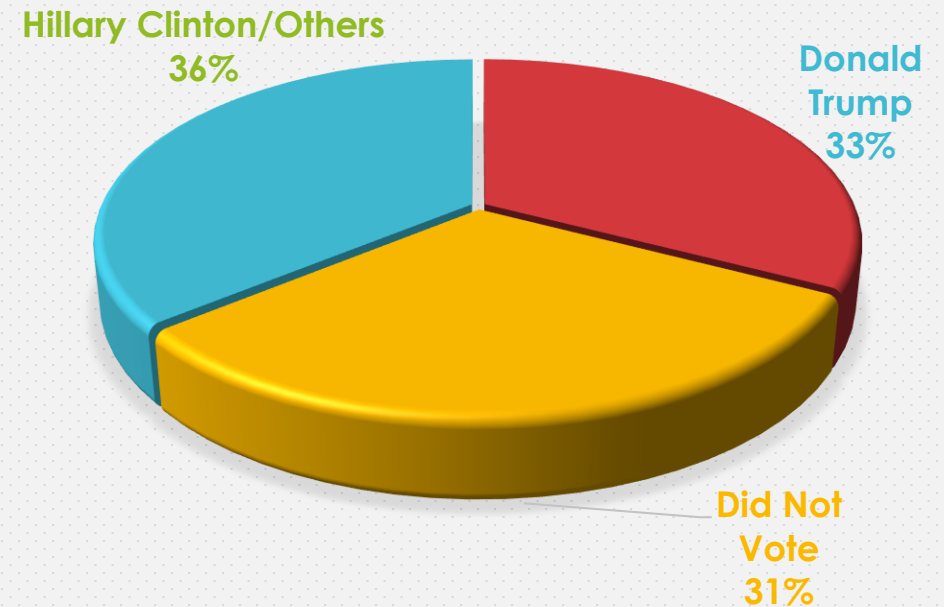
ANALYTICAL MODELS – MULTINOMIAL LOGIT

- US Overall Model Performance
 - Actual votes from the sample compared to predictions from the overall model

ACTUAL VOTES



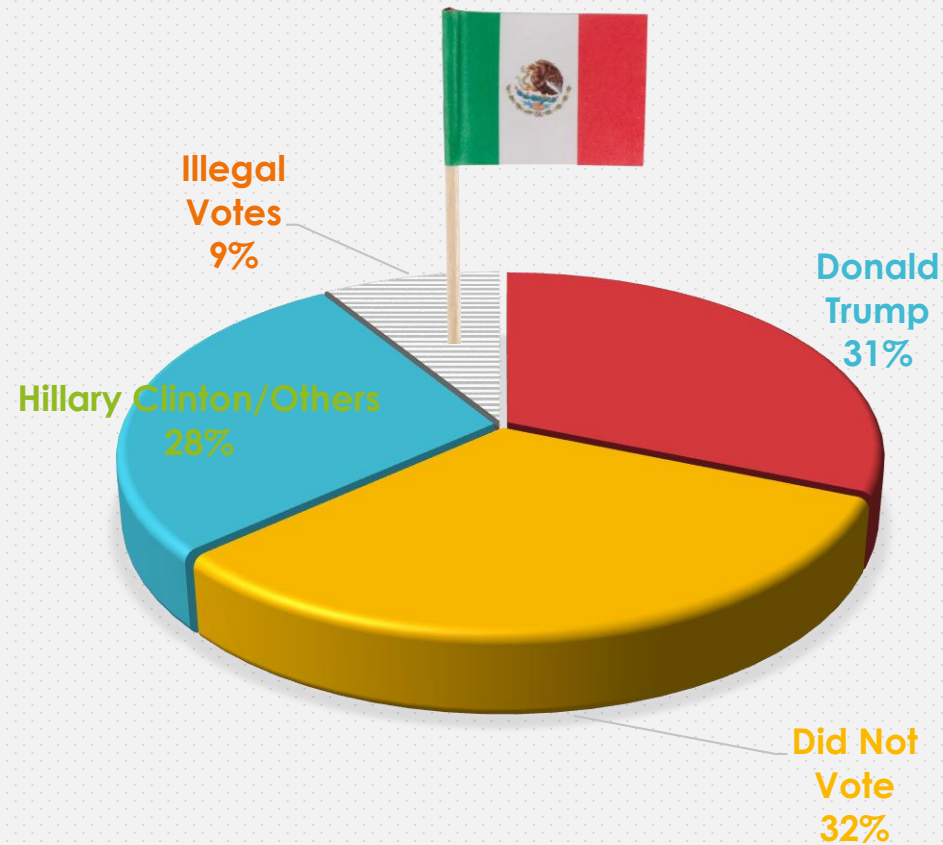
PREDICTED VOTES



ANALYTICAL MODELS – MULTINOMIAL LOGIT

- US Overall Model Performance

ALTERNATIVE MODEL



ANALYTICAL MODELS – MULTINOMIAL LOGIT

- UK Model
 - Response - Vote on British Referendum to Leave the EU
 - Leave the EU / Populist Conservative
 - Did Not Vote / Swing
 - Remain in the EU / Liberal
 - *Predictors – 47 original predictors*
- Best Model Selection
 - Forward Stepwise Model Selection based on minimizing AIC.
 - Number of predictors reduced to 16

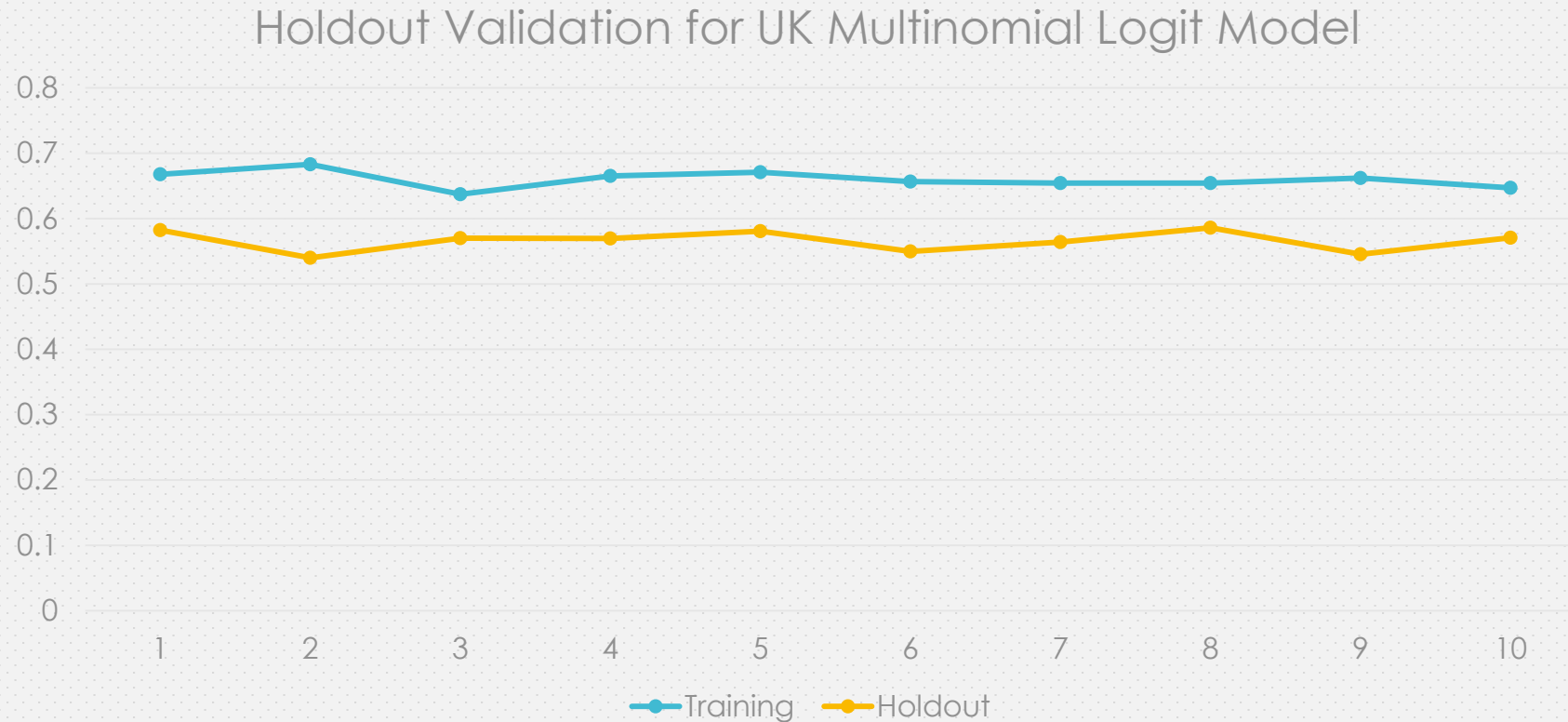
	aic <dbl>	deviance <dbl>
UK_null_model	2964.317	2960.317
UK_full_model	2583.505	1731.505
UK_step_model	2381.485	2097.485
3 rows		

Final Predictors - UK Vote

important_issues_when_voting
employment_status
perceived_effect_of_diversity
country_comes_first
economy_country_next12months
work_type_routine
currentplace_change_past5years
likelihood_to_demonstrate
age
conspiracy
worldview
consideration_or_good_behaviour
education_level
country_direction_past5years
improving_life_by_hardwork

ANALYTICAL MODELS – MULTINOMIAL LOGIT

- UK Model Holdout Validation
 - 70:30 split
 - 10 random samplings



ANALYTICAL MODELS – MULTINOMIAL LOGIT

- UK Model Overall Model Performance

ACTUAL VOTES



PREDICTED VOTES



ANALYTICAL MODELS – MULTINOMIAL LOGIT

- Comparison of predictors between US and UK set
- Variables that factor in the decision to vote “populist” vs. other are vary between US and UK reflecting cultural differences and peoples aspirations.

Final Predictors - UK Vote	Final Predictors – US Vote
important_issues_when_voting	economy_country_next12months
employment_status	worldview
perceived_effect_of_diversity	education_level
country_comes_first	important_issues_when_voting
economy_country_next12months	country_direction_past5years
work_type_routine	employment_status
currentplace_change_past5years	next_generation_opportunities
likelihood_to_demonstrate	likelihood_to_demonstrate
age	media_tv_hours
conspiracy	opinion_government
worldview	change_economy_country_past12months
consideration_or_good_behaviour	country_comes_first
education_level	family_friends_highereducation
country_direction_past5years	preferred_type_of_political_leader
improving_life_by_hardwork	religion
	lgbtq
	currentplace_change_past5years
	conspiracy
	household_finances_past12months
	government_controlled_by_elite

ANALYTICAL MODELS – MULTINOMIAL LOGIT

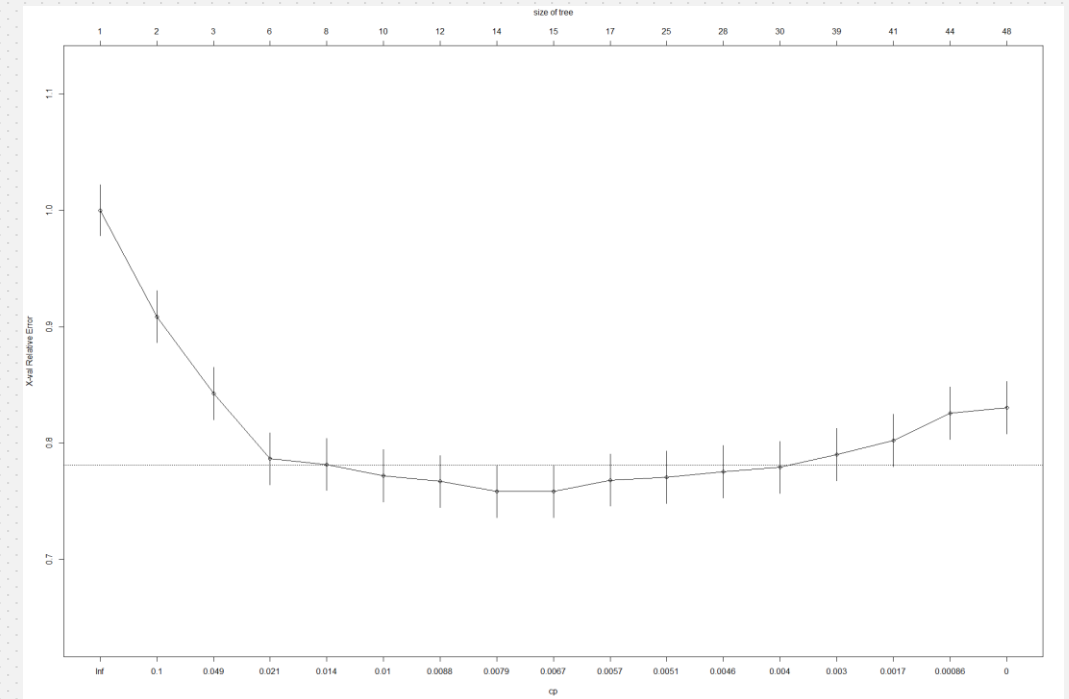
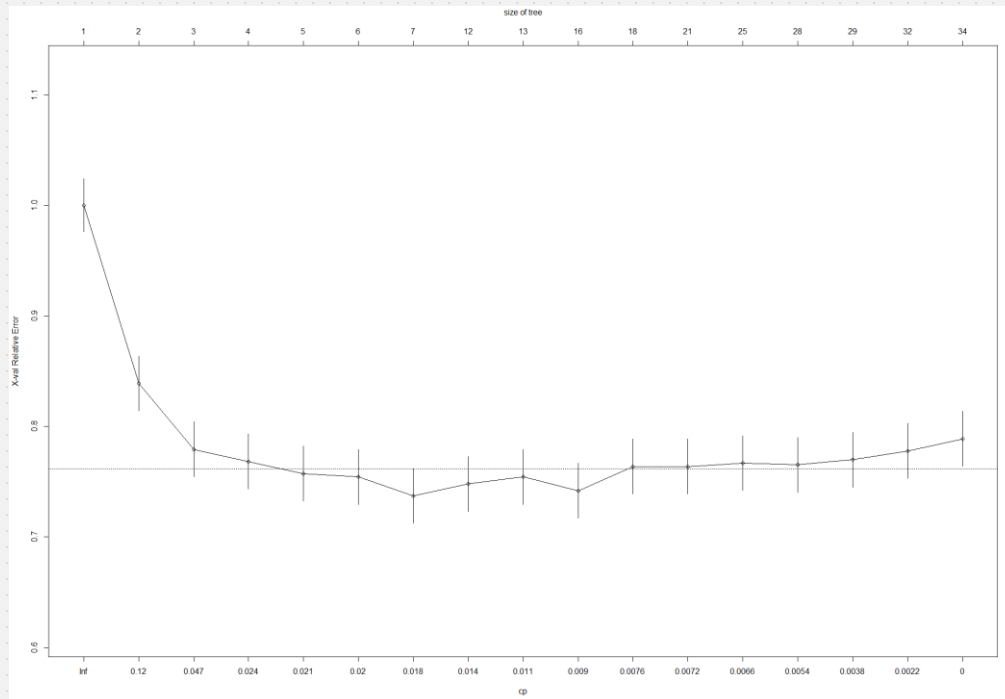
- "Transatlantic Holdout"
 - US and UK models were used to predict the outcome of the UK referendum and US Presidential Election, respectively.
 - UK model appears to have more predictive power on the US elections, than the opposite.

		actual		
predicted_UK_ref		For the UK to leave the EU	I did not vote	For the UK to stay in the EU
Donald Trump		0.322	0.184	0.188
I did not vote		0.317	0.571	0.351
other candidate		0.361	0.245	0.461

		actual		
predicted_US_elec		Donald Trump	I did not vote	other candidate
For the UK to leave the EU		0.517	0.235	0.238
I did not vote		0.140	0.410	0.154
For the UK to stay in the EU		0.343	0.355	0.607

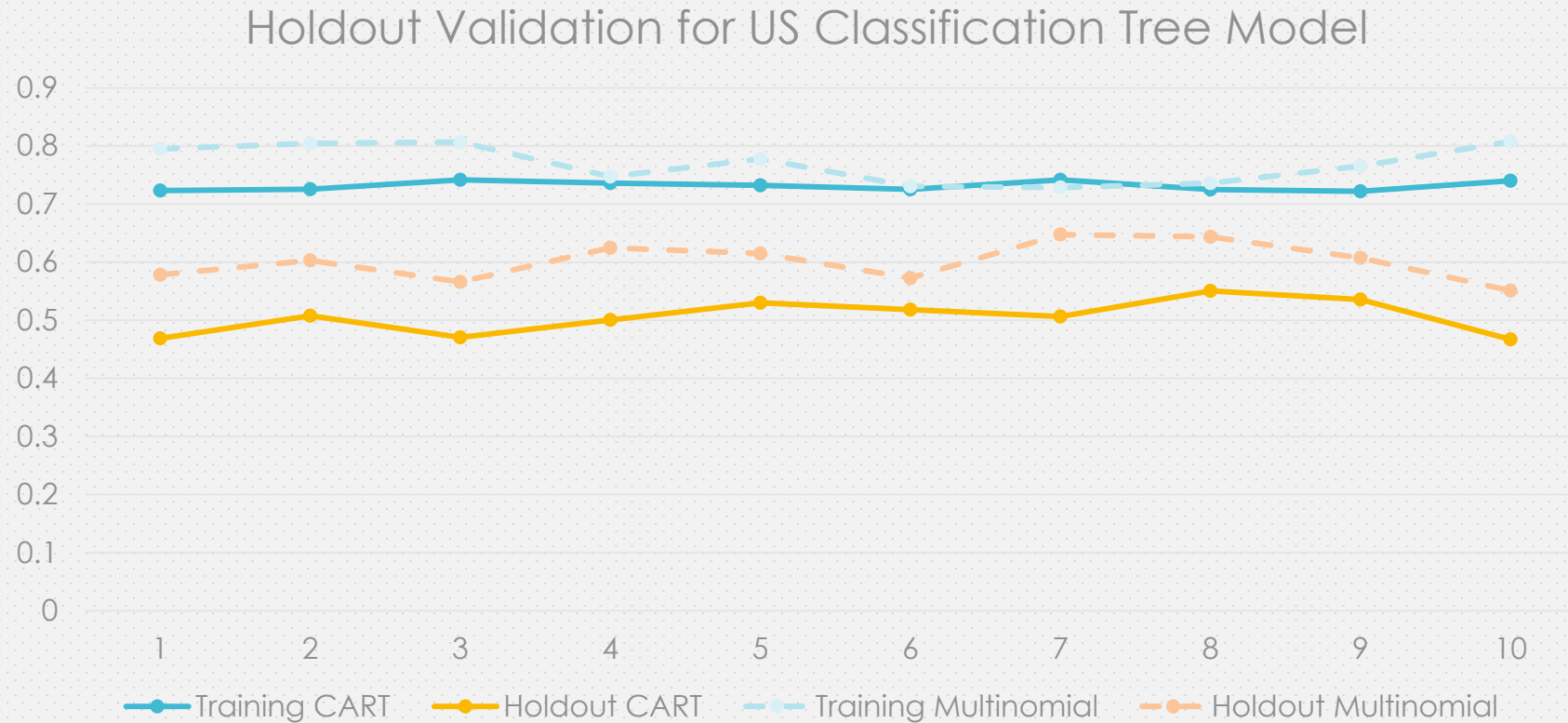
ANALYTICAL MODELS - CLASSIFICATION TREES

- Second Approach - CART
 - Two classification trees are grown for the US and UK dataset as an alternative model to the multinomial logit.
- Best Model Selection
 - Complexity parameter corresponding to lowest cross-validation error is selected and trees are pruned at this level.



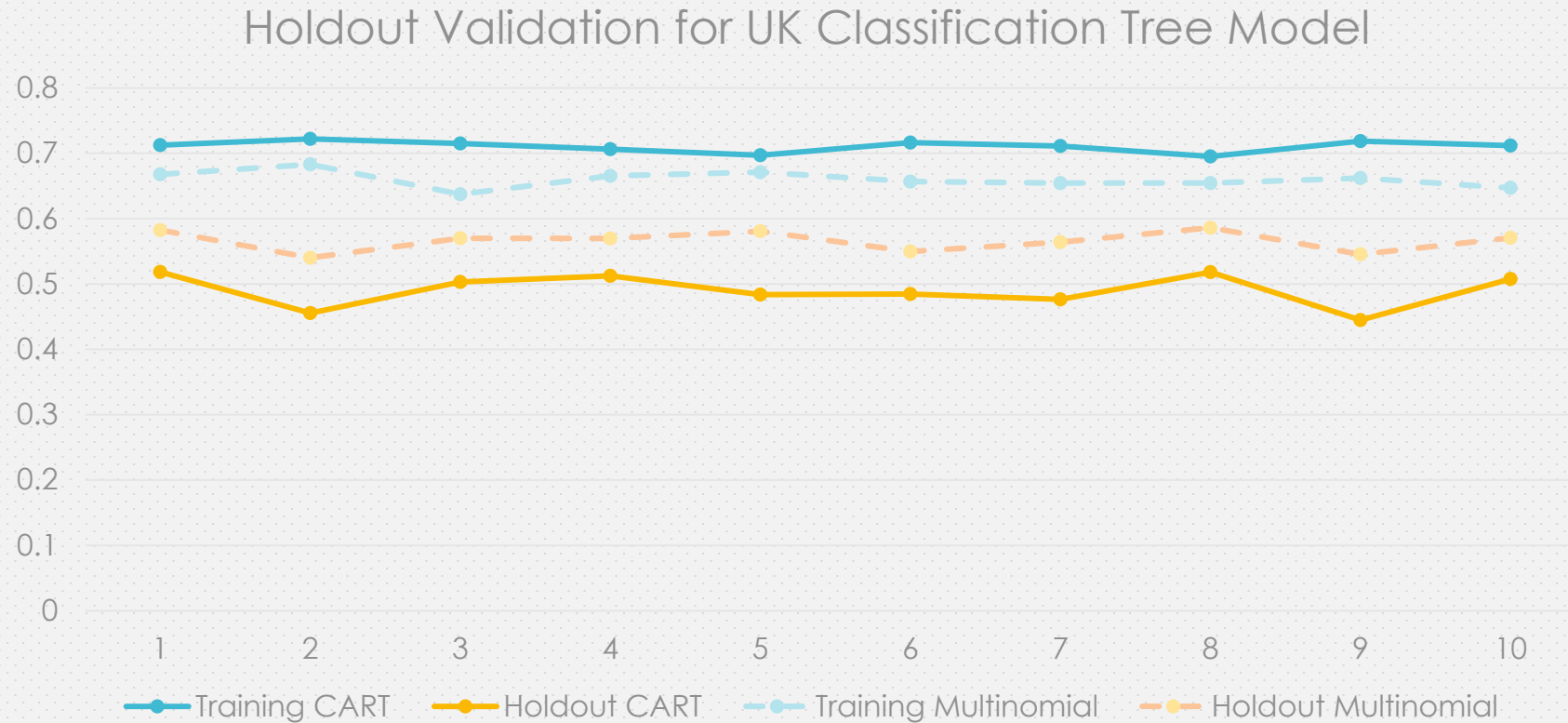
ANALYTICAL MODELS - CLASSIFICATION TREES

- US Model Holdout Validation
 - 80:20 split
 - 10 random samplings

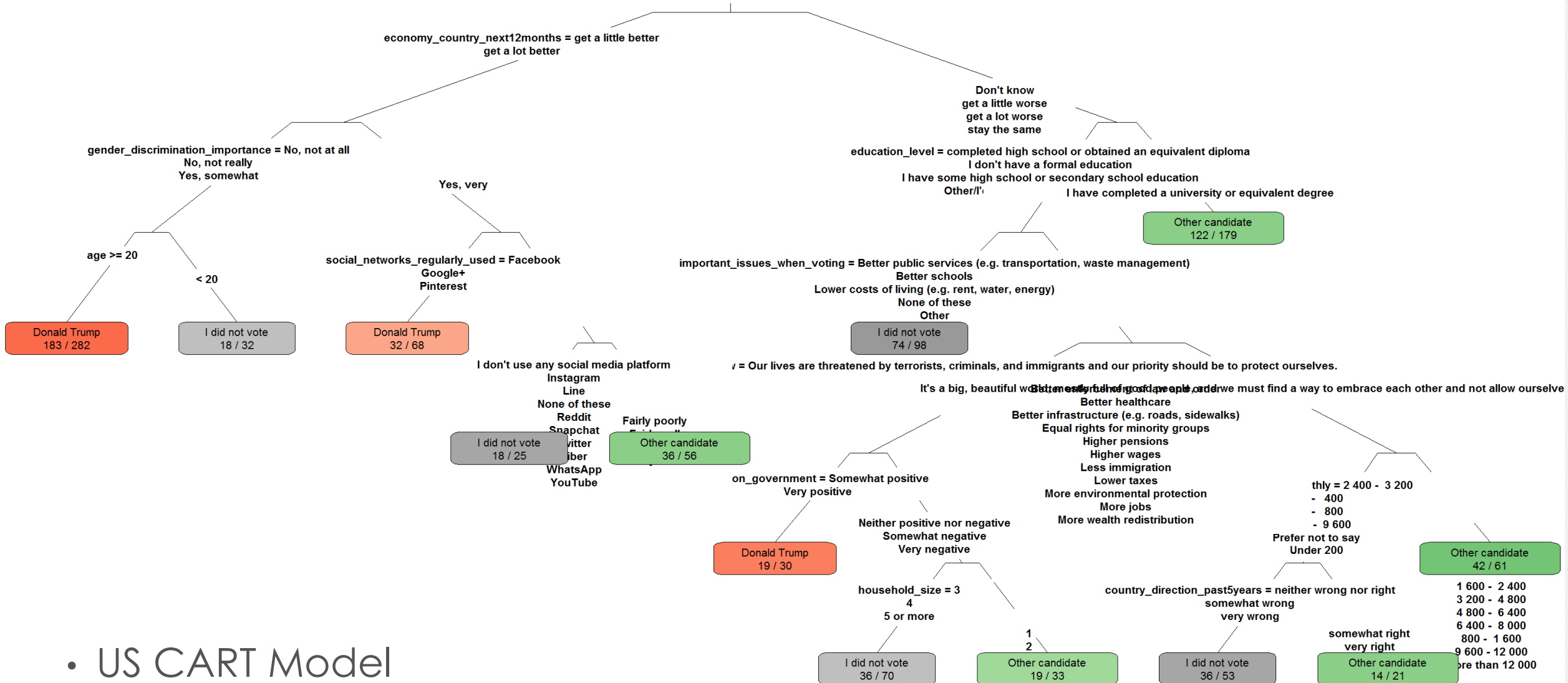


ANALYTICAL MODELS - CLASSIFICATION TREES

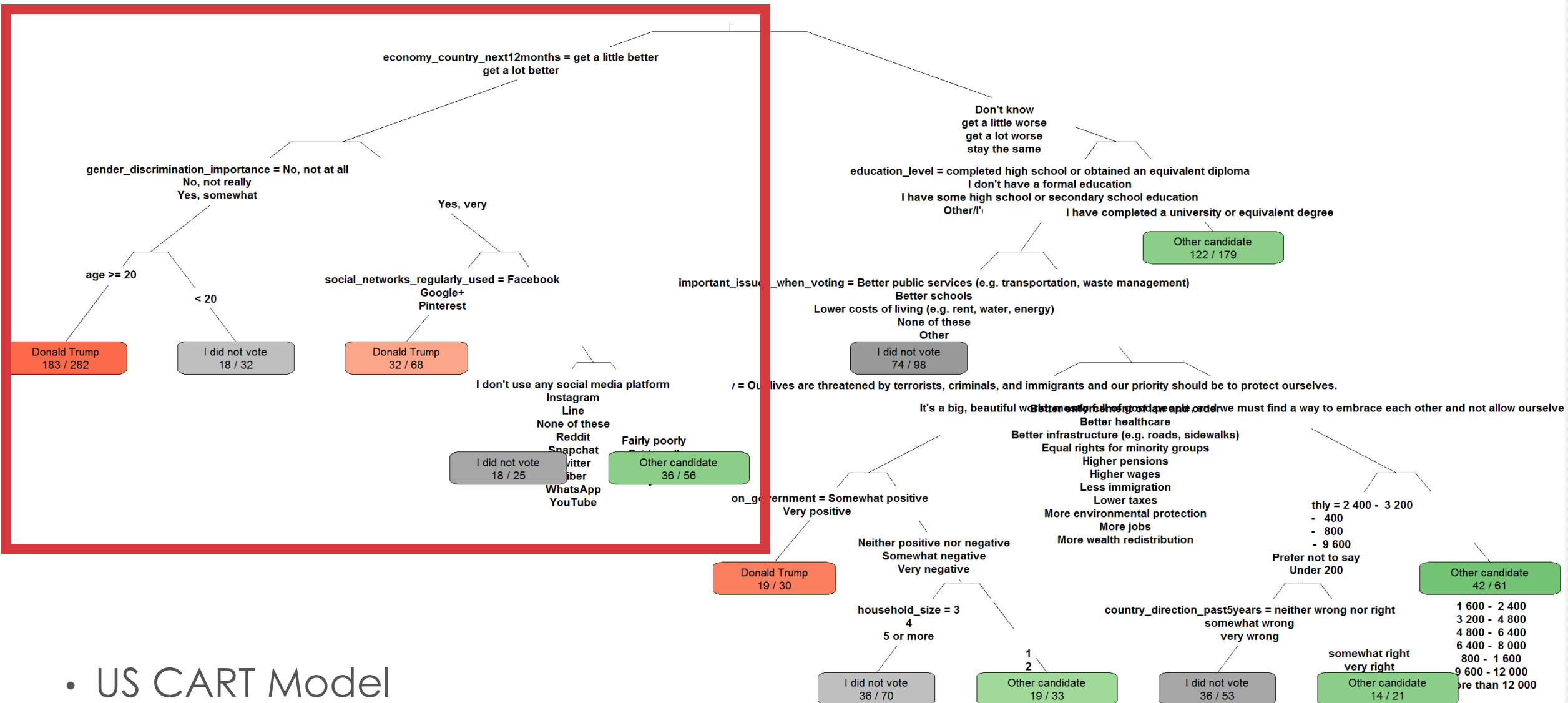
- UK Model Holdout Validation
 - 70:30 split
 - 10 random samplings



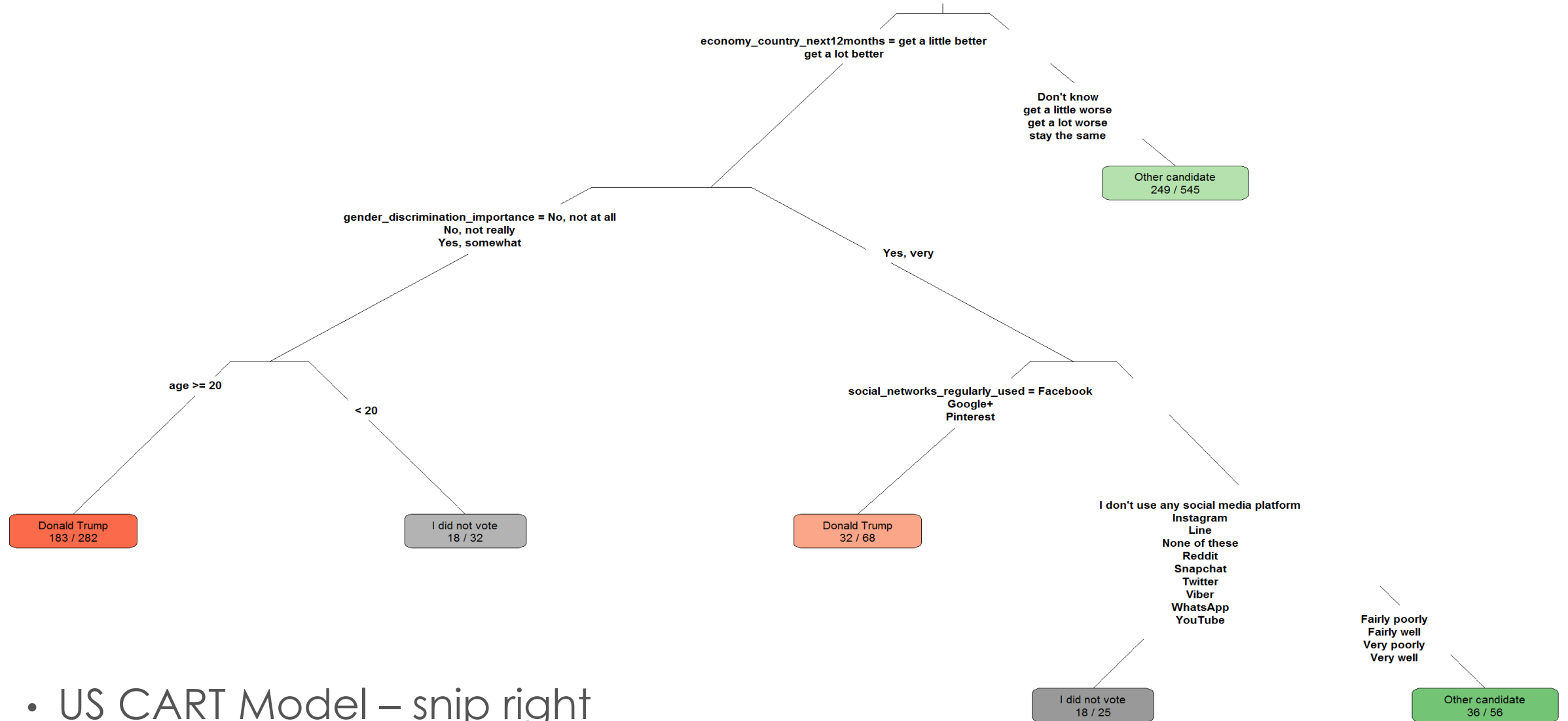
ANALYTICAL MODELS - CLASSIFICATION TREES



- US CART Model

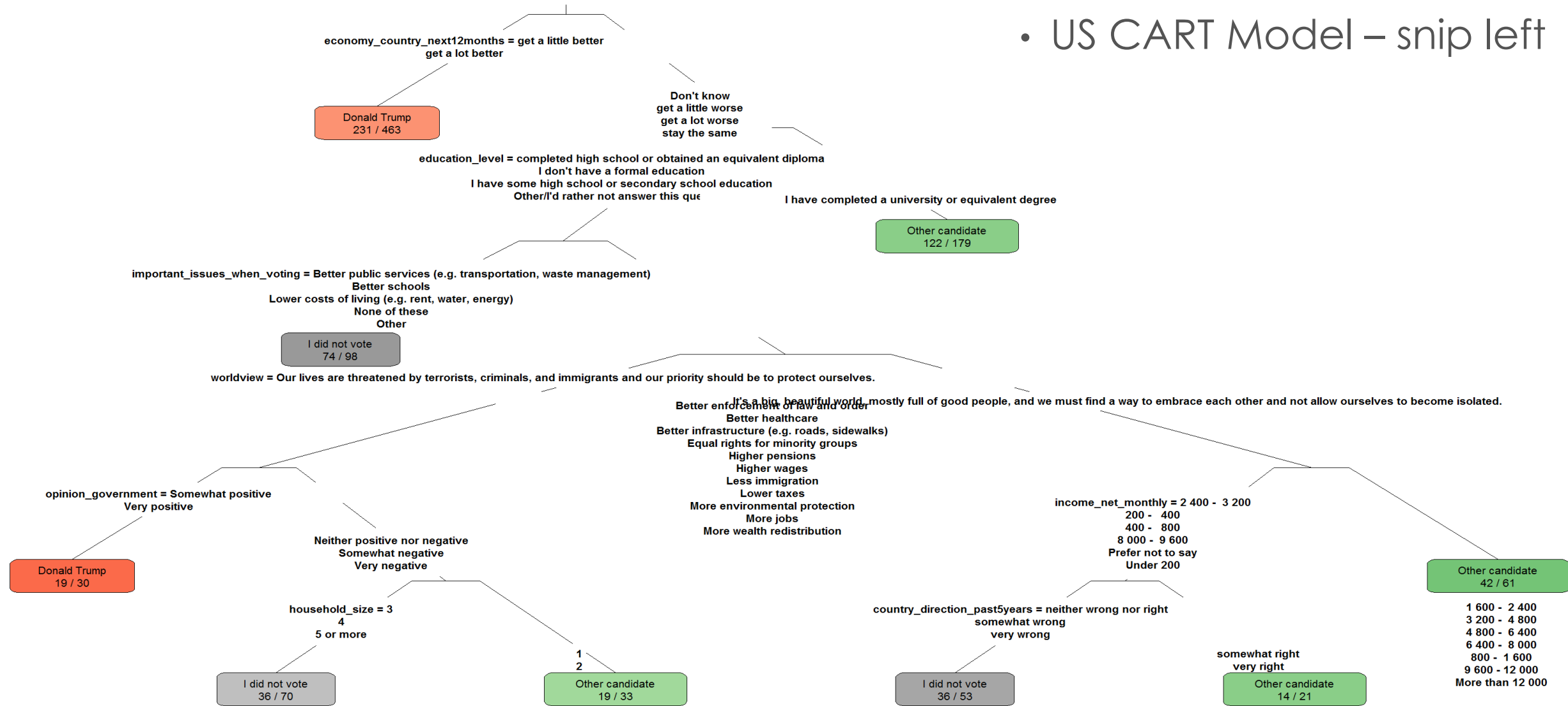


ANALYTICAL MODELS - CLASSIFICATION TREES



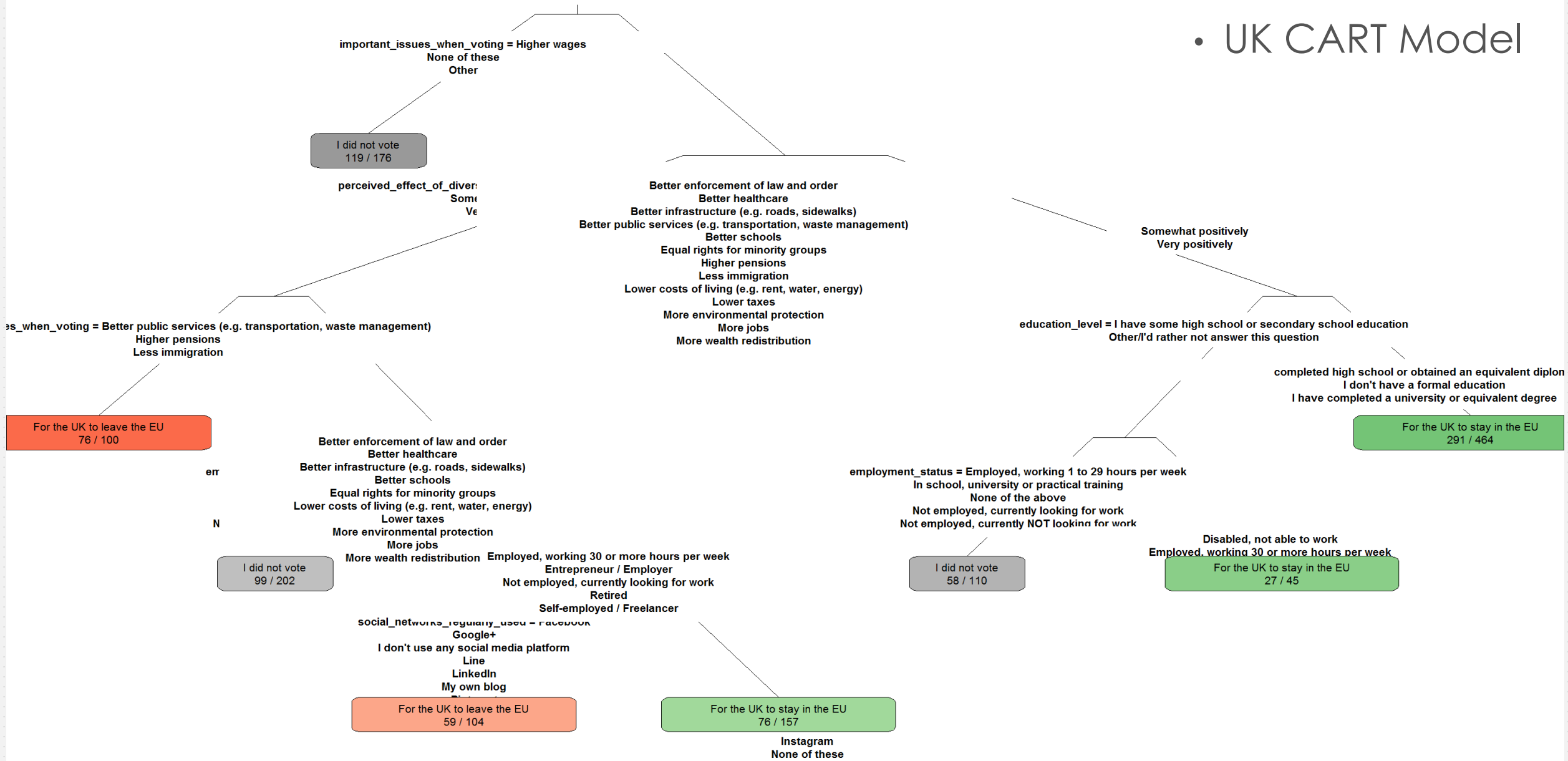
ANALYTICAL MODELS - CLASSIFICATION TREES

- US CART Model – snip left



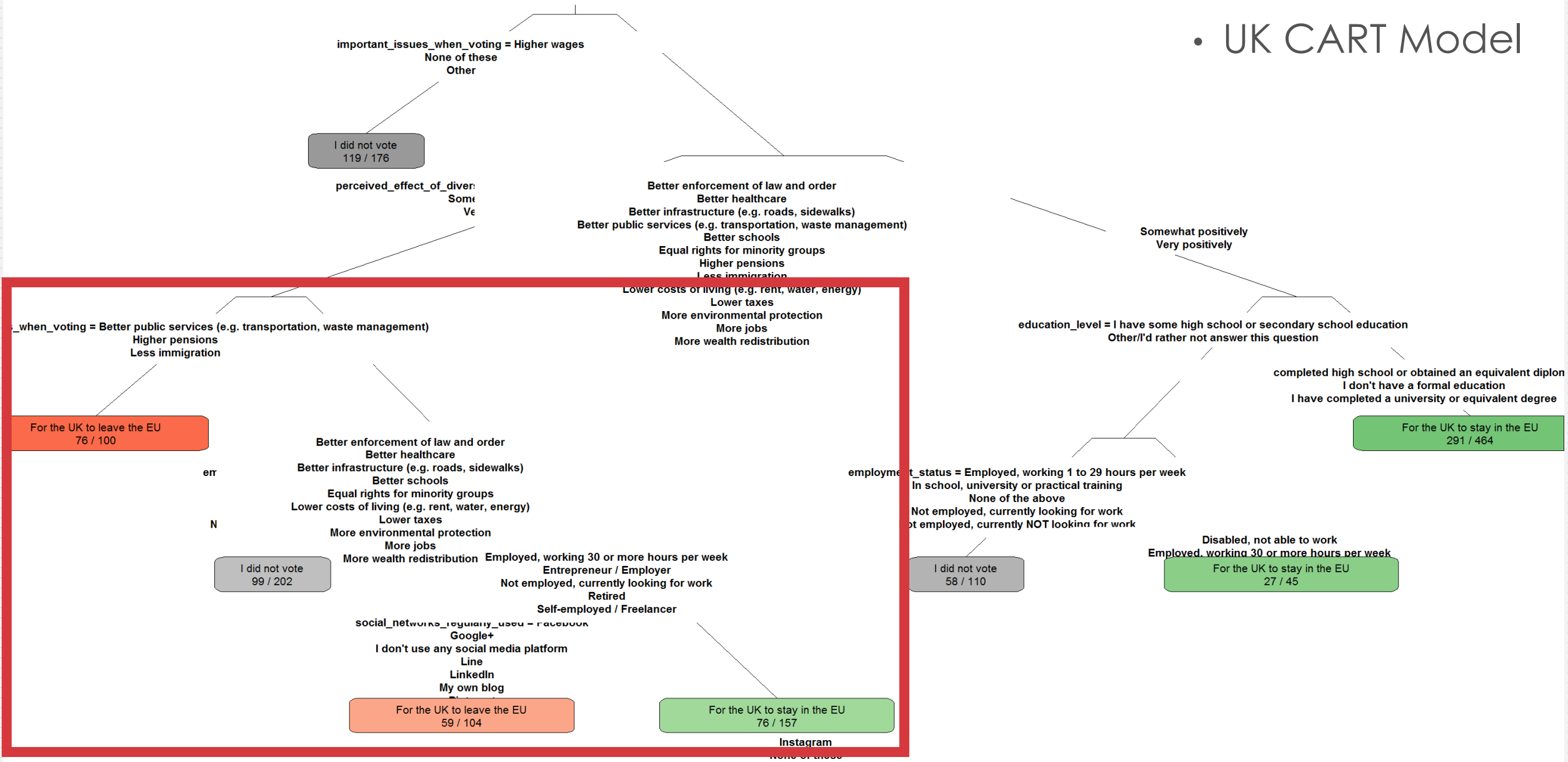
ANALYTICAL MODELS - CLASSIFICATION TREES

• UK CART Model



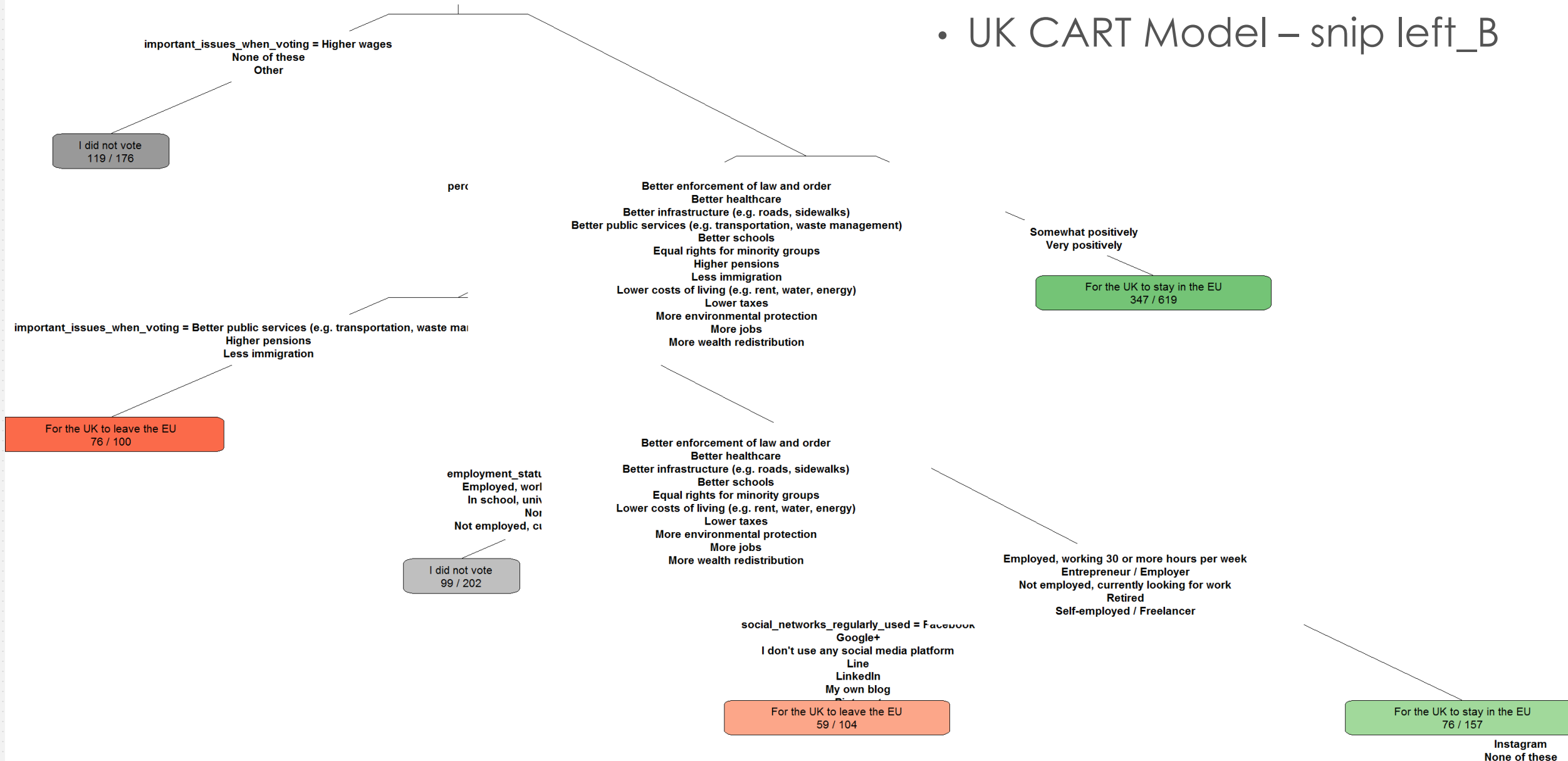
ANALYTICAL MODELS - CLASSIFICATION TREES

• UK CART Model



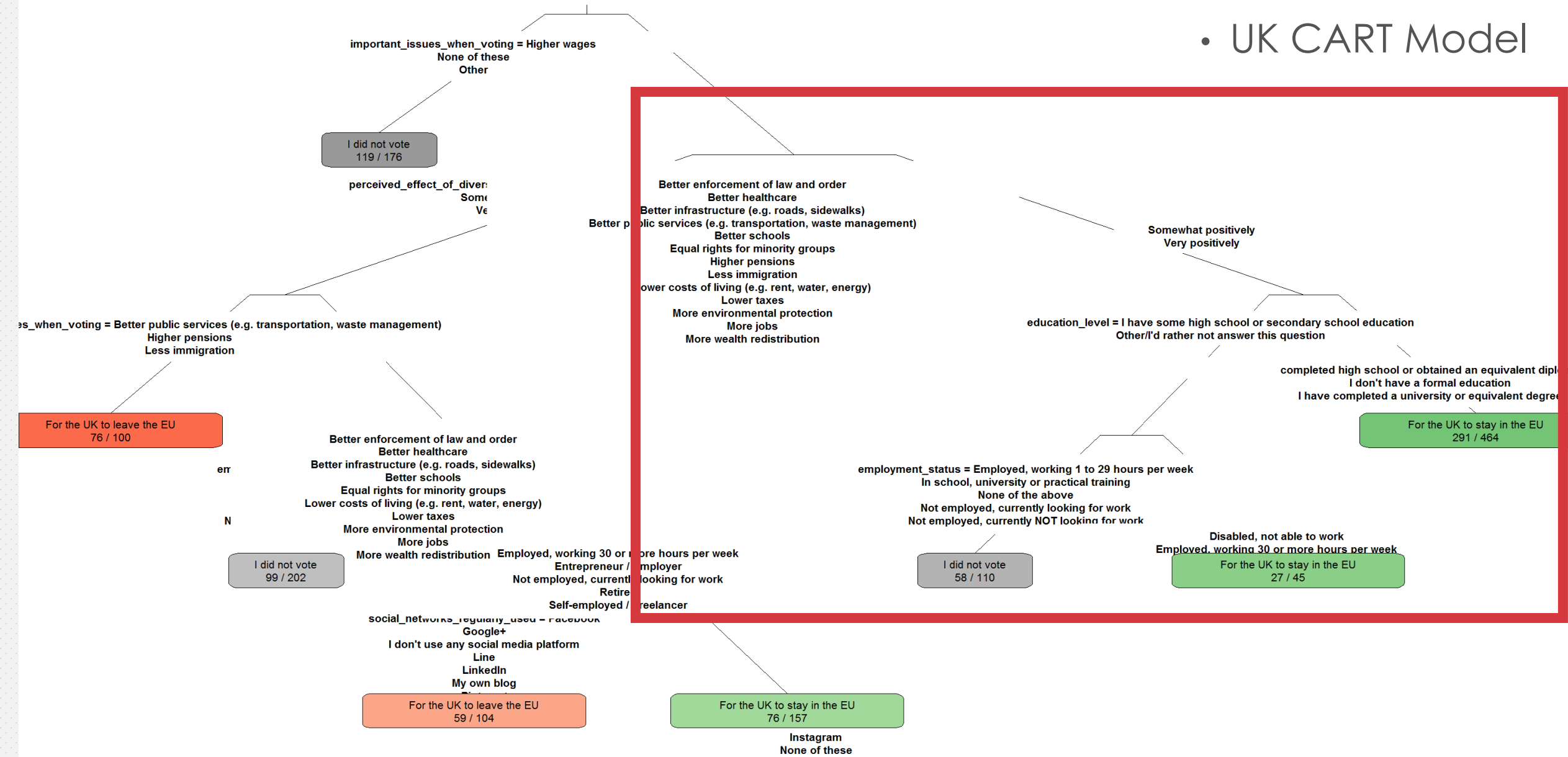
ANALYTICAL MODELS - CLASSIFICATION TREES

- UK CART Model – snip left_B



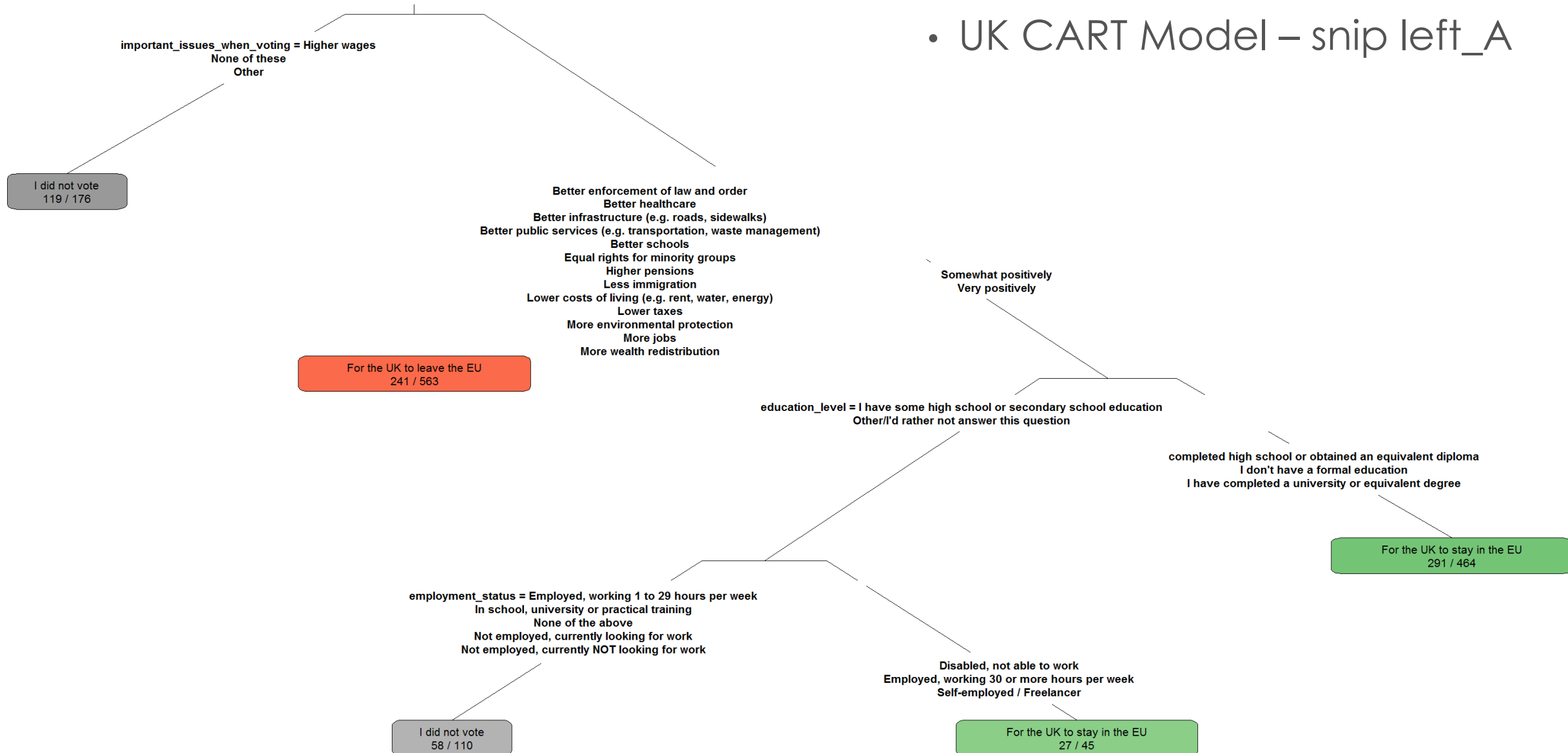
ANALYTICAL MODELS - CLASSIFICATION TREES

• UK CART Model



ANALYTICAL MODELS - CLASSIFICATION TREES

- UK CART Model – snip left_A



PREDICTIONS

- Multinomial Logit US and UK models were used to predict conservative populist tendencies in three major European countries.
- Based on the “Transatlantic Holdout” these results should be taken with more than a grain of salt.
- Since elections have not taken place yet in these countries, the ‘did not vote’ predictions are considered to be a swing vote and are assigned as such to the prediction with the second highest probability.

PREDICTIONS

- *France*
- The **Front National** leader (Marine Le Pen) is now **polling** at 26.5 per cent

	populist <dbl>	swing.populist <dbl>	swing.non.populist <dbl>	non.populist <dbl>
FR_UK_predictor	0.296	0.061	0.142	0.501
FR_US_predictor	0.183	0.114	0.280	0.423

2 rows

- France has a multi-party parliamentary system with currently 6 parties represented in the National Assembly.



PREDICTIONS

- *Germany*
- **Alternative for Germany** is a recent far right political party which has gained traction recently in Germany.

	populist <dbl>	swing.populist <dbl>	swing.non.populist <dbl>	non.populist <dbl>
DE_UK_predictor	0.335	0.064	0.161	0.440
DE_US_predictor	0.191	0.133	0.311	0.366

2 rows

- Germany also has a multi-party parliamentary system.
- Afd is polling currently at 11%.



PREDICTIONS

- *Italy*
- **5-Star Movement** is currently polling at 30% in Italy. The party was founded in 2009.

	populist <dbl>	swing.populist <dbl>	swing.non.populist <dbl>	non.populist <dbl>
IT_UK_predictor	0.393	0.047	0.088	0.473
IT_US_predictor	0.214	0.098	0.242	0.446

2 rows

- Italy also has a multi-party parliamentary system.
- The leader of the party, Beppe Grillo shown on the right, is a popular former comedian.



PREDICTIONS

- *Poland*
- **Law and Justice (PiS)** is currently polling at 35% in Poland. It currently holds 37.58% of the seats in the parliament.

	populist <dbl>	swing.populist <dbl>	swing.non.populist <dbl>	non.populist <dbl>
PL_UK_predictor	0.326	0.062	0.113	0.498
PL_US_predictor	0.210	0.063	0.184	0.544

2 rows

- Italy also has a multi-party parliamentary system with 6 major competing parties.
- With 235/460 seats in the Sjem and 56/100 in the Senate, it is currently the largest party in the Polish Parliament.



CONCLUSIONS

- **Stakeholders:**

- Electoral campaigns
- Political Organizations
- The Media
- Billionaires/Oil companies etc.

- **Cultural Differences**

- *The flavor of populism in the US is different than its counterpart in the EU.*
- *Despite the difference less tolerance to diversity, anti-immigration policies and isolationism are common trends between US and EU.*

- **Confounding Variables**

- The response to some of the questions might change drastically before and after the election (questions on future expectations)

THANK YOU
QUESTIONS?