

Understanding Voter Turnout Through Big Data Analytics

Kueh Chong Hua

Universiti Utara Malaysia

Jastini Mohd Jamil

Universiti Utara Malaysia

Izwan Nizal Mohd Shahraneer (✉ izwan.nizal@gmail.com)

Universiti Utara Malaysia <https://orcid.org/0000-0002-3566-7257>

Ang Jin Sheng

Universiti Utara Malaysia

Research

Keywords: Voter Turnout, Data Mining, Decision Tree Classification

Posted Date: April 5th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-380811/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Title Page

Title:	Understanding Voter Turnout through Big Data Analytics
Research Design:	This study's objective is to predict Malaysian turnout in the general election in Malaysia using classification tree algorithms. This research work has been approached as a binary classification problem; the model needed to predict whether each citizen entitles to vote would cast their vote in the general election. The Asian Barometer Survey 2010 and 2014 datasets were used in this study. A decision tree algorithm with a different experimental setting of CHAID, CART, and C5.0 was used in this study. The best prediction model was selected based on several popular performance matrices. The important variables in signaling voter turnout were reveals based the best model selected.
Complete author and co-author email addresses, ORCID and initial.	<ol style="list-style-type: none">1. Kueh Chong Hua (KCH) chonghua@merdeka.org2. Jastini Mohd Jamil (JMJ) https://orcid.org/0000-0002-6262-6223 jastini@uum.edu.my3. Izwan Nizal Mohd Shahraneer (INMS) https://orcid.org/0000-0002-3566-7257 nizal@uum.edu.my4. Ang Jin Sheng (AJS) angjinsheng@gmail.com
Author Affiliations	^{1,2,3,4} School of Quantitative Sciences, Universiti Utara Malaysia, 06010 UUM Sintok, Kedah, Malaysia.
Corresponding Author	Izwan Nizal Mohd Shahraneer (INMS) School of Quantitative Sciences, Universiti Utara Malaysia, 06010 UUM Sintok, Kedah, Malaysia Tel: +60125354691 nizal@uum.edu.my

Understanding Voter Turnout through Big Data Analytics

Abstract. In the concept of democracy, citizens can come out to vote and choose who they would prefer to lead the country in a competitive election. Malaysia is one of the countries that practicing democracy in shaping the country's future. There is a total of 14 general elections held in Malaysia. The turnout rate of the 2013 general election, 84.8% is the highest record in Malaysia history. However, the actual participation rate is considered lower when voter turnout is compared to voting age population and the number of eligible voters. Thus, this study's objective is to predict Malaysian turnout in the general election in Malaysia using classification tree algorithms. This dataset used in this study are the Asian Barometer Survey 2010 and 2014 dataset. The three selection decision tree algorithms used in this study are CHAID, CART, and C5.0. Among these three methods, CHAID perform the best to predict the Malaysian turnout in Malaysia's general election.

Keywords: Voter Turnout, Data Mining, Decision Tree Classification

1. Introduction

Malaysia has been practicing democracy and allow citizen to choose their desired leader to shape the country's future over sixty years. Malaysia has held 14 General Elections (GEs) after granted its independence on 31 August 1957. Generally, Malaysia's GE is held every five years. Malaysians who aged above 21 years old is eligible and responsible to vote on the condition that they need to register with Election Commission (EC).

However, Malaysia does not enforce compulsory voting. In year 2013, Malaysia's population who fulfil the requirements to vote are estimated 17.9 million people. However, only 13.2 million voters were registered with the EC as depicted in Table 1. Although there is a significant increase in terms of number of registered voters and number of voters turnout, the turnout rate is decreased if it is compared with voting age population (VAP).

Fowler (2013) observes that election outcomes and public policies can be change when the numbers of voters increases during the election. Thus, some policies may only be beneficial to particular groups of people if the political participation or voter's turnout is imbalance. As a result, government's performance as a whole will drop due to the policies are made not allot well and focus on specific cluster of people. Therefore, voter turnout plays an important role in deciding which political party or leader who will lead the country in a competitive election. It is crucial for both ruling and opposition political parties to identify who will vote because the numbers of voters decide their victory.

2. Related Works

2.1. Voter Turnout Model

There are three contextual political participation models used to explain the voting patterns in Malaysia including sociological model, political mobilization and psychological involvement. The sociological model describes voter's socio-demographic background, for instance age, gender and education. Political mobilization discloses the context of communication between political candidates and the citizen. On the other hand, psychological involvement defines the amount of interest about politics among voters.

2.1.1. Sociological Model

The sociological model suggests that sociodemographic variables such as age and education are important factors in explaining voter turnout. There are numerous studies have been conducted to connect socio-demographic factors and voting behavior in past decades. For instance, Mohd Hed and Grasso (2020) reveal that the numbers of young people are less compare to older people in participating

political engagement. Although education, ethnic group and gender does not affect political activism in Malaysia significantly, Blais (2000) and Tenn (2005) suggest that education is another factor that contribute to the rate of turnout in United States (US). Besides, Blais (2000) points out that more variables including government employees, marital status, religiosity and income are also correlated with political participation in US. In a study conducted by Norris (2004), demographics including age, gender, income level, education level, membership, religiosity and cultural attitudes are proven significantly related to voter turnout. Therefore, focus on societal context plays an important role to understand the voter turnout.

2.1.2. Political Mobilization

The concept of political mobilization is the people who practicing to manipulate the existing distribution of power (Nedelmann, 1987). Political mobilization model was developed by Rosenstone and Hansen (1993). They point out that campaigns and interpersonal conversations about politics have causal influence on voting. Political parties try to mobilize voters by using a variation of campaigning methods from offline and traditional to online and modern campaigning. Traditional campaigning requires many resources to get support and trust from the people (Farrell, 1996) whereas modern campaigning used information and communication to influence voters. Welsh (2018) claims that modern campaigning which utilize social media such as Facebook and Whatsapp encourages more people engage in politics and indirectly bring changes to country political landscape in Malaysia.

2.1.3. Psychological Involvement

The voter turnout is strongly related to the voter's political interest (Blais, 2007). In other words, the more interest a voter in politics, the more likely the voter will vote in an election. Verba, Schlozman, and Brady (1995) imply that educational and parental influence affect one person's political interest. In addition, Internet and social media are crucial variables in influencing voters' turnout. The Internet

is the fastest way to engage with young voters and contribute higher voter turnout among young voters (Hirzalla, Van Zoonen, & de Ridder, 2011). An interesting study conducted by Rauf, Hamid, and Ishak (2016) shows that a positive relationship between ability to access to political information and participation in politics. Wang (2007) maintain that the role of Internet as a medium in politics as a place of information seeking and opinion expressing when he conducted a study in year 2007. Thus, the internet plays an important role to enhance and give voter information about the elections, political parties and candidates. Furthermore, stimulate the turnout rate and participation of young people in an election. Put differently, the changes and improvement of communication and technology are influencing the electoral behavior and political landscape.

2.2. Decision Tree Classification Algorithms

2.2.1. CART

The Classification and Regression Trees (CART) is developed by Breiman, Friedman, Stone, and Olshen (1984). CART can be applied in both categorical and continuous data. CART performs binary splits and using Gini or Entropy splitting rules to achieve an optimal purity node.

The following are the steps taken to implement CART algorithm:

- (1) Start at the root node.
- (2) Find the split set that minimizes the sum of Gini indexes and use it to split the node into two child nodes to achieve an optimal purity node;

$$\text{Gini}(D) = 1 - \sum_{i=1}^n (p_i)^2 ; \quad (3.1)$$

where p_i is the relative frequency of class i in D ,

D is the dataset

- (3) If a stopping criterion is reached, then exit;
- (4) Prune the tree based on cost-complexity pruning
- (5) A test will be performed to examine the accuracy of the Model, if model evaluation criteria is not satisfying, repeat step 2 – 4.

2.2.2. CHAID

CHAID or Chi-Squared Automatic Interaction Detection is developed by Kass (1980). CHAID can only be conducted in categorical data. CHAID determines the size tree based on the Chi- Square test for independence.

The following are the steps taken in modelling using the CHAID algorithm:

- (1) Start at the root node.
- (2) Find the split set based on *Likelihood ratio Chi-Squared Statistics* and use it to split the node into two child nodes to achieve an optimal purity node;

$$\text{Chi Square, } \chi^2 = \sum_{j=1}^n \frac{(e-o)^2}{e}; \quad (3.2)$$

where e is the expected value, o is the observe value

p_i is the relative frequency of class i in D

D is the dataset

- (3) If a stopping criterion is reached, then exit;
- (4) Prune the tree based on chi-square test for independence
- (5) A test will be performed to examine the accuracy of the Model, if model evaluation criteria is not satisfying, repeat step 2 – 4.

2.2.3. C5.0

Quinlan (1993) combine and modify early version of C4.5 and ID3 and invented C5.0. The C5.0 offers new features such as the winnowing, boosting, generate smaller tree and unequal costs for different types of errors (Kuhn & Johnson, 2013). C5.0 uses information gain theory as the purity criterion to split the dataset and applied pessimistic pruning for the pruning process.

The following are the steps taken in implementing the C5.0 algorithm:

- (1) Start at the root node.
- (2) Find the split set based on *Entropy* measure and use it to split the node into two child nodes to achieve an optimal purity node;

$$\text{Entropy (D)} = - \sum_{i=0}^n -p_i \log_2 p_i; \quad (3.3)$$

where p_i is the relative frequency of class i in D

D is the dataset

- (3) If a stopping criterion is reached, then exit;
- (4) Prune the tree based on pessimistic pruning

A test will be performed to examine the accuracy of the Model, if model evaluation criteria is not satisfying, repeat step 2 – 4.

3. Research Procedure

3.1. Data

The Asian Barometer Survey (ABS) is collecting public opinion on issues such as political values, democracy and governance in 14 East Asia countries. Therefore, the dataset used in this study are secondary data that come from the third and fourth waves of ABS. The third wave and fourth wave of

ABS was carried in October year 2010 and year 2014 respectively. These surveys usually carried out after 17 or 18 months of Malaysia's general election. Table 2 summarizes the dimensions, attributes, measure level of the dataset that are selected for this study. Participation in the general election is the target variable in this study whereas the other 19 attributes are treated as predictor variables.

Table 2. Dimensions, Attributes and Measure Level of Dataset

	Dimensions	Attributes	Measurement Level
1	Participation in Elections	– Did you vote in the ... General Election?	– Nominal
2	Socio-economic background	– Gender	– Nominal
		– Ethnic group	– Nominal
		– Age group	– Ordinal
		– Marital status	– Nominal
		– Highest education level	– Ordinal
		– Religion	– Nominal
		– Employment status	– Nominal
		– Settlement	– Nominal
3	Political Mobilization	– Attend a campaign meeting or rally?	– Nominal
		– Try to persuade others to vote for a certain candidate of party?	– Nominal
		– Try to help out or work for a party or candidate running in the election?	– Nominal
		– <i>*Did you or a member of your family receive BRIM or any other government hand-out before the election period?</i>	– Nominal
		– Receive any reward or anything in exchange for your vote or support from any party?	– Nominal

4	Psychological Involvement	– How interested would you say you are in politics?	– Nominal
		– How often do you follow news about politics and government?	– Ordinal
		– When you get together with your family members or friends, how often do you discuss political matters?	– Ordinal
		– How often do you use the Internet including social media networks to find information about politics and government?	– Ordinal

* These questions only appear in ABS wave 3 Dataset

3.2. Experimental Setting

Figure 1 shows the experiment flow in this study. First stage of this study is preprocessing the data. The main reason of the existence of missing values is respondent refuse or unable to respond to the survey question. The missing values usually recorded as “No Response” (NR) or “Don’t Know” (DK) and will be filtered from the dataset. Then, the data is partitioning into training set and testing set with the ratio of 70% and 30% respectively. Testing set is used to validate the pattern generated from the training sample. The records in the training set are selected through simple random sampling method. The third stage is building the classification model through the datasets. The classification model include CART, CHAID and C5.0. The control parameter of each model is summarized in Table 3, Table 4 and Table 5 respectively. Lastly, the model is evaluated through accuracy, sensitivity, specificity, positive prediction value (precision), negative prediction value, Area Under ROC curve (AUC).

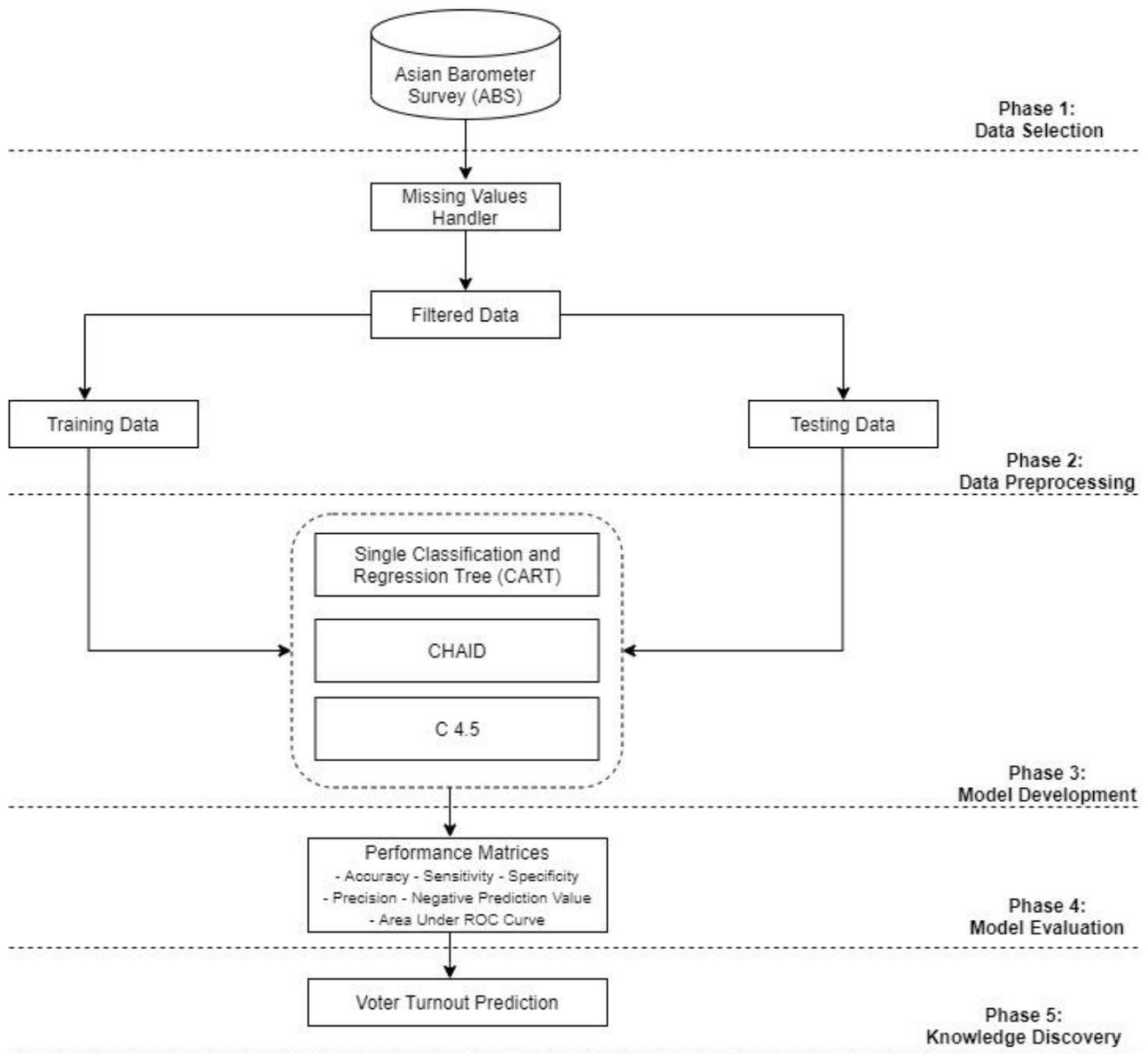


Figure 1. Experiment Setup

Table 3. Control parameters for CART algorithm

Parameters	Value	Description
minsplit	20	The minimum number of observations that must exist in a node in order for a split
minbucket	7	The minimum number of observations in any terminal node (which also equivalent to 1/3 of the minsplit)
maxcompete	4	The number of competitor splits retained in the output.
maxsurrogate	5	The number of surrogate splits retained in the output.
usesurrogate	2	Controls how surrogates are made use of in the model

xval	20	The number of cross-validations
surrogatestyle	0	Controls the selection of a best surrogate.
maxdepth	30	Limit the depth of a tree
cp	0.01	Control the size of the tree in order to select an optimal tree size.

Table 4. Control parameters for CHAID algorithm

Parameter	Value	Description
alpha 2	0.05	Level of significance used for merging of predictor categories
alpha 4	0.05	Level of significance used for splitting of a node in the most significant predictor
minprob	0.001	Minimum frequency of observations in terminal nodes
minsplit	20	Number of observations in split response at which no further split is desired
minbucket	7	Minimum number of observations in terminal nodes
Stump	False	Only root node splits are performed

Table 5. Control parameters for C5.0 algorithm

Parameter	Value	Description
Winnnow	TRUE	A logical: should predictor winnowing (i.e feature selection) be used?
noGlobalPruning	TRUE	A logical to toggle whether the final, global pruning step is needed to simplify the tree.
CF	0.5	A number in (0, 1) for the confidence factor. lower factor levels will likely prune away the leaves which over specify the classification
Mincases	7	An integer for the smallest number of samples that must be put in at least two of the splits
fuzzyThreshold	TRUE	A logical toggle to evaluate possible advanced splits of the data.
earlyStopping	TRUE	A logical to toggle whether the internal method for stopping boosting should be used.

3.3. Evaluation Metrics

The evaluation metrics used to measure the performance of classification models in this study are confusion matrix and area under curve (AUC) under receiver operating characteristic (ROC).

3.3.1 Confusion Matrix

A confusion matrix is a table that give information about the result of classification and actual values.

Table 6 shows a sample of confusion matrix. True negative (TN) and true positive (TP) represent the amount of voters that have been correctly classified between the actual and predicted values. Whereas, false negative (FN) and false positive (FP) are the number of voters that have been falsely predicted or incorrectly classified. Besides confusion matrix, the performance metrics of a predictive model is measured by accuracy rate, sensitivity rate, specificity, positive predictive value and negative predictive value. The formula of performance metrics stated is summarized in Table 7.

Table 6. Confusion Matrix Table

		Predicted	
		Negative	Positive
Actual	Negative	True Negative	False Positive
	Positive	False Negative	True Positive

Table 7. Performance Metrics

Performance Metric	Formula
accuracy rate	$\frac{TN + TP}{TN + FP + FN + TP}$
sensitivity rate	$\frac{TP}{FN + TP}$
specificity	$\frac{TN}{TN + FP}$
positive predictive value	$\frac{TP}{FP + TP}$
negative predictive value	$\frac{TN}{TN + FN}$

3.3.2 AUC under ROC

The Receiver Operating Characteristic can be illustrated in two-dimensional graph where sensitivity rate is plotted against one minus specificity rate (1- specificity). The graph contains values from 0 to 1 in each of the axis. Each point on the curve of the ROC curves provide the true-positive rate and false-positive rate and visually it in attractive way to summarize the accuracy of predictions. Using the same graph, the area under the ROC curve is called an Area Under Curve (AUC). The closer the value to 1, the better the performance of a classification model.

4. Results

Figure 2(A) – Figure 2(F) illustrates wave 3 and wave 4 rules representation for CART, CHAID and C5.0 classification models. Age variable is the first rule of the breakdown of the tree for the three classification models. In CART and C5.0, age is split into two groups, below 30 and above 30. On the other hand, CHAID split age variable into 21-30, 31-50 and above 50 years old. It is interesting to note that age below 30 years old who helped out or worked for a candidate or party in the election are most likely to vote in wave 3 dataset. In wave 4 dataset, young people who attending rally or campaign meeting have high probability to turn out and vote in an election.

Table 8 and Table 9 are the summary of model performance for wave 3 and wave 4 dataset respectively. Among three algorithms, CHAID perform the best because it provides 84.03% and 85.57% of accuracy on wave 3 and wave 4 dataset respectively. Although CART model has high amount of sensitivity (97.31%), the specificity is very low (33.85%) in wave 3 dataset. In wave 3 dataset, C5.0 have highest rate of specificity that is 60%. However, C5.0 have the lowest rate of negative predictive value in wave 3 dataset. C5.0 produce highest true positive at 97.15% in wave 4 dataset but has a very low specificity

at 13.33%. In other words, C5.0 predicts 86.67% of respondents most likely turnout in an election. The specificity of CHAID performs better compare to CART and C5.0 although the sensitivity of CHAID is slightly lower compare to CART and C5.0. Most model perform fairly good from the point of view of AUC under ROC curve in wave 3 and 4 dataset. The AUC ranges start from 0.5 to 1.0. The best rate of AUC under ROC curve is 0.9 to 1.0 whereas 0.5 means no predictive value.

Table 8. Summary of model performance for Wave 3 dataset

	CART	CHAID	C5.0
1 Accuracy [95% CI]	82.99% [78.14%, 87.14%]	84.03% [79.28%, 88.06%]	81.60% [76.63%, 85.90%]
2 Sensitivity (True Positive)	97.31%	95.07%	87.89%
3 Specificity (True Negative)	33.85%	46.15%	60.00%
4 Positive Predictive Value (Precision)	83.46%	85.83%	88.29%
5 Negative Predictive Value	78.57%	73.17%	59.09%
6 Area Under ROC Curve (AUC)	72.50%	80.00%	81.00%

Note: CI refers to confidence intervals

Table 9. Summary of model performance for Wave 4 dataset

	CART	CHAID	C5.0
1 Accuracy [95% CI]	84.88% [80.24%, 88.79%]	85.57% [81.00%, 89.40%]	84.19% [79.48%, 88.19%]
2 Sensitivity (True Positive)	95.62%	92.43%	97.15%
3 Specificity (True Negative)	17.50%	42.50%	13.33%
4 Positive Predictive Value (Precision)	87.91%	90.98%	85.97%
5 Negative Predictive Value	38.89%	47.22%	46.15%
6 Area Under ROC Curve (AUC)	73.30%	76.1%	67.85%

Note: CI refers to confidence intervals

5. Conclusions

In conclusion, three classification models including CART, CHAID and C5.0 are used to identify people who voted in the General Elections of Malaysia in the year of 2008 and 2013. Among three models, CHAID have the highest accuracy in both datasets compare to CART and C5.0. The other criteria of CHAID such as true positive, true negative, positive predictive, negative predictive value and AUC under ROC curve are acceptable as well. Age is the main predictor to differentiate voter turnout among young and older voters. In 2008 GE, young voters below 30 are less likely to vote if they are not interested to join or help during the election. The impact of rally towards young voters are high in GE 2013. On the other hand, factors that drive middle age group to vote are assist or help out the general election candidates or party, interest in politics and ability to join in the politics. In this study, not all variables can be used in predicting and classifying voter turnouts in Malaysia. For instance, sociological model variables such as religion, gender and employment status does not influence significantly across the algorithms used in this study.

6. Limitation and Future Research

There are few limitations in this study. The first limitation is only three theoretical models applied in examining the relationship between voter turnout and political participation models. In the future research, behaviour of voter turnout is suggested to include other theoretical models such as cultural modernization and rational choice theory. In terms of methodological aspect, there are many imbalance cases in both datasets. To illustrate, voter turnout in dataset 2013 have 84%. That is to say, only 16% people who did not vote in the GE year 2013. Thus, future study should consider this factor and apply some possible and feasible approaches such as resampling methods. Furthermore, more decision tree algorithm such as Support Vector Machine (SVM), random forest and Boosting C5.0 can be used and evaluated to predict voter turnout.

Declaration

Ethics Approval

Not Applicable

Consent to Participate

Not Applicable

Availability of Data and Material

Not Applicable

Competing Interests

To the best of my knowledge and belief this paper contains no material previously published by any other person except where due acknowledgment has been made. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. The authors have no conflicts of interest to declare that are relevant to the content of this article.

Funding

This research was funded by Universiti Utara Malaysia (UUM) through University Research Grant Scheme.

Authors' Contributions

Kueh Chong Hua extracted the data from Election Commission Malaysia. All authors conducted the analyses, wrote, and reviewed the manuscript.

Acknowledgements

We thank Universiti Utara Malaysia (UUM) for providing the research grant.

References

- Blais, A. (2000). *To vote or not to vote?: The merits and limits of rational choice theory*: University of Pittsburgh Pre.
- Blais, A. (2007). Turnout in elections. In *The Oxford handbook of political behavior*.
- Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*: CRC press.
- Farrell, D. M. (1996). Campaign strategies and tactics. *Comparing democracies: Elections and voting in global perspective*, 160-183.
- Fowler, A. (2013). Electoral and policy consequences of voter turnout: Evidence from compulsory voting in Australia. *Quarterly Journal of Political Science*, 8(2), 159-182.
- Hirzalla, F., Van Zoonen, L., & de Ridder, J. (2011). Internet use and political participation: Reflections on the mobilization/normalization controversy. *The Information Society*, 27(1), 1-15.
- Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 29(2), 119-127.
- Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling* (Vol. 26): Springer.
- Mohd Hed, N., & Grasso, M. T. (2020). Age group differences in political activism in Malaysia. *Journal of Youth Studies*, 23(6), 765-779.
- Nedelmann, B. (1987). Individuals and parties—Changes in processes of political mobilization. *European Sociological Review*, 3(3), 181-202.
- Norris, P. (2004). *Electoral engineering: Voting rules and political behavior*: Cambridge university press.
- Quinlan, J. (1993). C5. 0 machine learning algorithm. In.
- Rauf, A. A. A., Hamid, N. A., & Ishak, M. S. (2016). Modelling the effect of access to information, political interest and policy satisfaction on youth online political participation in Malaysia. *Jurnal Komunikasi: Malaysian Journal of Communication*, 32(2).
- Rosenstone, S. J., & Hansen, J. M. (1993). *Mobilization, participation, and democracy in America*: Longman Publishing Group.
- Tenn, S. (2005). An alternative measure of relative education to explain voter turnout. *The Journal of Politics*, 67(1), 271-282.
- Verba, S., Schlozman, K. L., & Brady, H. E. (1995). *Voice and equality: Civic voluntarism in American politics*: Harvard University Press.
- Wang, S.-I. (2007). Political use of the Internet, political attitudes and political participation. *Asian Journal of Communication*, 17(4), 381-395.
- Welsh, B. (2018). "Saviour" Politics and Malaysia's 2018 Electoral Democratic Breakthrough: Rethinking Explanatory Narratives and Implications. *Journal of Current Southeast Asian Affairs*, 37(3), 85-108.

Table 1: Historical Statistics of General Elections (1955-2013)

Date of General Election	VAP ¹ (‘000)	VEP (‘000)	Turnout (‘000)	Increase of voters (‘000)	Increase of voters (%)	Turnout (VAP) (%)	Turnout (VEP) (%)	Variance Turnout (%)
1955-07-27	-	1,280.9	1,060.6				82.8%	
1959-08-19	-	2,177.7	1,596.3	896.8	+70.0%	-	73.3%	-9.5%
1964-04-25	-	2,681.9	2,116.0	504.2	+23.2%	-	78.9%	+5.6%
1969-05-10	-	3,450.0	2,539.2	768.1	+28.6%	-	73.6%	-5.3%
1974-08-24	5,265.0	4,178.9	3,138.4	728.9	+21.1%	59.6%	75.1%	+1.5%
1978-07-08	6,067.2	5,059.7	3,810.0	880.8	+21.1%	62.8%	75.3%	+0.2%
1982-04-22	6,828.2	6,081.6	4,585.5	1,021.9	+20.2%	67.2%	75.4%	+0.1%
1986-08-02	7,893.9	6,791.4	5,052.8	709.8	+11.7%	64.0%	74.4%	-1.0%
1990-10-20	8,882.0	8,000.0	5,784.0	1,208.6	+17.8%	65.1%	72.3%	-2.1%
1995-04-24	10,175.0	9,012.4	6,155.5	1,012.4	+12.7%	60.5%	68.3%	-4.0%
1999-11-29	13,411.5	9,564.1	6,627.9	551.7	+6.1%	49.4%	69.3%	+1.0%
2004-03-21	13,802.5	9,756.1	7,209.8	192.0	+2.0%	52.2%	73.9%	+4.6%
2008-03-08	15,283.3	10,922.1	8,300.8	1,166.0	+12.0%	54.3%	76.0%	+2.1%
2013-05-05	17,883.7	13,268.0	11,256.6	2,345.9	+21.5%	62.9%	84.8%	+8.8%

Note:

VEP: adult citizens that are registered and eligible to vote; VAP: adult citizens that are within the prescribed age (21 years old and above) to vote regardless of being registered with the EC

¹ <http://www.idea.int/data-tools/question-countries-view/441/221/ctr>



Figure 2(A). CART Rules Representation Wave 3 dataset

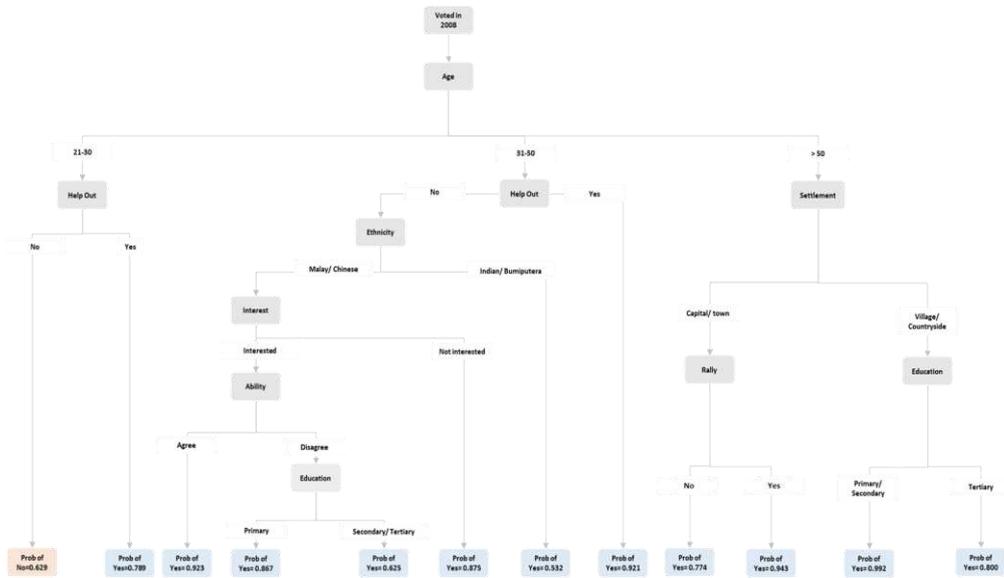


Figure 2(B). CART Rules Representation Wave 3 dataset

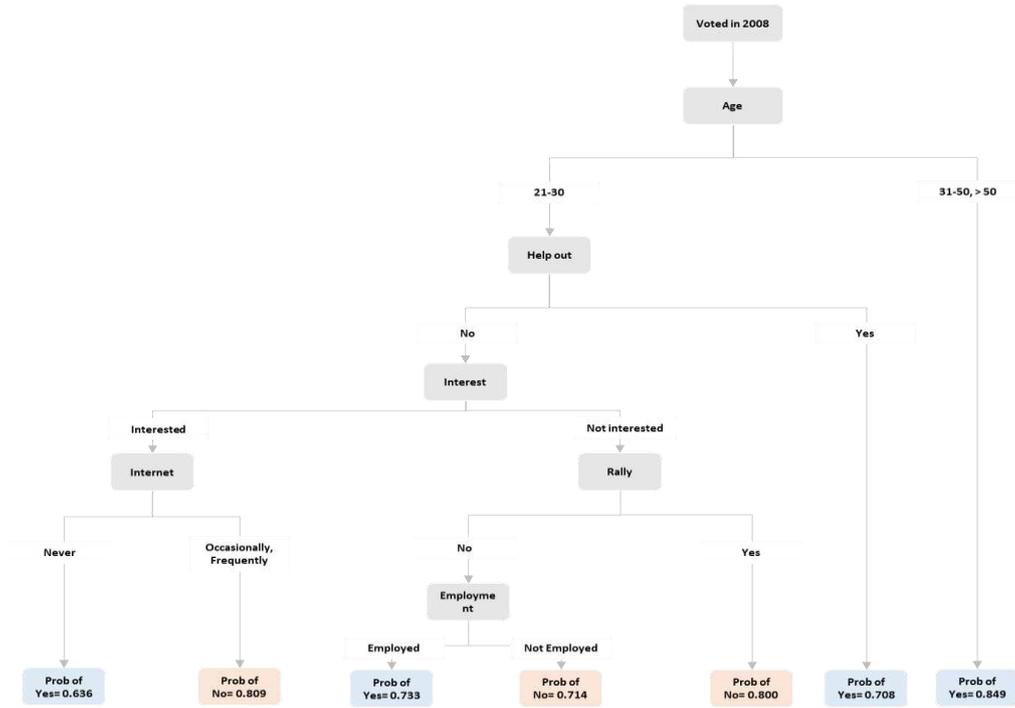


Figure 2(C). C5.0 rules representation Wave 3 dataset

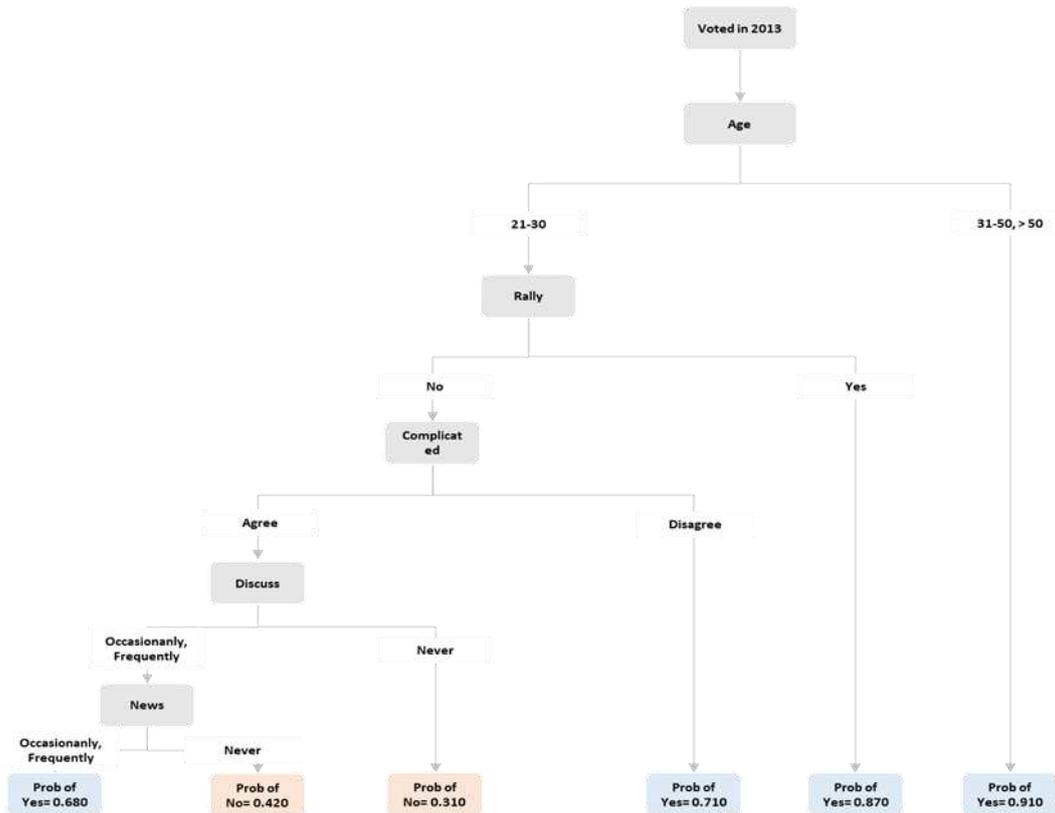


Figure 2(D). CART rules representation Wave 4 dataset

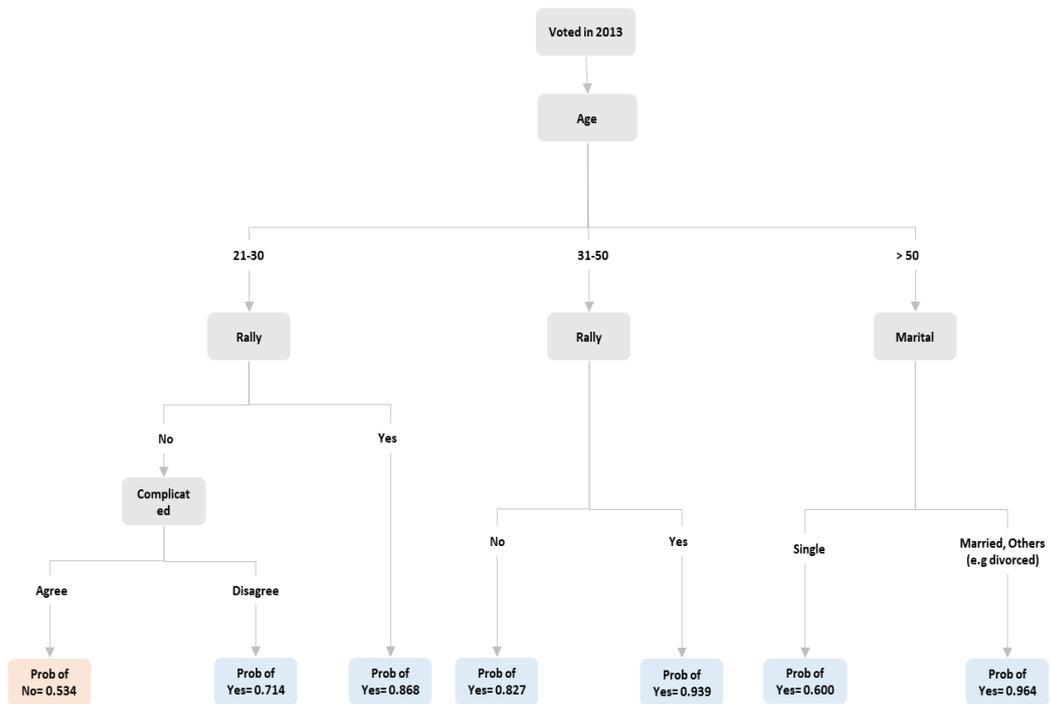


Figure 2(E). CHAID rules representation Wave 4 dataset

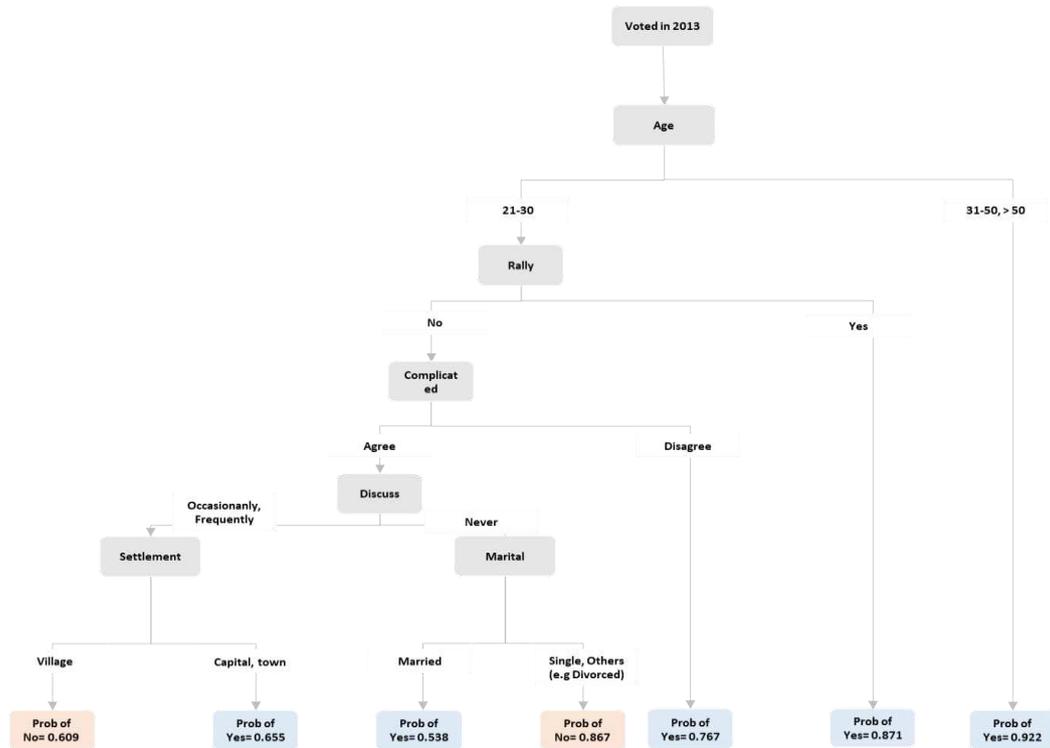


Figure 2(F). C5.0 rules representation Wave 4 dataset

Figures

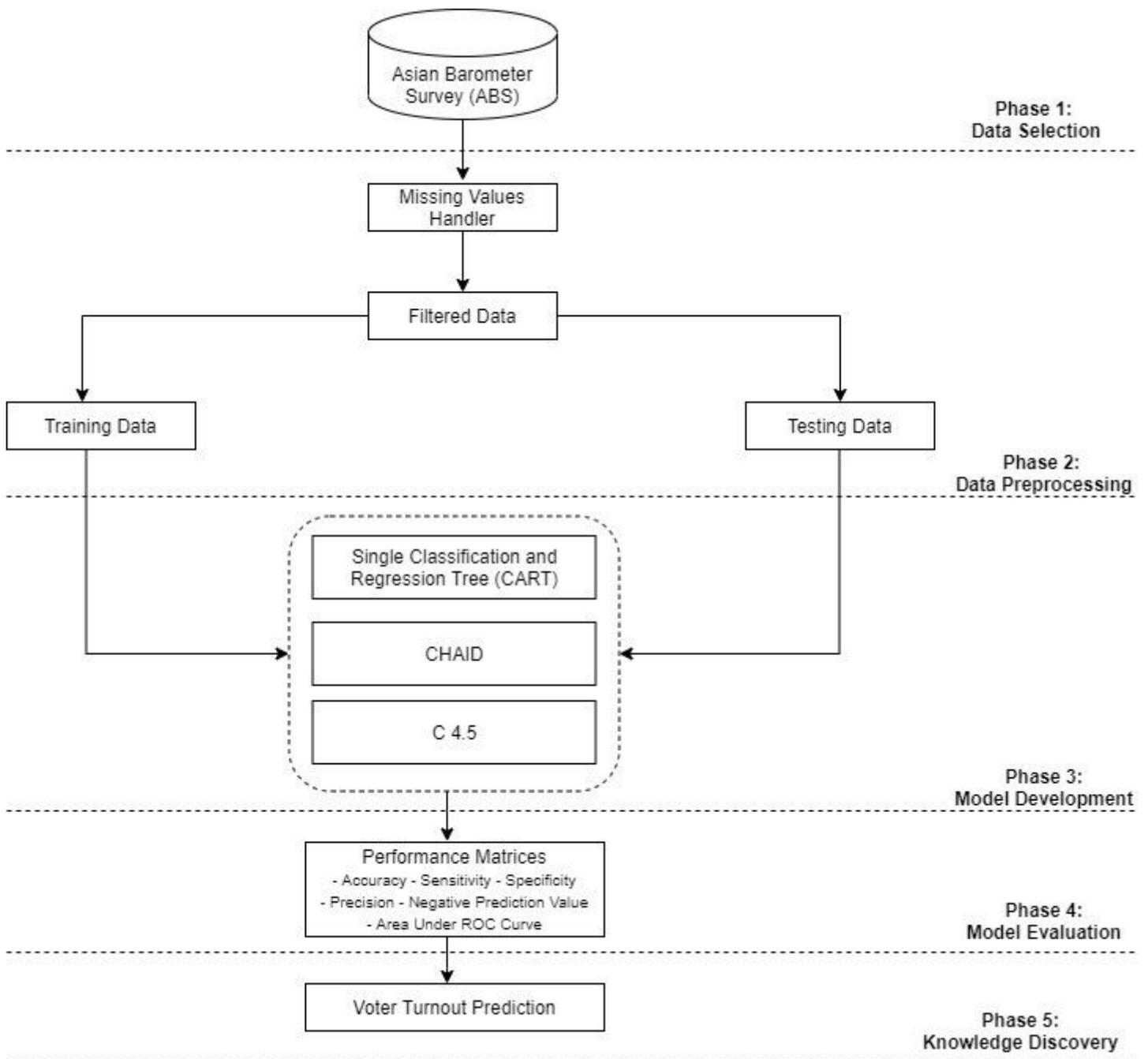


Figure 1

Experiment Setup

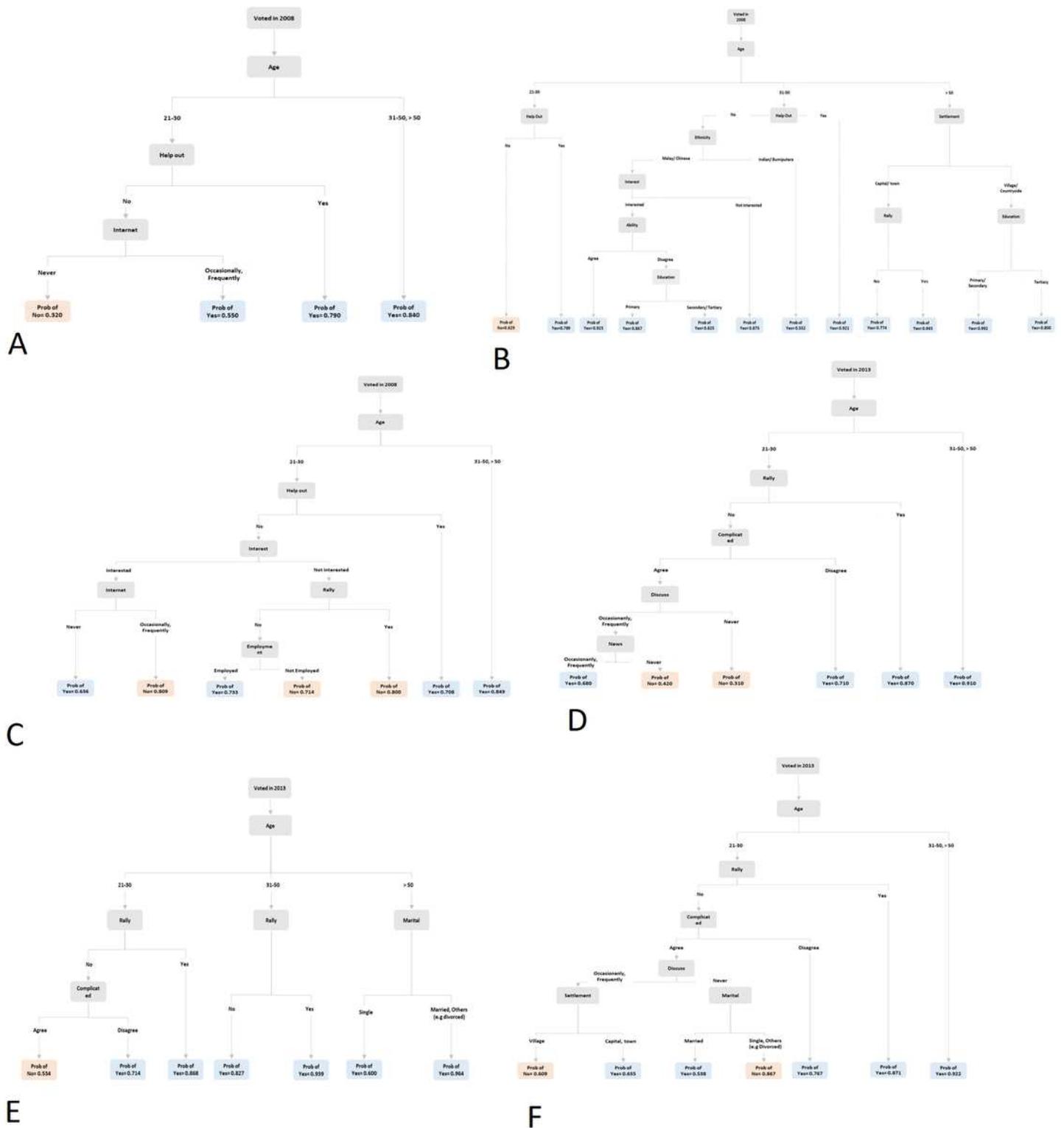


Figure 2

(A). CART Rules Representation Wave 3 dataset (B). CART Rules Representation Wave 3 dataset (C). C5.0 rules representation Wave 3 dataset (D). CART rules representation Wave 4 dataset (E). CHAID rules representation Wave 4 dataset (F). C5.0 rules representation Wave 4 dataset