Journal of Ambient Intelligence and Humanized Computing
https://doi.org/10.1007/s12652-021-03123-x

ORIGINAL RESEARCH

Discovering abstentionist profiles in 2015 Catalan elections

Eva Armengol[1] · **Àngel García-Cerdaña**[1]

Received: 17 February 2020 / Accepted: 5 March 2021

Abstract

One of the main democratic act is to elect representers for a country Government. Voters have to choose the representers of political parties that best fit their interests. However, it is also possible to abstain, i.e., to give up choosing the representers. Lastly, the percentage of abstention has increased in many democracies and, consequently, many studies have been carried out in order to discover the causes of the abstention. Causes such as alienation and indifference for politics have been found and characterized, but it is not clear *who* is abstaining. Some studies in specific countries have related aspects such age, education or wealth with voting. Nevertheless, all these studies are based on questionnaires that could not be completely reliable. In the present paper we use *decision trees*, an artificial intelligence technique, to extract patterns that characterize the profile of abstentionists. Differently from other studies, we used a data base elaborated from public data bases with both socio-economic and electoral data. In particular, we focused our study in Catalonia and the electoral results to the Catalan Parliament held in 2015. We also propose the use of decision trees to analyze the database and to extract patterns to characterize different profiles of abstentionist people.

Keywords Catalan Parliament · Electoral analysis · Abstention · Decision trees

1 Introduction

Democracies are based on political parties and election systems allowing voters to put the confidence in representers of these political parties to defend their interests. Representers belong to political parties that have a political program and voters choose the one that best fits their social, economic, and ethic values. Parties are interested in knowing who are their voters and try to adapt their electoral purpose to enlarge the target of voters. However, political parties seems not to be worried by abstention, why voters do not use their right to elect representers? Lastly, abstention has increased in most of the western democracies and many analysis have been performed to determine the causes of this increment (see for instance Adams et al. 2006; Blais and Kostelka 2015; Ferreira and Dionisio 2008).

The causes of the abstention are multiples and for different reasons attending to particular context of the constituency at hand. Blais and Kostelka (2015) have analyzed abstention in the European elections held in 2014. Among other more general conclusions, they state that the procedure of vote can also favours the abstention (see also Blais et al. 2014). Also, voters have to face different kinds of elections (for instance, general elections, municipalities, European Parliament, etc) and seems clear that they act differently in each one of them. Electors are more willing to vote in elections that consider closest to them, (for instance, municipalities), and the abstention is higher when the voter perceives that the elections are far away of its interests (see for instance Gill 2005; Kirchgässner 2003).

Ferreira and Dionisio (2008) associates the increment of the abstention to political disaffection, and no representativeness of political parties. In fact it is increasing the percentage of population that does not identify theirselves with a particular party (see Holland and Miskin 2002). There are several reasons for this lack of representativeness, for instance, lack of confidence in democracy in general, corruption, etc. (as it is explained in Castela and Villardón 2011; Cazorla-Martín et al. 2017; Nwankwo et al. 2017). It seems clear that by means of adequate

✉ Eva Armengol
eva@iiia.csic.es

Àngel García-Cerdaña
agarciacerdanya@gmail.com

[1] CSIC - Spanish Council for Scientific Research, IIIA - Artificial Intelligence Research Institute, Campus UAB, 08193 Bellaterra, Catalonia, Spain

Published online: 20 March 2021

questionnaires is possible to identify two kind of abstention as is mentioned by Adams et al. (2006) and Ferreira and Dionisio (2008):

- Alienation when parties or candidates are too distant from the voter.
- Indifference when parties or candidates are too similar between them.

Blais et al. (2014) performed an experimental analysis to examine the reasons that voters can have to abstain and to determine alienation and indifference. In this experimentation, the authors simulate a controlled scenario and quantify the benefits that abstention could produce to each voter of the experiment.

Castela and Villardón (2011) use the *ecological inference* method proposed by King (1997) as a basis to use the HJ-Biplot method (see Villardón 1986) to determine groups of population and their electoral behaviour from data of Portuguese elections held in 2002 and 2005. This work is interesting to determine the evolution of votes. In our case we only handle data of one electoral cite: the one of 2015 to Catalan Parliament.

A different approach is the one of Nwankwo et al. (2017) that, from a questionnaire including socio-demographic questions, use *Principal Component Analysis* (PCA) to determine the main aspects that influence the abstention in South Eastern Nigeria. In particular, they obtained eight main components: socioeconomic status, lack of trust in the electoral process, social trust and unemployment, registration and demographic factor, corruption and inadequate security, deception and intimidation, indigene status and electoral manipulation and poverty.

The most of abstention analysis try to explain why voters decide not to vote, however few of them try to characterize *who* is abstaining. Nwankwo et al. (2017) have show that, in Nigeria, people with high education, income and occupational status spur their participation in political activities (especially in voting). Rosema (2007) claims that when people is good represented at the polls they have a tendency to vote in order to have political voice in government. Blais and Kostelka (2015) have also found that people with high education tends to vote more than people less educated. Particularly, in this study the authors also conclude that the abstention in European elections is due to the absence of an European identity among citizens, reinforcing in that way the idea that this kind of elections are far of the voter's day-to-day life. Daoust and Blais (2017) have reported an interesting analysis to detect which population groups are more interested in the local elections than in the national ones. They have found that people with stronger attachment to their municipality are more interested on participating in local elections.

These studies are based on questionnaires before and during elections. In addition to directly ask for the vote, sometimes the voter is also questioned about aspects of his life in order to take into account some sociologic aspects such as age, gender, studies, economic situation, etc. Statistical methods give valuable information such as correlations between these variables. However, correlations do not provide a clear characterization of the people that is abstaining in a particular election. Results from questionnaires take globally all the data but cannot focus on profiles of potential voters. In our opinion, in order to provide a more detailed exam of the data it is necessary to fragment them in significant groups that follow a pattern. This kind of pattern is the one that can be obtained using *decision trees*. The advantage of decision trees is that their results are easily understood by experts and can be seen as explanations or descriptions of classes of objects. Decision trees have proved to be an useful tool to extract patterns that explain in a clear way relations that can also be found by means of an statistical analysis. The main difference is that the statistical analysis shows a general trend whereas decision trees give a better relation between attributes taking into account its values. In other words, a regression model shows that two variables A and B explain a percentage p of the variable C whereas a pattern extracted from a decision tree shows which of the values of A and B explain better the variable C.

In Armengol and Vicente (2019) we proposed the use of decision trees to analyze the abstention of the elections to the Catalan Parliament held in 2015. The results of such analysis were that the percentage of abstention in Catalonia directly depends on the percentage of votes to the socialist party: the more percentage of votes to the socialist party, the high percentage of abstention. Also, we reported the statistical analysis showing a correlation of $r = 0.580$ between the abstention and the percentage of votes to the socialist party. Using the patterns extracted from a decision tree, we have shown that the correlation is between 75 and 82% in extreme values (*very high* or *very low*) of both variables, whereas such correlation is not so clear (around 50%) in intermediate values (*high* or *low*). Therefore, decision trees give us more information about the relation between variables. This result was a first surprise since we expected the abstention was primarily related to some sociologic aspect of the population (i.e., immigration, unemployment, income, etc.).

In the present paper we propose to discover the abstentionist profile of the Catalan population, based on an analysis of the population of each electoral section. The decision trees is the method we use to obtain these profiles since in our previous work they have been proved to be useful and understandable.

In the next section there is a brief explanation of decision trees and then, in Sect. 3, we describe the data base used in the experiments.

2 Decision trees

A *Decision tree* (DT) is a directed acyclic graph in the form of a tree. The root of the tree has not incoming edges and the remaining ones have exactly one incoming edge. Nodes without outgoing edges are called *leaf* nodes and the others are *internal* nodes. A DT is a classifier expressed as a recursive partition of the set of known examples of a domain (see Maimon and Rokach 2010). The goal is to create a domain model predictive enough to classify future unseen domain objects.

Each node of a tree has associated a set of examples that are those satisfying the path from the root to that node. The leaves determine a partition of the original set of examples, since each domain object only can be classified following one of the paths of the tree (see the example below). The construction of a decision tree is performed by splitting the source set of examples into subsets based on an attribute-value test. This process is repeated on each derived subset in a recursive manner. Figure 1 shows the ID3 algorithm (see Quinlan 1979, 1986) commonly used to grow decision trees. From a decision tree we can extract rules (i.e., patterns) giving descriptions of classes, since each path from the root to a leaf form a description of a class. When all the examples of a leaf belong to the same class such description is *discriminant*. Otherwise, the description is *no discriminant*.

A key issue of the construction of decision trees is the selection of *the most relevant attribute* to split a node. Each measure uses a different criteria, therefore the selected attribute could be different depending on it and thus the whole tree could also be different. In our experiments we use the López de Mántaras' distance (1991), which is an entropy-based normalized metric defined in the set of partitions of a finite set. It compares the partition induced by an attribute, say a_i, with the *correct partition*, i.e., the partition that classifies correctly all the known examples. The best attribute is the one inducing the partition which is closest to the correct partition. Given a finite set X and a partition $\mathcal{P} = \{P_1, \dots, P_n\}$ of X in n sets, the entropy of $\mathcal{P}$ is defined as ($|\cdot|$ is the cardinality function):

$$H(\mathcal{P}) = -\sum_{i=1}^{n} p_i \cdot \log_2 p_i, \text{ where } p_i = \frac{|P_i|}{|X|}$$

and where the function $x \cdot \log_2 x$ is defined to be 0 when $x = 0$. The *López de Mántaras*' distance (LM) between two partitions $\mathcal{P} = \{P_1, \dots, P_n\}$ and $\mathcal{Q} = \{Q_1, \dots, Q_m\}$ is defined as:

$$\mathrm{LM}(\mathcal{P}, \mathcal{Q}) = \frac{H(\mathcal{P}|\mathcal{Q}) + H(\mathcal{Q}|\mathcal{P})}{H(\mathcal{P} \cap \mathcal{Q})}, \tag{1}$$

where

$$H(\mathcal{P}|\mathcal{Q}) = -\sum_{i=1}^{n}\sum_{j=1}^{m} r_{ij} \cdot \log_2 \frac{r_{ij}}{q_j},$$
$$H(\mathcal{Q}|\mathcal{P}) = -\sum_{j=1}^{m}\sum_{i=1}^{n} r_{ij} \cdot \log_2 \frac{r_{ij}}{p_i},$$
$$H(\mathcal{P} \cap \mathcal{Q}) = -\sum_{i=1}^{n}\sum_{j=1}^{m} r_{ij} \cdot \log_2 r_{ij},$$
$$\text{with } q_j = \frac{|Q_j|}{|X|}, \text{ and } r_{ij} = \frac{|P_i \cap Q_j|}{|X|}.$$

Decision trees can be useful for our purpose because their paths give us *patterns* describing classes of objects (electoral sections in our approach) in a user-friendly manner. A drawback of decision trees is *overfitting*, meaning that there are few objects in most of the leaves of the tree. In other words, paths are actually descriptions that poorly represent the domain. The responsible of overfitting is the stopping condition of the algorithm: the set of examples has to be partitioned until all the examples of a node belong to the same class.

A way to either avoid or reduce overfitting is by pruning the tree i.e., to expand all the nodes and then, with a post-process to merge two or more nodes; or, under some

Fig. 1 ID3 algorithm for growing a decision tree

```
ID3 (examples, attributes)
   create a node
   if all examples belong to the same class  return class as the label for the node
   otherwise
        A ← best attribute
        for each possible value v_i of A
                add a new tree branch below node
                examples_vi ← subset of examples such that A = v_i
                ID3(examples_vi, attributes - {A})
  return node
```

conditions, a node is no longer expanded. However, in both cases, this means that leaves can contain objects belonging to several classes and, therefore, paths do not represent discriminatory descriptions of classes. In other words, the descriptions or patterns represented by the branches (paths) of the tree are satisfied by objects of more than one class.

In our approach, we managed overfitting by controlling the percentage of elements of each class. Let S_N be the set of objects associated with an internal node N, the stopping condition in expanding N (the *if* of the ID3 algorithm) holds when the percentage of objects in S_N that belong to the majority class decreases in one of the children nodes. In such a situation, the node N is considered as a leaf.

As example, let us suppose that the examples of a data base can be classified in one of the following classes: Very low, Low, High, and Very high. Let us suppose now that when we grow a decision tree, we find that the most relevant attribute is A1 and that for a given value $v1$ of such attribute, the examples of the data base are divided by classes in the percentages show in the left side of Fig. 2. This means that the 43.75% of examples that have the value $v1$ in the attribute A1 belong to the class Very low. Because the goal is, if possible, to achieve groups of examples belonging to the same class, the next step is to select another attribute to specialize the description $[\mathsf{A1} = v1]$. Let us suppose that S_1 is the set of examples satisfying the description above and that now the most relevant attribute is A2 and that for a given value $v2$ of such attribute, the examples of the data base are divided by classes in the percentages show in the center of Fig. 2. This means that the 46.51% of examples that have the value $v1$ in the attribute A1 belong to the class Very low. Notice that this percentage is higher than the one obtained with only A1. Therefore, now we have the tree path (description) $[[\mathsf{A1} = v1], [\mathsf{A2} = v2]]$ that could be specialized by adding to it another attribute. Let us suppose that S_{12} is the set of examples satisfying the description above that now the most relevant attribute is A3 and that for a given value $v3$ of such attribute, the examples of the data base are divided by classes in the percentages show in the right side of Fig. 2. This means that the 38.09% of examples that have the value $v3$ in the attribute A3 belong to the class Very low. Notice that this percentage has decreased with respect to the previous one. Therefore, now the procedure stops and the tree path is $[\mathsf{A1} = v1], [\mathsf{A2} = v2]]$, i.e., A3 is not included.

Commonly, the procedure to classify an unseen object is to go through the tree according to the value that the object has in each attribute included in the paths. The object is classified when a leaf is reached. In our case, we do not want to classify but only want to analyze the patterns that form the tree paths. Thus, if an object satisfies the central node in Fig. 2 then it means that it has the value *v1* in the attribute A1 and the value *v2* in the attribute A2.

3 The data base

Previously to describe the data base of our experiments we briefly explain the administrative organization of Catalonia and the political context in 2015.

3.1 Catalan organization and political context

Catalonia is located in the N.E. of Spain near the French frontier (Fig. 3). In Catalonia, there are four different kind of elections: Municipalities, Catalan Parliament, Spanish Parliament and European Parliament. Our goal is to characterize electoral sections according to their abstention level focusing on the elections to the Catalan Parliament held in September 27, 2015, and extract patterns giving some kind of explanation of the result. In the future we plan to extend this analysis to other elections.

From the administrative point of view, Catalonia is composed of four concurrencies: Barcelona, Girona, Tarragona and Lleida (Fig. 3). In Armengol and Vicente (2019), we also reported the results of an experiment where we forced to build a decision tree using only socio-economic attributes. In that experiment we found that abstention is related with immigration: the more percentage of immigration, the high percentage of abstention. In the present paper, we report experiments focused on discover more accurate patterns explaining abstention in Catalonia. In particular, we have analyzed separately each Catalan constituency and compared whether or not all of them exhibit the same behavior with respect to the abstention.

Electoral landscape of Catalonia is formed by 5048 *electoral sections* each one of them composed of a minimum of 500 potential voters and a maximum of 2000. Most of times, each electoral section corresponds to an *electoral table*

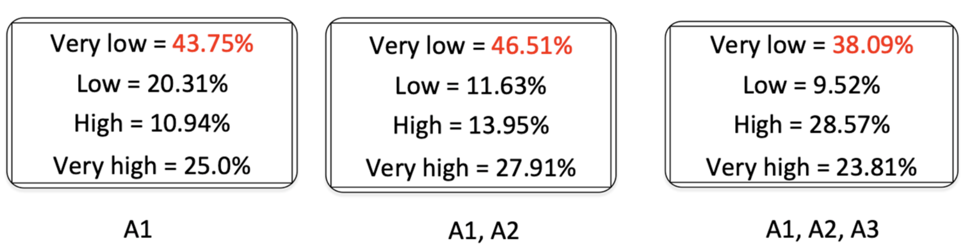

Fig. 2 Example of the stopping condition we used when growing a decision tree

Fig. 3 Left: location of Catalonia. Right: concurrencies composing Catalonia

although this can variate if the number of voters of a table is either too high or the population is scattered throughout the territory represented by the electoral section. In such cases, an electoral section is divided in several electoral tables with the constraint that no electoral table can have less than 200 voters. Considered by concurrencies, Barcelona has 3588 electoral sections, Girona 531, Tarragona 542, and Lleida 387. Notice that the results of Barcelona have a great weight when the electoral results are considered globally.

In 2015 there was a complex political framework in Catalonia. In addition to the traditional ideologies left–right a new issue appear: the independence of Catalonia from Spain. Some of the historical Catalan parties already had the independence in their program, however it was not the main objective. Nevertheless, from a set of reasons that are out of the scope of this paper, the independence of Catalonia has become a priority for many population and for some parties, to the point that the choice independence/ no independence has put the left-right dichotomy in a second term. It will be interesting to show if this factor has influenced the profile of people who abstain. Particularly, a first effect of this new issue is the increment of turnout with respect to previous elections. Figure 4 shows the percentage of abstention and votes to each party that had concurred to elections held in 2015. Results are given globally and also for each concurrency. Notice that abstention is around the 20% in all the concurrencies. This percentage is actually low compared with the one produced in previous Catalan elections: around 30% in 2012 and around 40% in 2010 (source: IDESCAT).

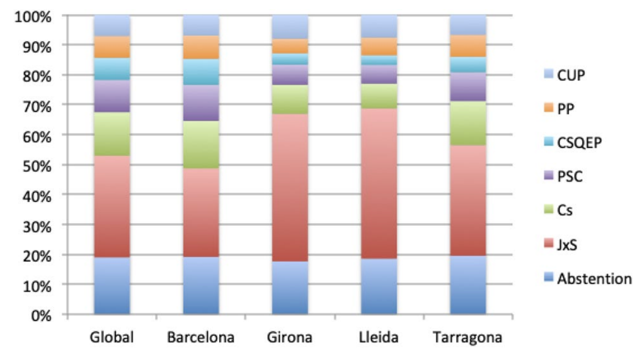

Fig. 4 Percentage of both abstention and votes to each one of the parties that concurred to the Catalan elections in 2015. This graphic shows the global results and also the results for each one of the Catalan concurrencies. (source: IDESCAT www.idescat.net)

3.2 The data

The data base we have is composed of 5048 records, each one of them corresponds to an electoral section in Catalonia. Each record has two kinds of information: electoral data and socio-demographic data. The file containing the electoral results for all the Catalan electoral sections is public and contains (in absolute numbers and in percentage) the votes to each political party, the null votes, the blank votes, abstention and how many potential voters has an electoral section (*electoral census*). This file does not contain socio-demographic information associated to the electoral sections.

Socio-demographic data in our data base has been formed from an aggregation of the data coming from several public data bases that have different granularity, therefore they are not easy to merge. The AIS Group company has used ecological inference processes to put all the data base in the same granularity (electoral sections). The result of such aggregation is a set of typologies of families (shown in Table 1) contained in the data base Habits© from the AIS Group. In that way, we can assess how many families of each typology are in each electoral section. In AIS AG (2019) there is a description of the method used to assess how many families of each typology are in each electoral section. The resulting file containing the socio-demographic information related to each electoral section is the one we used in our experiments.

All the attributes are numerical and we have discretized them. Such discretization has been done by computing the quartiles of each one and assessing a label to each attribute of the resulting intervals (*VL, L, H*, and *VH*).

Table 1 Typology of families considered in our experiments

Typology	Description
AF	Families with adolescents (until 18 years old) and children
BG	Families with young sons (from 18 to 35 years old)
CH	Families with children
DK	Families where the main salary is a pension
EI	Families with two working persons and without children (DINK)
J	Families with one or two members, no children, only one salary
LP	Families with one or two members, no children, one or two pensions
M	Singles older than 35 with a salary
N	The main person is a student, a housework or a permanent disability
O	The main person of the family is unemployed
Q	Family formed by only one person receiving a pension
R	The main person of the family is an immigrant
Expenses	Average of the expenses of a family
Income	Average of the income in a family

They have obtained as it is explained in https://www.ais-int.com/marketing-y-ventas/geomarketing-habits-big-data/

4 Experiments

The experiments reported in Armengol and Vicente (2019) shown that the percentage of abstention was highly correlated to the percentage of votes to the socialist party (PSC), i.e., the more votes to the socialist party, the more percentage of abstention. It is difficult to interpret these results in sociological terms and taking into account only one electoral call. However, it seems to indicate that in electoral sections with low percentage of vote to the socialist party was low abstention because voters of other parties were highly motivated to vote. Conversely, electoral sections with high percentage of votes to the socialist party has high percentage of abstention, meaning that voters of other parties were not motivated to vote.

In the present paper we have grown decision trees using only the socio-economic attributes shown in Table 1, and the percentage of abstention as solution class. Our goal is to extract patterns characterizing the electoral sections according to their percentage of abstention. In order to avoid overfitting the algorithm cuts the node expansion when the percentage of the majority class decreases if the node was expanded (see Sect. 2 where the decision trees are explained). Then, we have manually analyzed the tree paths (patterns) corresponding to leaves that contain around the 25% of the population. Most of significant patterns are not discriminant, i.e., they are satisfied by electoral sections having different abstention levels.

4.1 Results

In one of the experiments reported in Armengol and Vicente (2019) we used only socio-economic attributes. In that experiment we found that abstention is related with immigration: the more percentage of immigration, the high percentage of abstention. In the current paper, we have also found that the immigration (Typology R) is the most relevant aspect to explain the abstention. Table 2 shows the patterns obtained taking all the data, i.e., without separating them by concurrencies. These patterns have to be read as follows:

- the 69.73% of electoral sections having very low immigration have a low or very low percentage of abstention,
- the 61.37% of electoral sections having low immigration have a low or very low percentage of abstention,
- the 57.0% of electoral sections having high immigration have a high or very high percentage of abstention,
- the 74.43% of electoral sections having very high immigration have a high or very high percentage of abstention.

It is interesting to remark that for an extreme number of immigrant families (*very low* or *very high*) the patterns are clear, whereas the others, although they show a clear trend, are not accurate enough to explain the abstention. A regression model shows that immigration only explains the 16.11% of the abstention, and the correlation between them is $r = 0.431$. These values are due, as the patterns show, to the intermediate values of immigration. We could extract patterns having more attributes, although in that case they are not satisfied by a significant enough number of electoral sections.

For instance, see the patterns shown in Table 3. Notice that both patterns are good discriminant since the first of them can classify a section as having *low* or *very low*

Table 2 Patterns taking into account only the Typology R with all its values

Pattern	#ES	Abstention					
		VL	*L*	*VL + L*	*H*	*VH*	*H+VH*
[TypologyR, *VL*]	1262	44.14	25.59	**69.73**	18.86	11.41	30.27
[TypologyR, *L*]	1266	31.83	29.54	**61.37**	23.14	15.48	38.62
[TypologyR, *H*]	1280	15.86	27.11	42.97	30.15	26.87	**57.02**
[TypologyR, *VH*]	1240	8.06	17.5	25.56	27.82	46.61	**74.43**

For each one it is shown the number of electoral sections that satisfy the pattern (#ES), the percentage of sections belonging to each one of the classes (*VL, L, H* and *VH*) and also the aggregation of abstention in two levels very low plus low (*VL* + *L*) on one hand and high and very high (*H* + *VH*) on the other hand

Bold values indicate the highest percentage of satisfaction of the pattern

Table 3 Patterns more specific than the ones shown in Table 2

Pattern	#ES	Abstention	
		VL + L	*H+VH*
[[TypologyR, *L*], [TypologyAF, *VL*]]	546	**81.5**	18.49
[[TypologyR, *H*], [TypologyCH, *L*]]	333	16.51	**83.48**

Bold values indicate the highest percentage of satisfaction of the pattern

abstention with 81.5% of accuracy whereas the second one classifies a section as having *high* or *very high* abstention with 83.48% of accuracy. However, both patterns have very low support since they are satisfied by the 10.82% and 6.60% respectively, of the whole electoral sections. A different point of view in favour of these patterns is that although the first one has a global support of 10.82%, it represents around the 43% of electoral sections having low immigration. Similarly, the pattern with a support of 6.60% represents around of the 26% of electoral sections having a high immigration.

It is also interesting to note that when an electoral section has *very low* immigration (i.e., [TypologyR, *VL*]) the next more relevant attribute is Typology AF (corresponding to families with no children) with value very low; whereas when the immigration is high, the next more relevant attribute is Typology CH (families with children) with value low. In other words, of all electoral sections with *low* immigration and *very low* number of families with no children, around the 81% of them have a *very low* or *low* percentage of abstention, i.e., they are electoral sections with high turnout. Also, of all electoral sections with *high* immigration and *low* number of families with children, around 83% of them have a *very high* or *high* percentage of abstention, i.e., they are electoral sections with low turnout. From a statistic analysis also have found that both Typology R and Typology AF explain the 21% of abstention.

In the next sections the abstention of each concurrency is analyzed.

4.2 Abstention in Barcelona

Barcelona has 3588 electoral sections. By growing a decision tree, we find that the relevant attribute is the Income average of the electoral section. Table 4 shows the patterns we obtained. These patterns have to be read as follows:

- The 77.32% of electoral sections having a very low income in average, have a high or very high percentage of abstention,
- the 68.59% of electoral sections having a low income in average, have a high or very high percentage of abstention,

Table 4 Patterns for explain the abstention in Barcelona

Pattern	#ES	Abstention					
		VL	*L*	*VL + L*	*H*	*VH*	*H+VH*
[Income, *VL*]	732	7.51	15.16	22.67	28.41	48.91	**77.32**
[Income, *L*]	871	11.14	22.27	33.41	32.72	35.87	**68.59**
[Income, *H*]	972	21.40	33.64	**55.04**	27.98	16.97	44.95
[Income, *VH*]	1013	46.20	29.81	**76.01**	16.98	7.01	23.99

Bold values indicate the highest percentage of satisfaction of the pattern

- the 55.04% of electoral sections having a high income in average, have a low or very low percentage of abstention,
- the 76.01% of electoral sections having a very high income in average, have a low or very low percentage of abstention.

These patterns can be interpreted as follows: the highest the income average is, the lower is the percentage of abstention. The worst pattern is the one associated to electoral sections with high income ($[\mathsf{Income}, H]$) where little more than 50% of the electoral sections have low or very low abstention. If we want to find specializations of that pattern we found that all of them have an accuracy lower than 60%, therefore it is not possible to find more accurate patterns to describe these electoral sections. The pattern $[\mathsf{Income}, L]$, has two specializations show in Table 5.

The pattern $\big[[\mathsf{Income}, L], [\mathsf{TypologyCH}, L]\big]$ (where Typology CH corresponds to families with children) is satisfied by 321 electoral sections representing almost the 9% of the total in Barcelona; however, it represents the 36.85% of sections having *low* income. According to that pattern, around the 70% of electoral sections satisfying it have *high* or *very high* abstention.

Concerning the pattern $\big[[\mathsf{Income}, L], [\mathsf{TypologyCH}, VL]\big]$, it is satisfied by 212 electoral sections, representing the 0.58% of the electoral sections of Barcelona and the 24.34% of those having low income. According to that pattern, around the 81% of electoral sections satisfying it have *high* or *very high* abstention.

4.3 Abstention in Girona

Girona has 531 electoral sections. By growing a decision tree, we find that the relevant attribute is the immigration (Typology R). Table 6 shows the patterns we obtained. These patterns have to be read as follows:

- The 92.12% of electoral sections having very low immigration have a low or very low percentage of abstention,
- the 81.09% of electoral sections having low immigration have a low or very low percentage of abstention,
- the 62.85% of electoral sections having high immigration have a high or very high percentage of abstention
- the 74.58% of electoral sections having very high immigration have a high or very high percentage of abstention.

There is a clear trend showing that the percentage of abstention in Girona increases with the immigration. The patterns are clearly discriminant since they correctly classify more than 70% of the electoral sections. The only exception, although with a percentage upper than the 62% is the one associated with electoral sections having high immigration. Notice that abstention in Girona tends to be always *low* or *very low*, and only when the immigration is *very high* the percentage of abstention is clearly *high* or *very high*.

4.4 Abstention in Tarragona

Tarragona has 542 electoral sections. By growing a decision tree, we find that the relevant attribute is the immigration (Typology R). Table 7 shows the patterns we obtained. These patterns have to be read as follows:

Table 5 Patterns more specific than the ones show in Table 2

Pattern	#ES	Abstention	
		VL + L	*H+VH*
$\big[[\mathsf{Income}, L], [\mathsf{TypologyCH}, L]\big]$	321	29.28	70.71
$\big[[\mathsf{Income}, L], [\mathsf{TypologyCH}, VL]\big]$	212	18.86	81.13

Table 6 Patterns for explain the abstention in Girona

Pattern	#ES	Abstention					
		VL	*L*	*VL + L*	*H*	*VH*	*H+VH*
[TypologyR, VL]	127	77.95	14.17	**92.12**	7.09	0.79	7.88
[TypologyR, L]	105	58.09	23.81	**81.09**	10.48	7.62	18.10
[TypologyR, H]	105	38.09	24.76	**62.85**	20.00	17.14	37.14
[TypologyR, VH]	194	9.79	14.95	24.74	26.29	48.29	**74.58**

Bold values indicate the highest percentage of satisfaction of the pattern

Table 7 Patterns for explain the abstention in Tarragona

Pattern	#ES	Abstention					
		VL	*L*	*VL + L*	*H*	*VH*	*H+VH*
[TypologyR, VL]	93	59.14	20.43	**79.57**	13.98	6.45	20.43
[TypologyR, L]	115	27.83	29.56	**57.59**	24.35	18.26	42.6
[TypologyR, H]	180	10.00	25.00	35.00	34.44	30.55	**64.99**
[TypologyR, VH]	154	2.60	13.64	16.24	25.32	58.44	**83.76**

Bold values indicate the highest percentage of satisfaction of the pattern

– The 79.57% of electoral sections having very low immigration have a low or very low percentage of abstention,
– the 57.59% of electoral sections having low immigration have a low or very low percentage of abstention,
– the 64.99% of electoral sections having high immigration have a high or very high percentage of abstention,
– the 83.76% of electoral sections having very high immigration have a high or very high percentage of abstention.

As in Girona, there is a clear trend that shows that the percentage of abstention in Tarragona increases with the immigration. The patterns are discriminant enough except the one corresponding to electoral sections with low immigration. The most discriminant pattern is the one associated with the electoral sections having *very high* immigration percentage. As in all the previous cases, more specific patterns do not have enough support, i.e., they are not satisfied by a significant number of electoral sections.

4.5 Abstention in Lleida

Lleida has 387 electoral sections. By growing a decision tree, we find that the relevant attribute is the immigration (Typology R). Table 8 shows the patterns we obtained. These patterns have to be read as follows:

– The 78.33% of electoral sections having very low immigration have a low or very low percentage of abstention,
– the 63.33% of electoral sections having low immigration have a low or very low percentage of abstention,
– the 60.52% of electoral sections having high immigration have a high or very high percentage of abstention,
– the 70.41% of electoral sections having very high immigration have a high or very high percentage of abstention.

In Lleida, there is also a clear trend that shows that the percentage of abstention increases with the immigration. All the patterns have a classification accuracy of more than the 60%.

4.6 Discussion

The patterns extracted from decision trees show that the most important aspect that explains the abstention is the immigration of the electoral sections. This is true globally and for all the concurrencies except Barcelona. Let us analyze in detail the different levels of abstention for each level of immigration.

4.6.1 Electoral sections with *very low* immigration

Figure 5 shows the results for electoral sections with very low percentage of immigration with respect to the different percentages of abstention. The graphic in the left hand side shows the abstention globally and for each one of the Catalan concurrencies. Clearly, the majority of electoral sections are those having very low abstention. This majority is specially clear in Girona (around the 80%) and Tarragona (around 60%). This is true also in Barcelona, although with lower difference with respect to the

Table 8 Patterns for explain the abstention in Lleida

Pattern	#ES	Abstention					
		VL	*L*	*VL + L*	*H*	*VH*	*H+VH*
[TypologyR, VL]	153	49.67	28.76	**78.33**	15.69	5.88	21.57
[TypologyR, L]	60	25.00	38.33	**63.33**	26.67	10.00	36.67
[TypologyR, H]	76	14.47	25.00	39.47	30.26	30.26	**60.52**
[TypologyR, VH]	98	5.10	24.49	29.59	28.57	41.84	**70.41**

Bold values indicate the highest percentage of satisfaction of the pattern

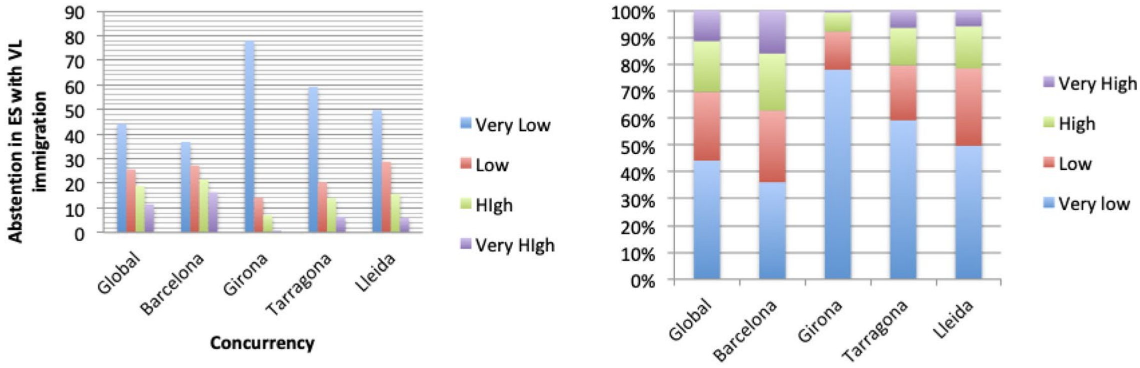

Fig. 5 Distribution of electoral sections with *Very Low* Immigration according to the percentage of abstention, globally and for each one of the Catalan concurrencies. Left hand side shows each percentage of abstention separately. Right hand side shows the contribution of each percentage of abstention

other abstention levels. The global difference between the percentages of abstention is not as high as should expect mainly due to the weight of Barcelona.

The graphic in the right hand side allows to see the contribution of each value of the abstention to the total for the electoral sections that have *very low* immigration. For instance, in Barcelona, the weight of electoral sections with *very low* abstention is around 35%, the weight for those with *low* abstention is around 25%, the weight of those with *high* abstention is around 20%, and the weight of those with *very high* abstention is around 15%. In Girona there is a high percentage of electoral sections with *very low* immigration having *very low* abstention (almost the 80%) and only around the 5% of electoral sections with *very low* immigration have *very high* percentage of abstention. In Lleida, there is a low percentage (around 10%) of electoral sections with *very high* abstention whereas the other levels are almost balanced.

Thus, the conclusion is that electoral sections with *very low* immigration have *very low* abstention in all the Catalan concurrencies.

4.6.2 Electoral sections with *low* immigration

Figure 6 shows the results for electoral sections with *low* immigration. Notice that, except in Girona, globally and for each one of the other concurrencies, the percentage of electoral sections with *very low* abstention has drastically decreased with respect the ones in Fig. 5. It is particularly interesting the case of Lleida where around the 50% of electoral sections with *very low* immigration have *very low* percentage of abstention (see Fig. 5) but when the percentage of immigration is *low*, then the percentage of electoral sections with *very low* abstention is reduced to around the 25% (see Fig. 6). It is also interesting to show that in Girona has increased the percentage of electoral sections with *very high* abstention to around the 5%.

Notice also that in Tarragona and in Lleida there is a high percentage of electoral sections (between 50 and 60%) having *very low* immigration with *very low* abstention (see Fig. 5), and that this percentage drastically decreases (around 25%) when there is a *low* immigration (see Fig. 6) being under the 30% in both cases.

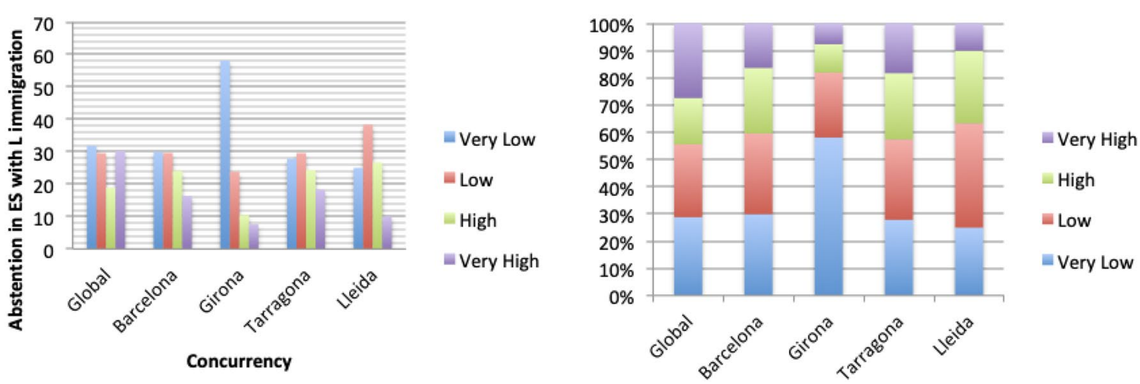

Fig. 6 Distribution of electoral sections with *Low* percentage of Immigration according to the percentage of abstention, globally and for each one of the Catalan concurrencies

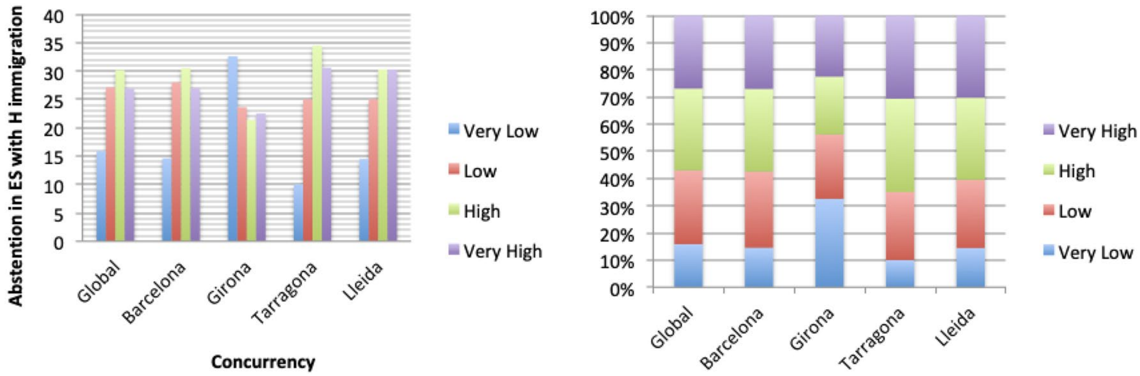

Fig. 7 Distribution of electoral sections with *High* Immigration according to the percentage of abstention, globally and for each one of the Catalan concurrencies

The conclusion here is that as the percentage of immigration increases, the number of electoral sections with high or very high abstention also increases

4.6.3 Electoral sections with *high* and *very high* immigration

Figure 7 shows the distribution of electoral sections with *high* immigration. Notice that globally and for all the concurrencies except in Girona, the percentage of electoral sections having *very low* abstention has decreased. In fact, now the majority class is *high* abstention. The only exception is Girona that, although the percentage has also decreased (now is between 30 and 40%) it is clearly the highest of all the concurrencies. The graphic in the right hand side clearly shows a balance between all the levels of abstention, being Girona the one having a different behaviour. However, notice that the percentage of electoral sections with *very high* abstention is similar in all the concurrencies (even in Girona).

Finally, Fig. 8 shows the distribution of electoral sections with *very high* immigration. From both graphics it is clear that the majority class are the electoral sections with *very high* abstention without exception. The percentage of sections with *very low* abstention has decreased below 10%.

As in for the previous sections, the conclusion is that the percentage of immigration is clearly correlated with the abstention. Now, we have seen that electoral sections with high or very high percentage of immigration mostly have high or very high abstention.

4.6.4 Final remarks

The set of all the graphs above clearly depicts our conclusion: the more immigration, the more percentage of abstention. This is true for all the Catalan concurrencies, although this trend seems to be not so clear in Barcelona where we have found that the most relevant attribute is the income.

In fact, Arowolo and Aluko (2013), and Nwankwo et al. (2017) have obtained similar results concerning to Nigeria, where although those who they call *strangers* (or

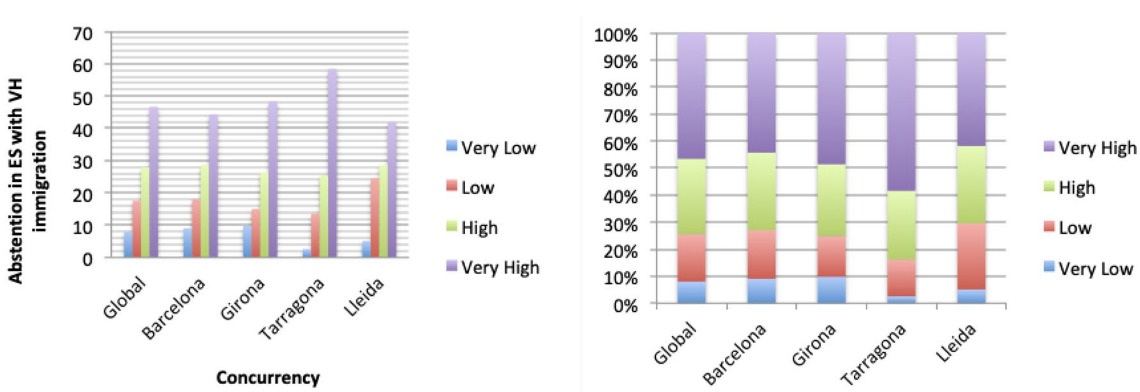

Fig. 8 Distribution of electoral sections with *Very High* Immigration according to the percentage of abstention, globally and for each one of the Catalan concurrencies

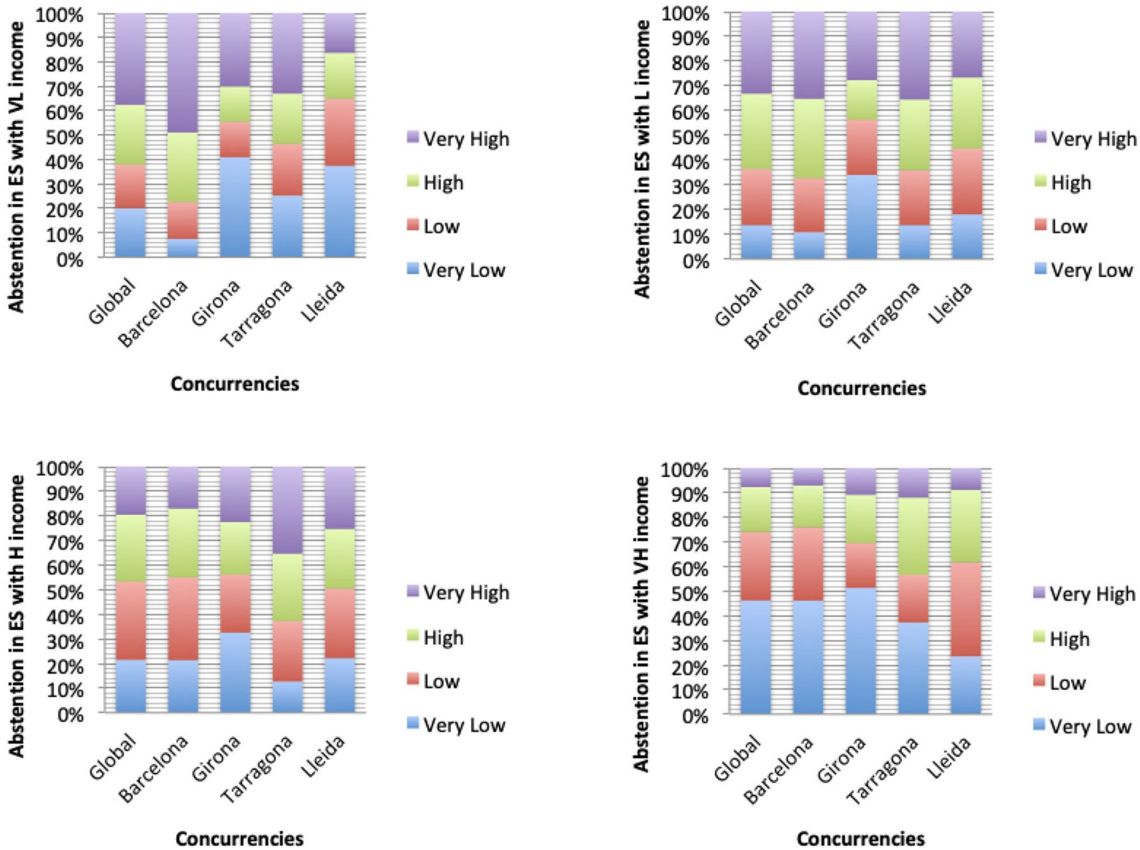

Fig. 9 Contribution that each level of Income to each percentage of abstention globally and for each one of the Catalan concurrencies

non-indigenous) are not restricted from voting, the social disconnection between them and the indigenous people may be the cause of low turnout. Also, Leighley and Vedlitz (1999) conducted studies to obtain models of political participation in U.S. based on socio-economic status, psychological resources, social connectedness, group identity or consciousness and group conflict. They find, among other aspects, that the longer one has lived in the same community, the greater the level of political participation. In fact, they perform an interesting analysis of the relation between race and political participation.

As we already mentioned, although Barcelona also follows this general trend, we have found that the most relevant feature for the abstention is the familiar income average. For this reason we have also constructed the graphs relating income and abstention for all the concurrencies. Figure 9 left shows the graphs with the contributions of each percentage of abstention to each one of the levels of income average. Notice that only Barcelona has a clear trend relating the increment of abstention with the increment of the income average. For instance, around the 77% of electoral sections with *very low* income average have *very high* or *high* abstention, and around the 76% of electoral sections with *very high* income have *low* or *very low* abstention. In Tarragona around the 25% of electoral sections with *very low* income average have *very low* percentage of abstention.

These results are in concordance with some studies performed in other countries. For instance, Solt (2008) conducted analysis trying to relate economic position with political engagement. One interesting result is that income inequality has a strong negative effect on the political interest, political discussion and electoral participation of those in the median quintiles or below. Also in Nigeria, Solt has found that the concentration of income and wealth favours the electoral participation: the lesser income people earn, the lesser the participate in voting. These results also have been confirmed by Nwankwo et al. (2017). In fact, in the case of Catalonia, clearly is Barcelona the concurrency that concentrates more power and more inequality between the population. So, we think that our results are also consistent with that model.

Finally, in Table 3 we have shown some more specific patterns that we consider not enough significant. However, it is interesting to note that one of them relates low immigration and low number of families without children with low or very low abstention. Reading this pattern in a inverse way, we could say that in electoral sections where the population are mainly families with children, i.e., with high engagement to the territory, there is a high turnout. This fact is also consistent with the studies of Solt (2008), Nwankwo et al. (2017), and Leighley and Vedlitz (1999) in the sense that although marital status is not a determinant factor for the interest in politics, it should be taken into account in the studies. The other pattern in Table 3 reflects the converse situation: electoral sections with high immigration and low families with children have high or very high abstention (i.e., low turnout). This could be read as electoral sections where the population has low engagement to the territory. Our conclusions are also similar to the ones achieved by Daoust and Blais (2017) since they have also found that people more integrated into their community have stronger attachment to their municipality and, as a consequence, they are more likely to participate. In principle, we could assume that the majority of immigrant people is not fully integrated to a concurrency.

As a final remark, we want to comment that when the tree is grown without taken into account the attribute Immigration, the most discriminant attribute is the Income, as in Barcelona. We have also seen that the correlation between the abstention and the income is $r = 0.2944$, significantly lower than the correlation between the abstention and the immigration. In addition, we have seen that the correlation between the immigration and the income is $r = -0.1593$, i.e., both variables are inversely correlated but such correlation is low. This first approach seems to suggest that the immigration is a powerful explanation for the abstention. However, in the future we plan to analyze this scenario in more depth for each one of the concurrencies.

5 Conclusions and future work

In the present paper we propose the use of decision trees to analyze the abstention of the elections to the Catalan parliament held in 2015. Different from the most of studies about abstention, our results are not based on questionnaires but in real data coming from socio-demographic information and public electoral results. The socio-demographic information is based on a typology of families given by the AIS Group Company.

Concerning the methodology used in the analysis, decision trees have proved to be a good tool to interpret the results. They provide comprehensible results and are flexible enough to search for different relations between the attributes and the solution classes.

The results of the study have been discussed globally and for each one of the Catalan concurrencies. Our conclusions is that the percentage of abstention depends on the level of immigration of the electoral sections: The more immigration, the more abstention. This is true globally and also for each concurrency. This result is coherent with studies performed by other authors in other countries. Therefore, despite of the political context lived in Catalonia in 2015, heavily polarized among the supporters and detractors of independence, we have seen that it only has influenced the level of turnout (almost the 80% of the population) but not the profile of voters.

Barcelona also follows the same trend, however, we have found that the most relevant feature is the familiar income average: the high income, the less abstention. This result is also consistent with other studies concluding that the economic inequality of the population favours the abstention. In our case, Barcelona is an economic pole and clearly the economic aspect is more important than the territory engagement. Also, it is the concurrency with highest economic inequality.

In the present work we have used decision trees because our goal was to find patterns. However, because of the resulting patterns are simplest than we expected (i.e., they are mainly composed of one attribute), we plan in the future to repeat the experiments using Random Forests and Principal Component Analysis to assess the relevance of all the attributes with respect to the abstention. In addition, it should be interesting to repeat the experiments of the present paper on data of other kind of elections: Spanish Parliament or European elections. The comparison of all these results will give us a complete picture of the abstention in Catalonia.

Acknowledgements The authors acknowledge the AIS Group enterprise for having given us the Habits© Data Base in a selfless way. This research is funded by the project PID2019-111544GB-C21 of the Spanish Ministry of Economy and Competitiveness.

Funding This research was supported by Consejo Superior de Investigaciones Científicas (Grant 201650E044) and Agència de Gestió d'Ajuts Universitaris i de Recerca (Grant 2014-SGR-118).

References

Adams J, Dow J, Merrill S (2006) The political consequences of alienation-based and indifference-based voter abstention: applications to presidential elections. Polit Behav 28(1):65–86

AIS AG (2019) www.ais-int.com/marketing-y-ventas/geomarketing-habits-big-data/

Armengol E, Vicente Z (2019) Analysis of abstention in the elections to the catalan parliament by means of decision trees. In: Modeling Decisions for Artificial Intelligence—16th International Conference, MDAI 2019, Milan, Italy, September 4–6, 2019, Proceedings, pp 248–258, https://doi.org/10.1007/978-3-030-26773-5_22

Arowolo D, Aluko OA (2013) Democracy, political participation and good governance in Nigeria

Blais A, Kostelka F (2015) The decision to vote or abstain in the 2014 European elections. Revue européenne des sciences sociales 2015/1(53–1):79–04

Blais A, Pilet J, Straeten K, Laslier J, Héroux-Legault M (2014) To vote or to abstain? An experimental test of Rational Calculus in First

Past the Post and PR elections. Electoral Stud. https://doi.org/10.1016/j.electstud.2014.07.001
Castela E, Villardón MPG (2011) Inferencia ecológica para la caracterización de abstencionistas: El caso de Portugal (in Spanish). Spatial Organ Dyn 3:6–25
Cazorla-Martín A, Rivera-Otero JM, Jaráiz-Gulías E (2017) Structural analysis of electoral abstention in the 2014 european parliamentary elections. Revista Española de Investigaciones Sociológicas pp 31–50
Daoust J, Blais A (2017) How much do voters care about the electoral outcome in their district? Representation 53(3–4):233–246. https://doi.org/10.1080/00344893.2018.1424028
Ferreira P, Dionisio A (2008) Voters dissatisfaction, abstention and entropy: analysis in European countries. Tech. Rep. 2008-11, University of Evora, CEFAGE-UE (Portugal), https://ideas.repec.org/p/cfe/wpcefa/2008_11.html
Gill J (2005) An entropy measure of uncertainty in vote choice. Electoral Studies 24(3):371–392, http://www.utdallas.edu/dept/socsci/hclarke/es/
Holland I, Miskin S (2002) Interpreting election results in Western Democracies. Current Issues Brief no2 2002-03 Politics and Public Administration Group Parliament of Australia https://www.aph.gov.au/About _Parliament/Parliamentary_Departments/Parliamentary_L
King G (1997) A Solution to the Ecological Inference Problem. Princeton University Press
Kirchgässner G (2003) Abstention because of indifference and alienation, and its consequences for party competition: a simple psychological model. University of St. Gallen Department of Economics working paper series 2003 2003–12, Department of Economics, University of St. Gallen
Leighley JE, Vedlitz A (1999) Race, ethnicity, and political participation: competing models and contrasting explanations. J Polit 61(4):1092–1114
López de Mántaras R (1991) A distance-based attribute selection measure for decision tree induction. Mach Learn 6:81–92
Maimon O, Rokach L (eds) (2010) Data mining and knowledge discovery handbook, 2nd edn. Springer
Nwankwo C, Okafor U, Asuoha G (2017) Principal component analysis of factors determining voter abstention in South Eastern Nigeria. J Pan Afr Stud 10:249–273
Quinlan JR (1979) Discovering rules by induction from large collection of examples. In: Michie D (ed) Expert systems in the microelectronic age. Edimburg Eniversity Press, pp 168–201
Quinlan JR (1986) Induction of decision trees. Mach Learn 1(1):81–106
Rosema M (2007) Low turnout: Threat to democracy or blessing in disguise? Consequences of citizens' varying tendencies to vote. Electoral Stud 26(3):612–623. https://doi.org/10.1016/j.electstud.2006.10.007
Solt F (2008) Economic inequality and democratic political engagement. Am J Polit Sci 52(1):48–60
Villardón MPG (1986) Una alternativa de representación simultánea: HJ-Biplot (in Spanish). Questiio 10:13–23