MATEMÁTICA
ESPOL - FCNM JOURNAL
2019, Vol. 17, No. 2 Publicado: 2019-12-27

The first construction of a quantitative profile for elected mayors in an Ecuadorian election

Construcción de un perfil cuantitativo de alcaldes en una elección en Ecuador

Wendy Chávez and Jessica Vera-Bermudez and Omar Ruiz-Barzola

Abstract This research puts into context statistical multivariate methods in the political arena, making the first contribution to the analysis of Ecuadorian politics on the definition of elected mayors' profile. In this regard, a logistic regression model was estimated to describe voters' preferences in terms of candidates' age, gender and political organization; and, a CHAID decision tree was built to delve into the preferred features depending on the political party. This latter analysis revealed interesting differences that also help to describe Ecuadorian political reality. The sample upon which the profiles were built included the voters' preferences from the four most populated Ecuadorian provinces in the mayors' election in 2014. This analysis combines the application of statistics in the political science field and proposes to be the first step for future research to analyze how the preferred profile of elected candidates changes over time.
Keywords: CHAID, logistic regression, Ecuadorian politics, mayors' profile.

Resumen Esta investigación matiza la utilidad de los métodos estadísticos multivariantes con la política, haciendo de este trabajo la primera contribución al análisis de la política ecuatoriana en lo que respecta a definición del perfil de alcaldes electos. En este sentido, se estimó un modelo de regresión logística para describir las preferencias de los votantes en términos de la edad, género y organización política

Wendy Chávez, Master in Public Administration
OPPG Observatorio de Políticas Públicas de Guayaquil, Kennedy Norte Mz. 801 V. 4 Guayaquil-Ecuador, e-mail: lachavez34@yahoo.com

Jessica Vera-Bermudez, BS. Computational Statistics
ESPOL Polytechnic University, Campus Gustavo Galindo Km. 30.5 Vía Perimetral P.O. Box 09-01-5863 Guayaquil-Ecuador, e-mail: jesmvera@gmail.com

Omar Ruiz-Barzola, PhD. Statistics and Optimization
ESPOL Polytechnic University, Escuela Superior Politécnica del Litoral, ESPOL, Facultad de Ciencias de la Vida (FCV), Facultad de Ciencias Naturales y Matemáticas (FCNM), Campus Gustavo Galindo Km. 30.5 Vía Perimetral P.O. Box 09-01-5863, Guayaquil, Ecuador, e-mail: oruiz@espol.edu.ec, https://orcid.org/0000-0001-8206-1744

de los candidatos; y, se construyó un árbol de decisión CHAID para profundizar en las características preferidas, según el partido político, revelando diferencias interesantes que además ayudan a describir la realidad política de Ecuador. La muestra empleada para construir los perfiles incluyó los votos de los ciudadanos de las cuatro provincias más pobladas de Ecuador en la elección de alcaldes de 2014. Este análisis combina la aplicación de estadísticas en ciencias políticas y propone ser el primer peldaño para futuras investigaciones acerca de cómo el perfil de candidatos electos cambia en el tiempo.
Palabras claves: CHAID, regresión logística, política ecuatoriana, perfil de alcaldes.

1 Introduction

Multivariate statistical methods have been widely used to predict behaviors and preferences. The political arena is no exception, since predicting voters and candidates' behavior has attracted society's attention. In this manner, logistic regression has been used as a modeler of preferences (Caprara et al., 2006), as well as a mechanism to test campaigns' effects (Kim, 2011). On the other hand, CHAID, one type of the classification trees, has also been used as a tool for politicians in the marketing context (Baines et al., 2002; Walker et al., 2017) as a method to segment.

Research demands not only a quantitative analysis, but it also requires the qualitative perception, so that a realistic analysis is conveyed. Then it is necessary to understand the complexity of the Ecuadorian political context and the major events that have characterized it; such as the successes and failures of traditional parties and the rise of non-traditional movements. Not only in Ecuador but in Latin-American, clientelism, corruption scandals, land-trafficking, narco-trafficking, terrorism and the detriment of life conditions of the population from 1970s to 2000 contributed to the strengthening of social movements that later became political movements or parties. In general, these social and political movements are disruptive to the status-quo (Somuano, 2007). These political organizations have evolved over time and new ones have been created. This evolution has been the focus of analysis of many political scientists.

Ecuador has some research on political evolution and behavior from the qualitative eyes (Mustillo, 2016; Mejía Acosta et al., 2005; Freidenberg and Sáez, 2001) but barely on the quantitative analysis. On the other hand, in most research articles of politics, the statistical methods described above are used in comparative or individual analysis. Since the wide usage of logistic regression in different fields has caused that its classification capacity is questioned, nowadays it is compared with classification trees, as Camdeviren et al. (2007); Abu-Hanna and de Keizer (2003); Fu (2004) analyzed in their works. However, further than a comparative or individual analysis, these two methods may be used as complements of each other so the capacity of both is taken.

This analysis aims to estimate the most preferred basic features of candidates in the majors' election through the representation of the votes from the four most populated Ecuadorian provinces by making the most of the two statistical methods mentioned in the previous paragraph along with the political reality of Ecuador. For this purpose, we employed a logistic regression model to estimate the probability for a candidate to win the local elections given candidates' gender, age and political party; and, CHAID decision tree to delve into preferred profiles.

2 Materials and Methods

2.1 Political Context

Freidenberg (2015) distinguishes between the traditional and non-traditional Ecuadorian parties, indicating that the traditional ones participated in electoral processes during and after 1978–1979—when the military government of Ecuador called for national open elections after a dictatorship period. The non-traditional parties, are the ones that did not participate in such elections but in the subsequent ones. Applying this classification to analyze the political process in 2014, we found that only three traditional parties participated in the local elections, and four non-traditional ones (Table 1).

Table 1: Ecuadorian Political Parties participating in 2014 local elections.

Traditional	Non-traditional
PSC: *Partido Social Cristiano* (created in 1951 as movement and becoming a political party in 1967)[a]	PS-FA: *Partido Socialista-Frente Amplio* (1995)[b]
MPD: *Movimiento Popular Democrático* (1978)	PSP: *Partido Sociedad Patriótica 21 de Enero* (2002)
PRE: *Partido Roldosista Ecuatoriano* (1984)	PRIAN: *Partido Renovación Institucional Acción Nacional* (2002)
	AVANZA: *Avanza* (2012)

[a] During the elections of 1978 the Partido Social Cristiano (PSC) participated as part of the Frente Nacional Constitucionalista.

[b] PS-FA was constituted in 1995 by the *Partido Socialista*, PSE, created in 1925, and the *Frente Amplio de Izquierda*, FADI, created in 1977 (Machado Puertas, 2007). Therefore, PS-FA is a non-traditional party due to its creation in 1995, but its origins are related to two traditional political parties.

2.1.1 Political power monopoly: rise and fall of traditional political parties

In Ecuador, the traditional parties controlled all executive positions at the national and local levels between 1978 and 1988. In the local elections of 1979, 1984 and 1988 the traditional parties were the only winning groups in most populated locations. Those parties had the monopoly of political competition; they also controlled positions, resources and relationships of power at local and national levels (Freidenberg, 2015). A monopoly of certain political parties meant that other organizations such as small social and/or political movements, had lower or no power to influence in decisions at the local or the national level.

Another characteristic of the traditional political parties was that their strength was based geographically on different areas of the Ecuadorian territory. For example, in 1984 the PSC and PRE were the dominant parties in the Coast and had relatively low support in the Sierra, while ID and DP were strong in the Sierra and not in the Coast. This territorial fragmentation or territorial presence of the parties (Sánchez, 2008) was also expressed at the Congress, where presidents had to manage a network of "ghost coalitions" (Mejía-Acosta, 2004, p.45) to ensure minimal governance. This situation was described by Collins (2006) who stated that "during the 1980s and 1990s the party system of Ecuador was one of the systems with the most heavily fragmented parties of the region, none of those was strong at the national level; and barely existed ideological coherence among them" (p.79-80).

This geographical fragmentation and consequent need of coalitions to survive contributed for the traditional political parties to lose power. Freidenberg (2015) developed the "index of predominance of traditional parties in executive position" (p.80) which is the proportion of national and subnational governments controlled by traditional parties in a determined election. The index of predominance of traditional political parties was one in the elections of: 1978-1979, 1984, 1988—showing absolute predominance of traditional political parties—and decreased until it got to its lowest value of 0.2 in 2009, showing predominance of non-traditional parties in power. Furthermore, in 2014 the old traditional organizations PRE, PRIAN and MPD were deregistered from the National Electoral Council because they could not get the minimum of four percent of the votes in 2013 and 2014. These facts show how traditional political parties lost power and the non-tradition ones became stronger.

The non-traditional movement that started this new era was MPAIS. It was structured on the basis of the small movements: *Iniciativa Ciudadana*, *Acción Democrática Nacional*, *Alianza Bolivariana Alfarista* and *Jubileo 2000*; and social organizations. Before gathering together, these small movements had a low participation in the decision-making process in the political or the public administration arena and had in common the frustration of not holding power for decades then they concentrated their efforts to achieve progressive social change (Roach and Roach, 1979, 1980; Gurr, 2010; Maier, 1942). Once they created the first structure of MPAIS, they became the biggest political movement of Ecuador, led by Rafael Correa, who won as President of Ecuador three times with over 50 percent of votes each election (MPAIS, 2006; Becker, 2013; Polga-Hecimovich, 2013), and formed

the Constituent Assembly with over 60 percent of MPAIS members (MPAIS, 2006; Polga-Hecimovich, 2013), supported by the vast majority of votes.

The presence of MPAIS broke the Ecuadorian pattern of being governed by traditional political parties that were born in 1978-1979, and became the first political force in the country but after the local elections in 2014 its level of power started to vary since opposing candidates won in some of the biggest cities of the country: Guayaquil, Quito, Cuenca, Manta and Santo Domingo (Vera Rojas and Llanos-Escobar, 2016). From MPAIS' constitution and success on, alliances became a popular option. In fact, Hernández and Buendía (2011), thinkers of MPAIS who documented in 2011 its successful path in their article "Ecuador: Advances and Challenges of MPAIS" recommended that the movement "would have to recompose its policy of alliances with the different sectors of society, in particular with the indigenous, peasants, workers and social groups movements in order to consolidate deep territorial agreements" (p. 142). This last idea is related to the need of alliances to maintain power at the municipal or local level.

The literature about Ecuadorian politics is mostly produced on the historical analysis of political movements and parties and their struggle for holding power in different periods. However, no previous analysis with a quantitative approach has been produced referring to the profiles of elected candidates, reason why this study is the first of its kind and proposes to be the first step for future political analysis. We use a logistic regression to build the winners' profile testing the attributes age, gender and political alliance in the four most populated provinces of the country: Guayas, Pichincha, Manabi and Azuay.

2.2 Dataset

The statistical information for this research has been produced based on the 2014 local elections for the position of Mayor in the Ecuadorian territory. The database gathers information at the polling station level and was provided by the National Electoral Council (CNE). Ecuador has 221 municipalities but this study focused on the provinces: Guayas, Manabi, Pichincha and Azuay, where the most populated counties are located: 70 municipalities and adding to 57 percent of the total population of Ecuador, according to the National Institute of Statistics and Census (INEC, 2010). They also concentrated 63.8 percent of voters of the 2014 election. The candidacies for mayors represented 79 alliances, 64 political movements and 7 political parties.

In order to analyze and build the profiles of the elected candidates, the dependent variable is binary and expresses if the candidate was elected ("Elected") or not ("No Elected"), and the independent variables are: organization, sex and age. Organization is a qualitative variable describing the twelve kinds of political organizations in the study. These include political parties, political, social and local movements (*movimientos locales*), and alliances/coalitions (*alianzas*) reported in the 2014 election. The variable sex has two response options: male and female. Age is a four-

categories variable: under 45 years old, between 45 and 64 years old and older than 64 years old.

2.3 Statistical Methods

We obtained the elected candidates' profile by performing logistic regression models (Cox, 1958) to explain the probability of a candidate to be elected, through the independent variables mentioned previously). Then, the logistic regression model is obtained as follows:

$$\pi(y) = \frac{e^{\beta_0 + \sum_{i=1}^{p} \beta_i x_i}}{1 + e^{\beta_0 + \sum_{i=1}^{p} \beta_i x_i}}$$

Where y is the response variable and $x_i, (i = 1, 2, ..., p)$ are p independent variables.

Now, the coefficients are estimated through a log-likelihood function, defined in terms of $\pi(x)$ and their significance values are verified with Wald test. which uses hypothesis testing on the estimated coefficients (Agresti and Kateri, 2011; Lemeshow et al., 2013).

We performed four different logistic regression models on a 70 percent of the data selected randomly. We chose the model that fulfills the following criteria:

1. to contain only variables that are significantly associated with the candidate status at significance level of 0.05;
2. to generate the lowest BIC value, Bayesian Information Criterion (Schwarz, 1978; Neath and Cavanaugh, 2012);
3. to be comprehensible, interpretable and realistic regarding this dataset and the Ecuadorian politics; and
4. to produce the highest sensitivity[1] (Carugo et al., 2016) obtained from

$$Sensitivity = \frac{\sum Tp}{\sum Positive_{Observed}}$$

Given a confusion matrix in a classification evaluation (Carugo et al., 2016) with observed values as rows and predicted values as columns (Table 2).

It should be pointed out that models built using compounded variables were the first ones dismissed under researchers' criteria for being complex and deficient in significant independent variables.

[1] A candidate was categorized as "elected", if the probability of winning was 0.20 or greater in the model; otherwise was named as "no elected". We chose this value because the minimum percentage of votes that was needed by an organization to win the 2014 elections (in the cities of our sample) was 21.11 percent.

Table 2: Confusion matrix structure

Observed/Predicted	Negative	Positive
Negative	Tn =True negatives	Fp =False positives
Positive	Fn =False negatives	Tp =True positives

We also performed a classification tree for categorized dependent and independent variables, named CHAID (Chi-Square automatic interaction detector) by Kass (1980). CHAID uses the p-value of the Chi-square as splitting criteria.

$$\chi^2 = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Where $i = 1,2,...,r$ and $j = 1,2,...,c$; O_{ij} represents the observed frequency of the cell; and E_{ij} represents the expected frequency of the cell.

In this manner, before splitting step, all predictors are evaluated at each branch, in reference to their independence with the output. Then, the predictor with the smallest p-value is chosen for splitting (Tsiptsis and Chorianopoulos, 2011). CHAID tree was built in order to prioritize characteristics that define a candidate as elected according to the position of the variable in the tree: the upper a variable is the greater association with the response variable it has; and, to observe candidates' behavior depending on the political organization to which they belong.

As a tree, CHAID receives stopping and splitting criteria so we set 100 as parent node' maximum size, 50 as child nodes' maximum size, and 0.05 as the statistical significance for the splitting step. Various authors thoroughly explained some parameters that modify the tree depth, levels below the root and some other rules for the splitting and stopping steps in their works (Tsiptsis and Chorianopoulos, 2011; Wilkinson, 1992; Magidson and Vermunt, 2005).

The data were processed by means of two statistical software: R (Team, 2013) and SPSS version 24.

3 Results

Table 3 illustrates the classification test of the final model.

Table 3: Confusion matrix of the selected model

Observed/Predicted	Negative	Positive
Negative	$Tn = 822$	$Fp = 332$
Positive	$Fn = 40$	$Tp = 168$

This model yielded the highest sensitivity rate

$$Sensitivity = \frac{168}{168+40} = 0.808$$

which is the probability that the model predicts correctly the status of a winner candidate. In general terms, the classification power of the selected model is

$$Accuracy = \frac{822+168}{822+332+40+168} \times 100\% = 72.7\%$$

Table 4 shows the coefficients summary of the selected logistic model. It points out that a candidate from Alliances, who is male and younger than 45 years old yields the highest probability of winning the elections, 0.493. On the other hand, local movements, MPD, MUPP, PRIAN, PSP, women and candidates older than 45 are characteristics that minimize the probability of being elected. For example, women from MPAIS younger than 45 years old had 0.139 probability to win local elections, and women from MPAIS between 45 and 64 years old, had yet less chances (0.079 probability). However, a less favorable set of conditions was the one defined by the characteristics: being a woman, from MUPP and between her 45 and 64 years old; the probability of winning the 2014 Elections was 7×10^{-10}.

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_{1k} \times Organization_k + \beta_2 \times Gender + \beta_{3l} \times Age_l$$

Table 4: Selected logistic regression model[b]

Variable	Coefficients	SE	p-value[c]
Constant	**-2.1567**	0.2366	< 0.0001 ***
Organization$_k$			
$k = 1$ Alliances	**0.9549**	0.1730	< 0.0001 ***
$k = 2$ CREO	-1.0841	0.2909	0.0002 ***
$k = 3$ Movimientos locales	-18.3532	764.1787	0.9808
$k = 4$ MPAIS	0.3344	0.2049	0.1027
$k = 5$ MPD	-17.9603	897.4928	0.9840
$k = 6$ MUPP	-18.2904	706.0891	0.9793
$k = 7$ PRE	-3.2183	1.0194	0.0016 **
$k = 8$ PRIAN	-17.9589	1075.4013	0.9867
$k = 9$ PS-FA	-0.9825	0.2656	0.0002 ***
$k = 10$ PSP	-18.0945	656.2583	0.9780
$k = 11$ SUMA	0.2898	0.2312	0.2102
(base=AVANZA)			
Gender			
Male	**1.1750**	0.2062	< 0.0001 ***
(base=Female)			
Age$_l$ *(years)*			
$l = 1$ Between 45 and 64	-0.6255	0.1189	< 0.0001 ***
$l = 2$ 65 and more	0.2672	0.1970	0.1749
(base=less than 45)			

[b] AIC: 2107,02 and BIC: 21.
[c] Signif. codes: 0 *** 0.001 ** 0.01

The logistic regression provided a general profile of the winners, but in order to analyze the preferences of the voters of each political organization it was necessary to use the three-level CHAID shown in Fig 1 and Fig 2, that illustrate the most and least elected candidates' profiles, respectively. Both Fig 1 and Fig 2 display that a candidate status is best predicted by the political organization since it is in the highest level of the tree. The elected candidates at the national level won in 720 polling stations and out of those, Fig 1 points out that the candidates of alliances won in 404 polling stations followed by candidates of MPAIS and SUMA that together won in 178 stations.

At this level, the CHAID analysis gathered the information of some organizations together because of its similarity of data behavior during the elections in 2014, as in node 1 (Fig 1) and node 5 (Fig 2). This is the case of MPAIS and SUMA that were put together in the same box. We analyzed the data and found that 75.3 percent of the candidates of MPAIS and 73.5 percent of the candidates of SUMA were elected at the polling stations level; and the percentages of non-elected were 24.7 and 26.5, respectively. The political differences between both movements are discussed later.

The most successful candidates from Alliances (node 2) were younger than 45 regardless their sex. However, a male in this age group has a higher chance of winning, opposite to women' situation: male candidates from Alliances won in 152 polling stations and women only in 28 (Fig 1). AVANZA (node 3) had also two con-

ditions for the candidates in order to be winners: most successful candidates were men of at least 45 years old.

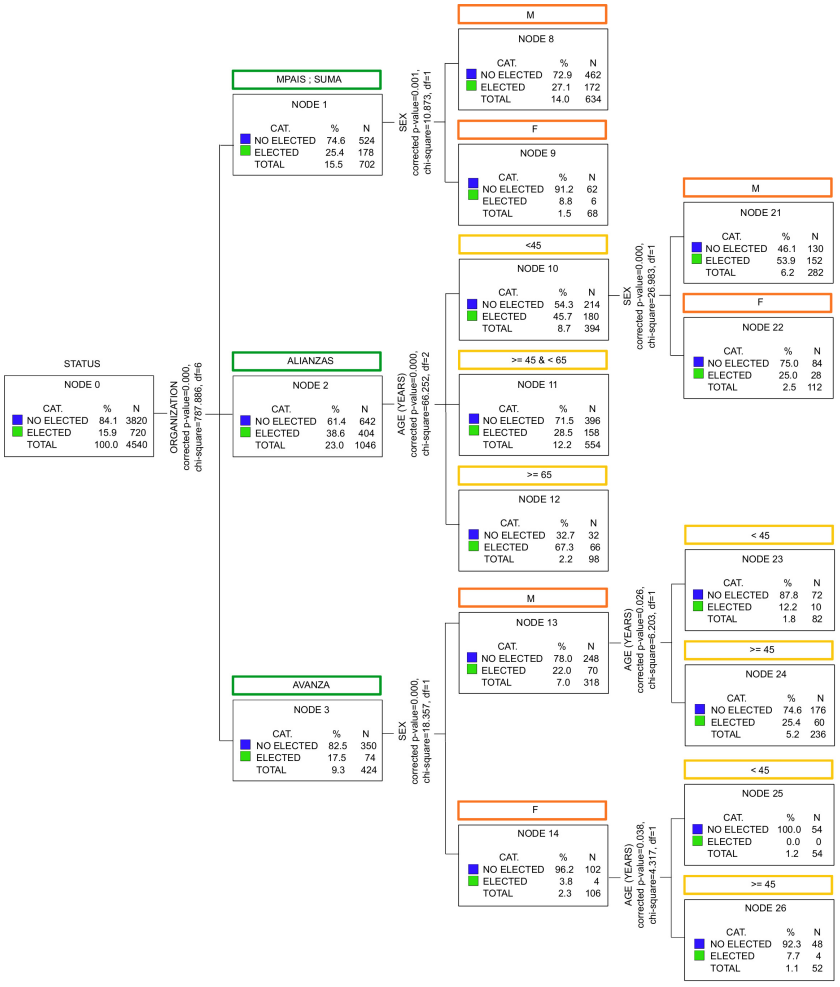

Fig. 1: CHAID decision tree including only the three most elected political parties.

Following with the CHAID analysis the second variable that characterized winners from MPAIS, SUMA (Fig 1), CREO and PRE (Fig 2) was sex. From PRE and CREO no woman was elected. Women candidates from MPAIS and SUMA won in only six polling stations. On the other hand, male candidates were elected in two (PRE), 22 (CREO) and 172 (MPAIS and SUMA) polling stations. Also, no other variable was important besides sex for these organizations. The political party PRE was the only one that hosted more female candidates than male candidates (Fig 2),

but none of them won the elections. Referring to age, it was the second-best predictor for PS-FA elected candidates. Mostly successful candidates were younger than 45 years old.

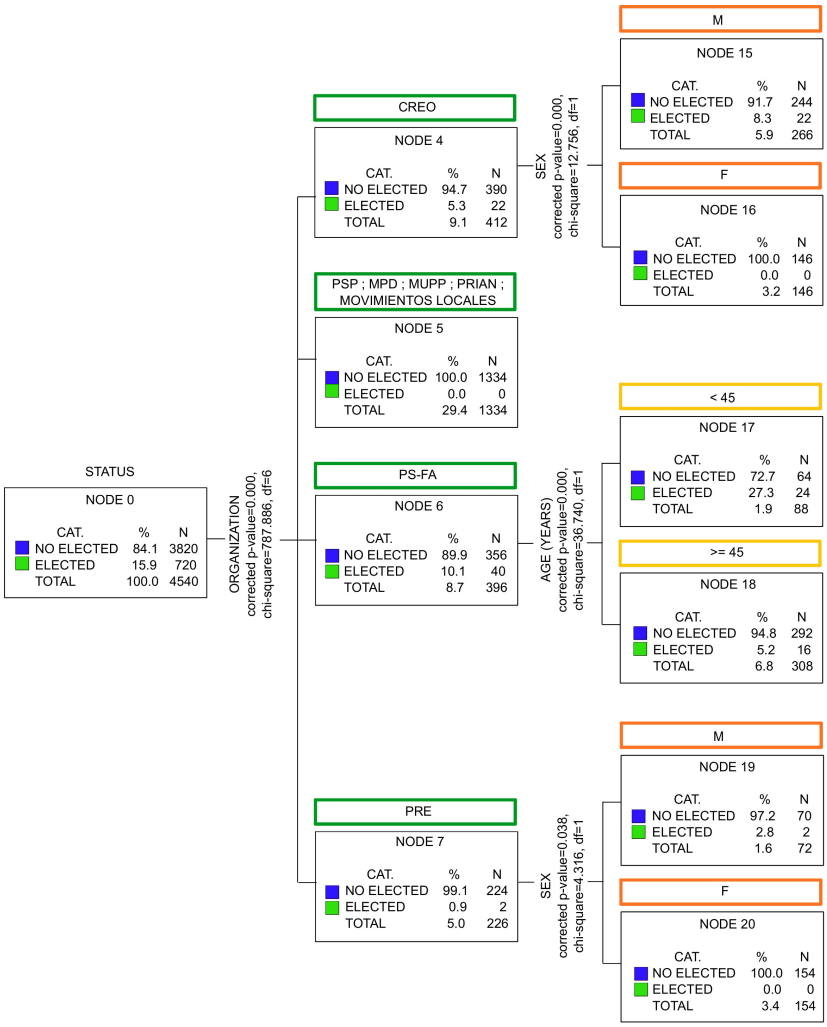

Fig. 2: CHAID decision tree including the least voted political parties.

Since the political organization through which candidates participate had strong influence on the election' results, according to the previous analyses, we also analyzed their individual behavior at counties level. Table 5 summarizes the number of counties where each political organization won, highlighting that alliances held

their position as preferred political organization at county level, followed by MPAIS and AVANZA.

Table 5: Winning percentages in counties by political organization.

Political organization	% Counties
PRE	1.4%
MPD	0.0%
MPAIS	22.9%
PSP	0.0%
MUPP	0.0%
AVANZA	14.3%
CREO	4.3%
PRIAN	0.0%
PS-FA	10.0%
SUMA	7.1%
MOVIMIENTOS LOCALES	0.0%
ALIANZAS	40.0%

In order to test the model accuracy and its forecasting capacity, we computed the probabilities of winning the 2019 local elections for the elected mayors in Cuenca, Guayaquil, Portoviejo and Quito, capital cities of the provinces Azuay, Guayas, Manabi and Pichincha, respectively. Table 6 also displays percentages of votes for each mayor.

Table 6: Evaluation of the model accuracy in the 2019 local elections of the four main capital cities

Cities	Elected Mayors	Political Organization	Age	Gender	Observed votes	Prediction
Cuenca	Pedro Palacios	Alliance	44	Male	28.06%	0.493
Guayaquil	Cynthia Viteri	Alliance	54	Female	52.6%	0.138
Portoviejo	Agustín Casanova	Alliance	59	Male	43.77%	0.342
Quito	Jorge Yunda	*Movimientos locales*	54	Male	21.38%	2.1×10^{-9}

The model was accurate enough to predict Pedro Palacios and Agustín Casanova's triumphs as mayors in Cuenca and Portoviejo, respectively. Pedro Palacios, a 44-year-old male candidate representing an alliance, fulfilled with the strongest attributes of a winner candidate in our model (probability of 0.493 of being elected). Agustín Casanova is the 59-year-old male elected mayor of Portoviejo, who ran for elections representing an alliance and according to our model he had 0.34 probability of winning. However, the model forecast low probabilities for the elected mayors in the two most populated cities of Ecuador: Guayaquil and Quito. For both, the estimation error is large enough to predict that these candidates would have lost the

local elections in their cities. In the former city, the model penalized the gender of the candidate, resulting in a probability of winning of 0.138. In the latter city, it punished the political organization the candidate represented, giving an approximately zero probability of winning the local elections.

The comparisons between the results of our model and the reality of the Ecuadorian elections in 2019, focused on these four cities, are described in the discussion section.

4 Discussion

We found that the most voted mayors were male, younger than 45 years old, elected through alliances; whereas, participating through local movements, MPD, MUPP, PRIAN, PSP, being a woman and/or older than 45 are attributes that minimize the probability of being elected. This means that participating through an alliance allowed the candidates to win elections (40 percent of counties won, irrespective of either candidates' sex or age) but at the same time let them maintain their identity. As defined by Somuano (2007), alliances involve close collaboration on specific issues, but both the party and the organization retain their own separate structure and general freedom of action. This was also discussed by Della Porta and Diani (2015), stating that: "actors instrumentally share resources in order to achieve specific goals; yet do not develop any particular sense of belonging and of a common future during the process. Once a specific battle has been fought, there need not be any longer-term legacy in terms of identity and solidarity, nor attempts to connect the specific campaign in a broader framework."

AVANZA and alliances were the only two political organizations in which both candidates' sex and age played a role in their candidacies' results: in AVANZA, a male candidate of 45 years old and older, has more possibilities of winning an election than a woman of the same age. For alliances, candidates who are between 45 and 64 have a higher chance of winning, because that range of age does not present any nodes down but if the candidate is younger than 45 and is male, his possibility of winning is still high, opposite to the situation of a woman. In fact, mostly men were elected in the political organizations where the candidates' sex was the defining attribute: MPAIS, SUMA, AVANZA, CREO and PRE. The two former organizations had no other attribute determinant, which means that the age of the candidates did not influence the decision of the voters of these two political movements. On the other hand, age is the second aspect that was considered by the voters—after the political organization—specifically for the ones who voted for Alliances and PS-FA: mostly successful candidates were younger than 45 years old. Apparently, citizens who voted for Alliances and PS-FA were mostly moved by the age instead of the sex, when voting for the winner.

We also found that MPAIS and SUMA behaved statistically similar so they appeared together in a box of the CHAID tree, which means that their results were similar (percentages of elected and non-elected candidates) but only at polling sta-

tions level since at counties level MPAIS won in the 22.9 percent of the counties and SUMA in 7.1 percent in the sampled provinces. Evidently these statistical analyses need the political context to be properly interpreted since these movements have different ideologies: MPAIS supports the Socialism of the Good Living (MPAIS, 2006) and SUMA (SUMA, 2012) is defined as a progressive center movement, thereby differing in the type of voters each one has, their set of principles, their ideas about how the free market should work and about the activities that need to be carried out by the public and the private sector.

It is fundamental to indicate that the results of this study cannot be compared to any other production of political analysis in Ecuador because there is no previous study that determines winners' profile and winning probabilities in mayor elections. However, some results of this analysis confront the political reality of this country, such as the polemic about female and male candidates: male candidates have been the favorite ones, according to the Ecuadorian history. This analysis supports that trend since citizens from Guayas, Pichincha, Azuay and Manabi who voted for these political parties, gave a high level of importance to the gender of the candidates, and specially preferred to vote for male candidates than for women. It is also important to point out that most candidates were male in all political parties, with the exception of PRE, which hosted more female candidates than male candidates though none of them won the elections.

By comparing the proposed model to the results of the local elections in 2019, we found that the model complies with the results of the elected mayors of Cuenca and Portoviejo, capital cities of the provinces Azuay and Manabi. The model estimated probabilities of winning the local elections as 0.49 and 0.34, respectively, which are enough to consider a candidate as a winner (at least 0.21), according to what was assumed and mentioned in previous sections.

In Guayaquil, capital city of Guayas, the results of the 2019 local elections, ended with the proclamation of Cynthia Viteri as female mayor of Guayaquil, who ran for elections representing an alliance. On one hand, her case complies with the model since alliances are the winners of the elections based on our predictions but on the other side, the model indicates that the female gender minimizes the probability of winning. She represented the alliance between the traditional political party she belongs to, *Partido Social Cristiano* (PSC), and the political movement *Madera de Guerrero* (MG). Nowadays, this political party has its headquarters in Guayaquil city, where relies its strongest political force which is relatively weak in other geographic areas of the country. It has had considerable voters' support since its first elected mayor in 1992, León Febres Cordero. Viteri is the third mayor in a row who belongs to the same political party, in power for 27 years now (1992-2019). These results confirmed what we stated in previous paragraphs about the capacity of the political organizations to become a decisive feature for winning elections, beyond sex and age.

The current mayor of Quito, capital city of Pichincha, has a virtually zero probability of winning, according to the model proposed. Jorge Yunda represents a new political movement that was created in 2014 (*Unión Ecuatoriana*). In 2019 this movement had candidates for different counties of Ecuador, reason why con-

ceptually it must not be considered as a local movement (El Comercio, 2018). Quito represents a case study that should include the analysis of the dynamics between the voters and the political parties in power, because voters elected mayors from the same political party: Democracia Popular DP (1984-2000) and Izquierda Democrática (2000-2009), and then their votes changed preferring candidates from different political movements and parties for each election since 2009: MPAIS (2009-2014), SUMA (2014-2019) and Unión Ecuatoriana UE (2019-2022).

The evaluation of women's role in politics had deserved numerous publications discussing especially the tough way that women had passed through to get involved in politics; since their right to vote to their right to be elected. For instance, only 11 women out 221 were elected as officials to local governments in 2008 (Ágora-Política, 2012); and, in the 2014 elections, women represented only 12.2 percent of the registered candidates for mayors and 7.2 percent of the winners (CNE, 2014). However, the triumph of the mayors in Guayaquil and Quito in 2019 displays the limitations of our results since the force of political organizations have not been finely measured. This is important for certain locations, where the dynamics between the voters and the political parties in power can be very specific and constitute a case study, being this one of the opportunities for future research that this study opens.

As researchers, we see tremendous value in this study to inspire the use of statistical methods in the political analysis. The methodology used and its output reflect the Ecuadorian reality, making our analysis of profiles consistent and adequate (sensitivity of 0.808 and accuracy of 0.727 in the logistic regression model) despite it is not based on quantitative variables. It also showed accuracy in the 2019 elections. However, this study was mostly made at polling stations level since the database' layout was given by candidates and his or her result in each polling station, thereby impeding to delve into real differences such as the behavior of MPAIS and SUMA at counties level. On the other hand, we analyzed the voters' sex but since it was not statistically associated with their vote, the analysis was based only on candidates' attributes.

This study needs to confirm the voters' preferred attributes depending on the political organization of their choice and women's performance in elections regardless of their low participation to show the evolution in the women's role in politics. Also, further research would be useful about how the coalitions are working on the territory and the types and levels of conflict they face during the exercise of power since coalitions came in handy for political parties to win, including the governing party. We also encourage researchers in the political arena to investigate the impact of candidates' campaigns in electoral results. And specific analysis is needed to study the case of Guayaquil and Quito, namely the political force of a traditional party and the impact of new political organizations.

We recommend developing datasets that are composed of the same variables to the corresponding entity so data are easier to manage, and scholars, in general, can benefit from them. It is possible and necessary to have a better processing of the dataset and increase the consistency of the structure for reporting future elections' results. The second recommendation is considering the changes that have been oc-

curring from 2017 on in the political structure of MPAIS, in the public sector and even the voters' opinions about these changes when analyzing future results based on this study. We also suggest that political experts and statisticians work together to analyze political theories, reality and present measurable results.

5 Bibliography

References

Abu-Hanna, A. and de Keizer, N. (2003). Integrating classification trees with local logistic regression in intensive care prognosis. *Artificial Intelligence in Medicine*, 29(1):5–23.

Ágora-Política (2012). Equidad poltica: Mujeres en el ejercicio del poder. *International IDEANIMD*, 3:32–34.

Agresti, A. and Kateri, M. (2011). Categorical data analysis. In *International encyclopedia of statistical science*, pages 206–208. Springer.

Baines, P. R., Harris, P., and Lewis, B. R. (2002). The political marketing planning process: improving image and message in strategic target areas. *Marketing Intelligence & Planning*, 20(1):6–14.

Becker, M. (2013). The stormy relations between rafael correa and social movements in ecuador. *Latin American Perspectives*, 40(3):43–62.

Camdeviren, H. A., Yazici, A. C., Akkus, Z., Bugdayci, R., and Sungur, M. A. (2007). Comparison of logistic regression model and classification tree: An application to postpartum depression data. *Expert Systems with Applications*, 32(4):987–994.

Caprara, G. V., Schwartz, S., Capanna, C., Vecchione, M., and Barbaranelli, C. (2006). Personality and politics: Values, traits, and political choice. *Political psychology*, 27(1):1–28.

Carugo, O., Eisenhaber, F., and Carugo (2016). *Data mining techniques for the life sciences*. Springer.

CNE (2014). Resultados de las elecciones seccionales 2014. Technical report, Consejo Nacional Electoral.

Collins, J. N. (2006). *Democratizing formal politics: indigenous and social movement political parties in Ecuador and Bolivia, 1978-2000*. PhD thesis, UC San Diego.

Cox, D. R. (1958). The regression analysis of binary sequences. *Journal of the Royal Statistical Society: Series B (Methodological)*, 20(2):215–232.

Della Porta, D. and Diani, M. (2015). *The Oxford handbook of social movements*. Oxford University Press.

El Comercio, D. (2018). El movimiento unión ecuatoriana proclamó la candidatura de jorge yunda para la alcaldía de quito. *Diario El Comercio*.

Freidenberg, F. (2015). dónde compiten los partidos en ecuador? estrategias territoriales multinivel en un sistema político unitario. *Revista Valor Agregado*, 2:81–84.

Freidenberg, F. and Sáez, M. A. (2001). *Los dueños del poder: los partidos políticos en Ecuador,(1978-2000)*. FLACSO, Sede Ecuador.

Fu, C. Y. (2004). Combining loglinear model with classification and regression tree (cart): an application to birth data. *Computational Statistics & Data Analysis*, 45(4):865–874.

Gurr, T. (2010). *Why Men Rebel*. Paradigm Publishers.

Hernández, V. and Buendía, F. (2011). Ecuador: avances y desafíos de alianza país. *Nueva Sociedad*, 234:12–142.

INEC (2010). Proyección de la población ecuatoriana, por años calendario, según regiones, provincias y sexo. Technical report, Instituto Nacional de Estadísticas y Censos.

Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied statistics*, pages 119–127.

Kim, Y. (2011). The contribution of social network sites to exposure to political difference: The relationships among snss, online political messaging, and exposure to cross-cutting perspectives. *Computers in Human Behavior*, 27(2):971–977.

Lemeshow, S., Sturdivant, R. X., and Hosmer, D. W. (2013). *Applied Logistic Regression (Wiley Series in Probability and Statistics)*. Wiley.

Machado Puertas, J. C. (2007). Ecuador: el derrumbe de los partidos tradicionales. *Revista de ciencia polí]tica (Santiago)*, 27:129 – 147.

Magidson, J. and Vermunt, J. K. (2005). An extension of the chaid tree-based segmentation algorithm to multiple dependent variables. In *Classificationthe Ubiquitous Challenge*, pages 176–183. Springer.

Maier, N. R. (1942). The role of frustration in social movements. *Psychological Review*, 49(6):586.

Mejía-Acosta, A. (2004). *Ghost Coalitions: Economic Reforms. Fragmented Legislatures and Informal Institutions in Ecuador (1979-2002)*. PhD thesis, University of Notre Dame.

Mejía Acosta, A., Freidenberg, F., and Pachano, S. (2005). La ciencia política en ecuador: un reflejo de su fragilidad democrática (1978-2005). *Revista de ciencia política (Santiago)*, 25(1):147–161.

MPAIS (2006). Manifiesto ideológico. Technical report, Movimiento Alianza País.

Mustillo, T. (2016). Party–voter linkages derived from the calculus of voting model: Electoral mobilization in ecuador. *Rationality and Society*, 28(1):24–51.

Neath, A. A. and Cavanaugh, J. E. (2012). The bayesian information criterion: background, derivation, and applications. *Wiley Interdisciplinary Reviews: Computational Statistics*, 4(2):199–203.

Polga-Hecimovich, J. (2013). Ecuador: Estabilidad institucional y la consolidación de poder de rafael correa. *Revista de ciencia política (Santiago)*, 33:135–160.

Roach, J. K. and Roach, J. L. (1980). Turmoil in command of politics: Organizing the poor. *The Sociological Quarterly*, 21(2):259–270.

Roach, J. L. and Roach, J. K. (1979). Disunity and unity of the working class: Reply to piven and cloward. *Social Problems*, 26(3):267–270.

Sánchez, F. (2008). *Democracia no lograda o democracia malograda? Un análisis del sistema político del Ecuador: 1979-2002*. FLACSO Andes.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464.

Somuano, M. F. (2007). Movimientos sociales y partidos políticos en américa latina: una relación cambiante y compleja. *Política y Cultura*, pages 31–53.

SUMA (2012). quiénes somos? Technical report, Movimiento Sociedad Unida Ms Acción.

Team, R. C. (2013). R: A language and environment for statistical computing.

Tsiptsis, K. K. and Chorianopoulos, A. (2011). *Data mining techniques in CRM: inside customer segmentation*. John Wiley & Sons.

Vera Rojas, S. and Llanos-Escobar, S. (2016). Ecuador: La democracia después de nueve años de la "revolución ciudadana" de rafael correa. *Revista de ciencia política (Santiago)*, 36:145–175.

Walker, L., Baines, P. R., Dimitriu, R., and Macdonald, E. K. (2017). Antecedents of retweeting in a (political) marketing context. *Psychology & Marketing*, 34(3):275–293.

Wilkinson, L. (1992). Tree structured data analysis: Aid, chaid and cart. In *Proceedings of Sawtooth Software Conference*.