

Grado en Estadística y Empresa  
2018/2019

*Trabajo Fin de Grado*

**“Procedimiento para estimar la  
intención de voto en España  
mediante árboles de decisión”**

---

**Autor: Álvaro Giménez Ramírez**

Tutor: Andrés M. Alonso Fernández

Universidad Carlos III de Madrid



## RESUMEN

En un momento de situación política muy convulso, las encuestas electorales y sus correspondientes predicciones son absolutamente necesarias. En este estudio se intentará encontrar un nuevo método de predicción de resultados electorales empleando árboles de decisión y las preguntas que el CIS realiza en sus encuestas. Para ello, se hará uso de las encuestas de marzo y abril que se llevaron a cabo para realizar la predicción de las elecciones generales de abril de 2019 y de las elecciones europeas de mayo del mismo año.

Primeramente, se analizan las variables empleadas para realizar el estudio, tomando en cuenta su significatividad en la intención de voto de los españoles. Posteriormente, se diseñan los árboles de decisión que sean necesarios. Por último, se lleva a cabo la imputación de los datos de modo que, se obtiene la intención de voto de aquellas personas entrevistadas que no la indicasen.

Los resultados obtenidos son bastante similares a los que se dieron en los dos procesos electorales dados y queda demostrado que, principalmente, la simpatía de voto, el recuerdo de voto y el nivel educativo son factores determinantes a la hora de votar.

**Palabras clave:** CIS; simpatía de voto; recuerdo de voto; nivel educativo; ideología; voto oculto.



## **DEDICATORIA**

A mi familia, en especial a mis padres, que han estado a mi lado desde mi primer día de universidad. Por todo el esfuerzo que han dedicado y sobre todo, por haberme apoyado en los momentos más difíciles y haber confiado en mí.

También a mi profesora Paula y a mi tutor Andrés, por todo el apoyo que me han brindado a cualquier hora del día y sobre todo, en verano.



# ÍNDICE

1. INTRODUCCIÓN.....	1
2. ANÁLISIS DESCRIPTIVO .....	3
2.1. Descripción univariante .....	4
2.2. Datos faltantes.....	13
3. METODOLOGÍA.....	17
3.1. Implementación .....	18
3.2. Árboles de decisión .....	19
4. RESULTADOS .....	25
5. CONCLUSIÓN.....	27
6. REFERENCIAS .....	29
7. ANEXOS.....	31

## ÍNDICE DE ILUSTRACIONES

Ilustración 2.1. Comparación de intención de voto.....	<b>¡Error! Marcador no definido.</b>
Ilustración 2.2. Comparación de simpatía de voto.....	5
Ilustración 2.3. Comparación de recuerdo de voto .....	6
Ilustración 2.4. Comparación de la ideología .....	7
Ilustración 2.5. Comparación de la situación económica.....	8
Ilustración 2.6. Comparación de la situación política.....	9
Ilustración 2.7. Comparación de la edad de los encuestados.....	9
Ilustración 2.8. Comparación de la situación laboral de los encuestados .....	10
Ilustración 2.9. Comparación del nivel educativo de los encuestados .....	11
Ilustración 2.10. Patrones de valor perdido de la encuesta de marzo de 2019.....	13
Ilustración 2.11. Patrones de valor perdido con mayor frecuencia en la encuesta de marzo de 2019.....	14
Ilustración 2.12. Patrones de valor perdido de la encuesta de abril de 2019.....	15
Ilustración 2.13. Patrones de valor perdido con mayor frecuencia en la encuesta de abril de 2019.....	15
Ilustración 4.1. Comparación de los resultados de la predicción respecto a la predicción del CIS y a los resultados de las elecciones del 28 de abril .....	25
Ilustración 4.2. Comparación de los resultados de la predicción respecto a la predicción del CIS y a los resultados de las elecciones del 26 de mayo .....	26

# 1. INTRODUCCIÓN

Las encuestas electorales son empleadas en la actualidad con gran asiduidad. Los medios de comunicación las muestran constantemente e incluso los partidos políticos hacen uso de ellas cuando necesitan tomar decisiones de calado para ver cómo respondería su electorado. Son muchas las empresas y organismos que se dedican a predecir los resultados electorales y, hoy en día, prácticamente nadie desconoce la existencia del tan afamado y criticado Centro de Investigaciones Sociológicas, más conocido como CIS.

En los medios de comunicación se habla de su *cocina electoral*, los partidos políticos incluso critican a dicho organismo y, en general, a todos aquellos que hacen las predicciones cuando éstas salen mal. Tal es así, que muchos recuerdan los fracasos de las predicciones en relación a la salida del Reino Unido de la Unión Europea. El fracaso que se produjo en Estados Unidos cuando ninguna encuesta pronosticó la elección de Donald Trump como presidente y por supuesto, en España, el *sorpasso* de Podemos al PSOE en las elecciones de diciembre de 2016 que nunca llegó a producirse.

Por este motivo, en este estudio se tratará de llevar a cabo una predicción, *a posteriori*, de los resultados electorales de las elecciones generales celebradas el pasado 28 de abril de 2019 y de las elecciones europeas celebradas el 26 de mayo del mismo año. Para llevarla a cabo, se emplearán las encuestas realizadas por el CIS en marzo y abril del mismo año que, el propio organismo, empleó para hacer sus predicciones. Estas encuestas se hicieron a españoles de más de 18 años de ambos sexos.

Para ello, el estudio comenzará analizando las preguntas que se tomarán en cuenta y, posteriormente, se emplearán los árboles de decisión como técnica estadística para la predicción de la intención de voto. De esta forma, se prevé que las preguntas mencionadas, sirvan de variables independientes que predigan la intención de voto de aquellas personas que no indican lo que votarán y que, dificulta la obtención de resultados veraces.

Este estudio tendrá la ventaja de poder ser comparado con los resultados que se dieron en dichas elecciones por lo que, se podrá comprobar la calidad de la predicción y por tanto, del método empleado.

Cabe destacar que las elecciones a estudiar vinieron precedidas por un momento convulso dentro de la política española. Un adelanto electoral por la falta de entendimiento de algunos partidos políticos para llegar a acuerdos como pudieran ser los presupuestos generales del Estado, la primera moción de censura con éxito en los más de 40 años de democracia española, la crisis catalana y por supuesto, la aparición

de nuevos partidos políticos o incluso las llamadas *fake news*, que tanto afectan a la política actual. Todas estas cuestiones, son factores que influyen en los votantes y que pueden llegar a generar dudas en cuanto a su intención de voto. Por ende, también influyen en las predicciones electorales que se pudieran hacer tomando en cuenta unas encuestas que se elaboraron un mes previo a la celebración de los comicios.

Para llevar a cabo este estudio se ha hecho uso de dos herramientas informáticas como son R y SPSS. Con la primera de ellas se ha llevado a cabo la recodificación de valores y la elaboración de ilustraciones para explicar las variables que se tomarán en cuenta. Por otro lado, SPSS será empleado para llevar a cabo la construcción de los árboles de decisión y para realizar la posterior imputación de datos que generará los resultados del propio estudio.

## 2. ANÁLISIS DESCRIPTIVO

Para llevar a cabo el estudio, se ha decidido hacer uso de dos encuestas llevadas a cabo por el Centro de Investigaciones Sociológicas (de aquí en adelante, CIS). La primera de ellas se realizó en marzo de 2019 de forma previa a las elecciones generales celebradas el 28 de abril y la segunda encuesta es la llevada a cabo en abril de 2019 con la vista puesta en las elecciones europeas del 26 de mayo.

El CIS realiza sus encuestas recorriendo “las secciones censales incluidas en la muestra, siguiendo unas rutas aleatorias, y contactan con los ciudadanos en sus casas”. Además, en todos los casos, la participación es absolutamente voluntaria de modo que, en cualquier momento de la entrevista, la persona entrevistada puede elegir no contestar alguna pregunta, si así lo desea (CIS, 2017).

El número de entrevistas que se llevaron a cabo en el caso de la encuesta de marzo fue de 16.191 y los entrevistadores tenían que hacer 31 preguntas a los encuestados. En abril se realizaron 39 preguntas a 17.574 encuestados. El motivo por el que en abril se formuló un mayor número de preguntas se debió a que en esa ocasión además de preguntar acerca de las elecciones europeas, también se celebraban en España las elecciones municipales y autonómicas.

Para realizar el estudio que se quiere tratar, se considera necesario realizar una recodificación de las respuestas para hacer más manejable su tratamiento estadístico. Para ello, se emplea el método de recodificación que el CIS propone y que se puede analizar en el **Anexo 1** y **Anexo 2**. De igual forma, se añade una recodificación especial para la variable de Edad que en un inicio toma la edad de los encuestados y que a través de la recodificación pasa a tomar distintos valores en función de unos determinados tramos de edad (menores de 28, entre 28 y 38, entre 38 y 48, entre 48 y 58, entre 58 y 68 y finalmente, mayores de 68 años). También se llevará a cabo una recodificación especial para aquellas respuestas relacionadas con Unidas Podemos. En todos los casos, ésta supondrá la suma de Podemos y todas sus confluencias (Izquierda Unida, EQUO y En Comú Podem).

## 2.1. Descripción univariante

El estudio tiene como principal objetivo conseguir llevar a cabo una predicción de los datos faltantes en el caso de la variable “intención de voto”. Ésta es la variable con la que se predicen los resultados electorales y por ese motivo adquiere una importancia especial.

- **En las elecciones convocadas para el (día) (mes) ¿a qué partido o coalición piensa votar?**

Esta pregunta es formulada del mismo modo en las encuestas de marzo y abril, sin embargo, existen distintas posibilidades de respuesta en función de los partidos que se presentan. Además también se incluyen las respuestas “no sabe”, “no responde”, “voto nulo”, “voto blanco”, “no votará” y “no lo tiene decidido aún”.

Para facilitar el estudio y su posterior análisis, se van a unir en la respuesta “Otros” aquellas partidos que no son mayoritarios (PP, PSOE, Podemos + confluencias, Ciudadanos y Vox).

De igual forma, se van a considerar como valores faltantes aquellas personas que no respondan a esta pregunta o bien, respondan con la opción “N.S.”, “N.C.”, “No lo tiene decidido aún”, “Voto en blanco” y “Voto nulo”. Se considera que es un mínimo porcentaje de población aquella que vota en blanco o nulo por lo que, deseando facilitar el estudio, no tiene sentido incluir dos posibles respuestas más que lo dificulten (entre el 1,6% y el 1,7% en las tres últimas elecciones celebradas (El País, 2019)).

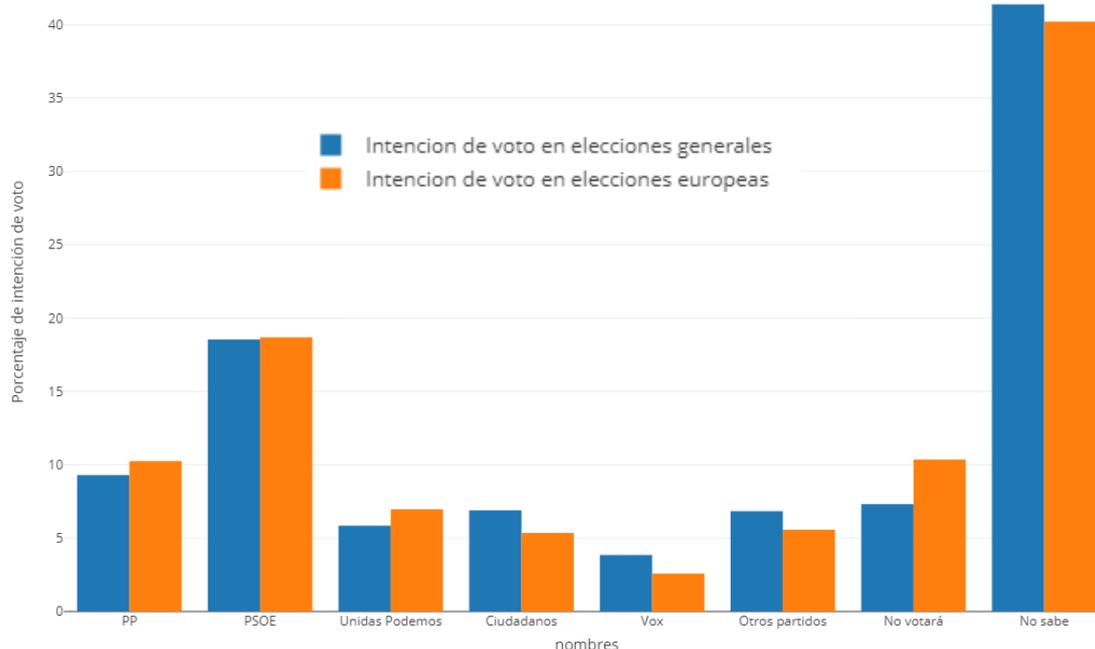


Ilustración 2.1. Comparación de intención de voto

En el gráfico se observa que, aproximadamente, el 40% de los encuestados no contestan a esta pregunta y por tanto, se toman como valores faltantes, que se desean predecir en este estudio.

Además, es necesario observar que la intención de voto por partidos es muy poco variable entre las dos encuestas, algo que se puede achacar al poco tiempo transcurrido entre ambas (tan solo un mes).

También es interesante ver que el PSOE es el partido con mayor intención de voto, seguido del PP. Posteriormente se encuentran Unidas Podemos y Ciudadanos con intenciones de voto similares. Por último se encontraría Vox con una intención algo menor. Este orden de fuerzas concuerda con los resultados de las elecciones generales y europeas (**Anexo 3** y **Anexo 4**). Sin embargo, es necesario mencionar que aunque el orden de fuerzas sea correcto, los porcentajes de intención de voto no lo son. Esto se debe a que los datos faltantes estarán distribuidos por todas las opciones políticas o en la abstención.

- **¿Con que partido simpatiza usted más?**



Ilustración 2.2. Comparación de simpatía de voto

Destaca la gran disminución de entrevistados que contestan “no sabe/ no contesta” (11,30%) respecto a la pregunta de intención de voto (40% aproximadamente). Esto se puede deber a que, a pesar de ser dos preguntas muy relacionadas, ésta no resulta tan evidente y explícita como la anterior y por tanto, puede resultar más sencilla de responder al encuestado. Por este motivo, se podría extraer que los resultados de esta pregunta podrían ser más similares a los resultados de las elecciones. Al analizar dichos resultados en el **Anexo 3** y **Anexo 4**, se puede observar que, ciertamente, los resultados de las elecciones son más similares a los resultados de la simpatía de voto. De hecho, éstos difieren en entre un 3-5% a la baja, algo que hace deducir que aquellos

encuestados que dicen no tener simpatía por ninguno de los partidos políticos se reparten de forma más o menos uniforme alrededor de todo el espectro político.

- **¿A quién votó en las últimas elecciones generales del 26 de junio de 2016?**

Es preciso tomar en cuenta que esta pregunta es la misma tanto para la encuesta que se refiere a las elecciones generales del 28 de abril como a la que se refieren a las elecciones europeas del 26 de mayo. Esto se puede deber a que las últimas elecciones europeas se celebraron hace cinco años y por tanto, el recuerdo de voto puede estar un poco olvidado o incluso, las situaciones personales del encuestado puede haberle hecho variar su ideología o sus razones de voto.

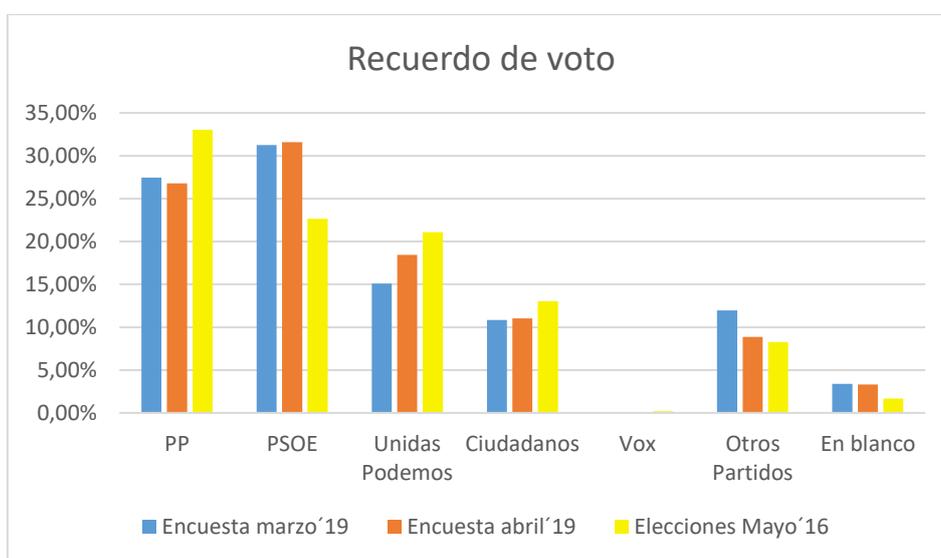


Ilustración 2.3. Comparación de recuerdo de voto

Además, en este caso se toma en cuenta, exclusivamente, aquellas personas que pudieron votar de forma válida. Se puede observar que existe un desajuste entre lo que los encuestados contestan y los resultados en las elecciones de 2016 (El País, 2019), principalmente en el PP y el PSOE. En el caso de los socialistas, su recuerdo de voto es muy superior respecto a los resultados que obtuvieron en dichas elecciones y sin embargo, en el caso del PP ocurre el caso contrario. Esto puede venir explicarse por la teoría de la espiral del silencio de Elisabeth Noelle-Neumann, que indica que los individuos de un grupo social adaptan su comportamiento y sus opiniones a las actitudes predominantes (Carranza, 2012) (Ferrero, s.f.). Esto hace indicar que el hecho de que en el momento en que se realizan las encuestas el partido del Gobierno y de mayor apoyo en otras encuestas sea el PSOE puede generar que los entrevistados indiquen que votaron al PSOE en las elecciones generales de 2016 a pesar de que no fuera así.

También cabe mencionar que el recuerdo de voto a Vox es nulo debido a que este partido era intrascendente durante las elecciones de 2016, cuando tan solo recibió un 0,6% de los votos de los españoles.

- **Cuando se habla de política se utilizan normalmente las expresiones izquierda y derecha. En esta tarjeta hay una serie de casillas que van de izquierda a derecha. ¿En qué casilla se colocaría Ud.? (Entre 1 izquierda y 10 derecha).**

La escala normal que se debería de emplear en esta variable sería de 0 a 10, sin embargo, desde el CIS se evita el 0 ya que, podría otorgarle a la izquierda connotaciones negativas. Esto supone que este indicador esté algo sesgado hacia la izquierda ya que, históricamente, se produce la moda en el valor 5. De ello se puede deducir que se produce porque se cree que éste es el punto medio e implica una ideología moderada. Esto se produce ya que, la población está habituada a que el punto central sea el 5, aunque en este caso no lo sea exactamente (CIS, 2019).

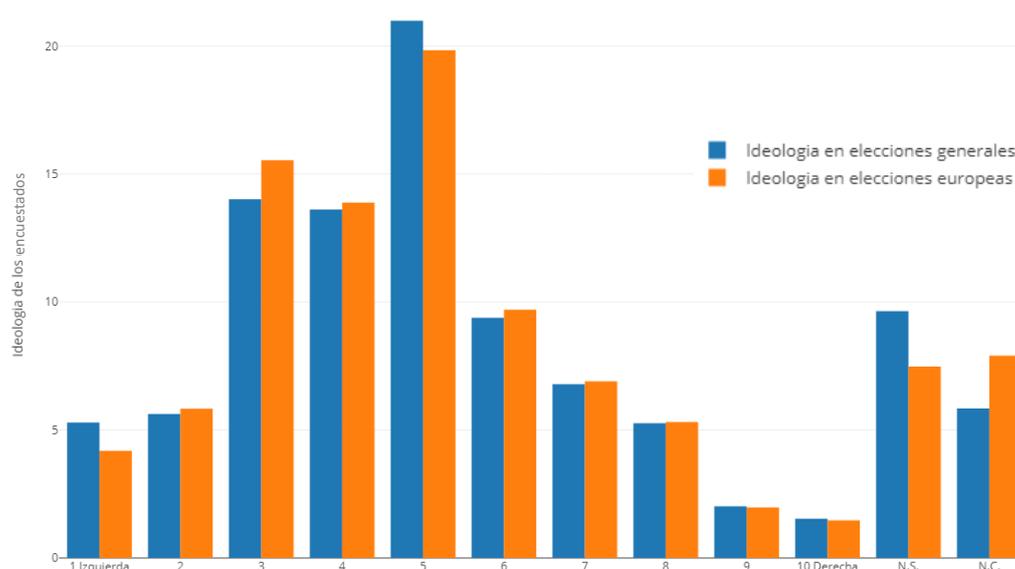


Ilustración 2.4. Comparación de la ideología

Como se menciona anteriormente, existe un sesgo que provoca que la mayoría de gente se auto-ubique en el 5 como centro ideológico. Además también se puede observar que existe una mayor concentración de entrevistados que se ubicarían en la izquierda ideológica.

También es destacable que, tal y como apuntan ciertos estudios, cerca del 70% de los españoles se ubican entre el centro-izquierda y el centro-derecha, algo que ocurre históricamente a pesar de la entrada de nuevos partidos políticos (Guisado & Ley, 2015).

Finalmente, se puede observar como las frecuencias ideológicas varían muy poco entre las dos encuestas estudiadas.

- **¿Cómo calificaría Ud. la situación económica general de España: muy buena, buena, regular, mala o muy mala?**

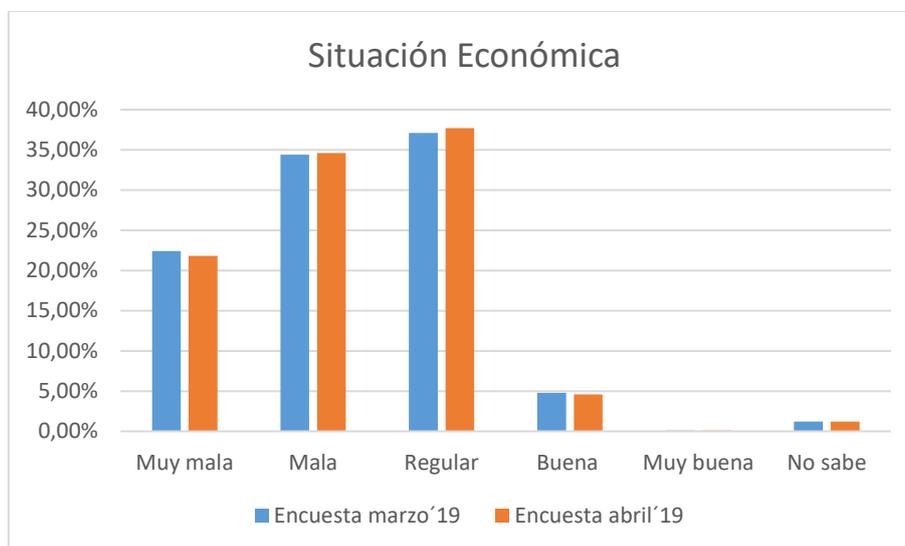


Ilustración 2.5. Comparación de la situación económica

En la mayor parte de los casos, los entrevistados contestan que la situación económica de España es regular seguida de la opción mala y muy mala. Tan solo un 5% de los encuestados indican que la economía española es buena o muy buena.

Además, los resultados son, prácticamente, idénticos para ambas encuestas, algo que viene motivado por el escaso periodo de tiempo que transcurrió entre la realización de éstas.

- **¿Cómo calificaría Ud. la situación política general de España: muy buena, buena, regular, mala o muy mala?**

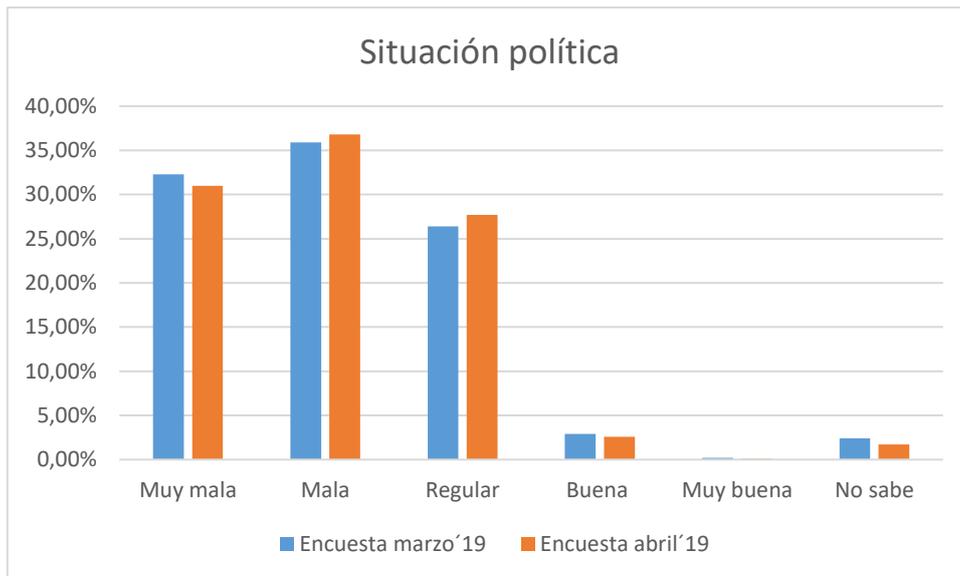


Ilustración 2.6. Comparación de la situación política

La mayor parte de los encuestados consideran que la situación política de España es mala, seguidos de aquellos que la consideran muy mala y, en tercer lugar, regular. Lejos queda el, aproximadamente, 4% que considera que ésta es buena o muy buena (opción que casi ningún encuestado responde).

Esta mala valoración de la política española puede explicarse debido a la inestabilidad que existe a la dificultad de pactos y acuerdos entre los partidos o al gran aumento de la tensión entre personajes políticos que cada vez es más patente.

• **¿Cuántos años cumplió Ud. en su último cumpleaños?**

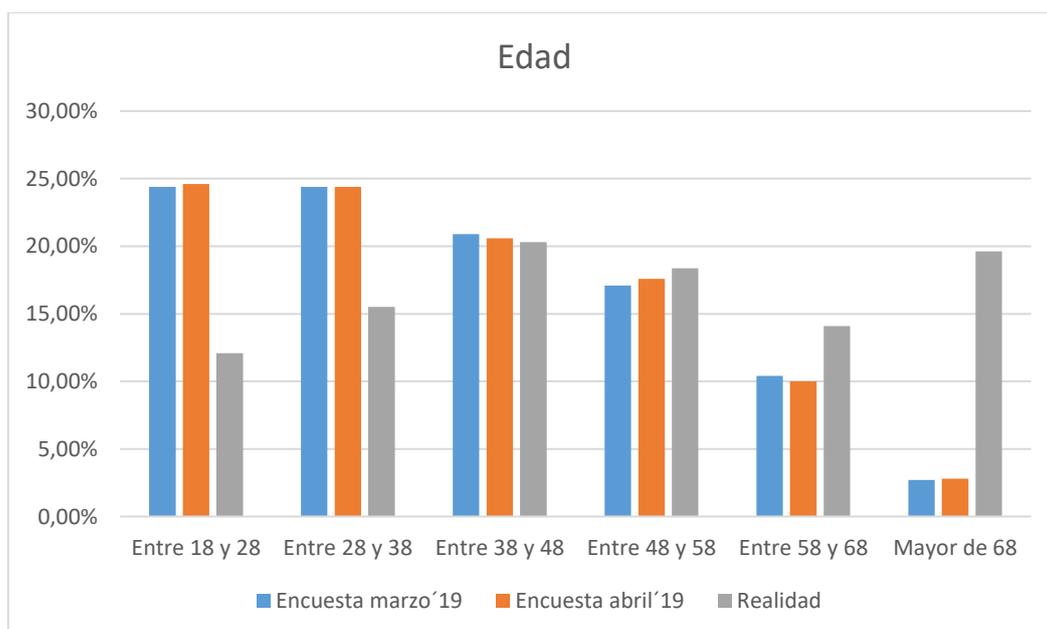


Ilustración 2.7. Comparación de la edad de los encuestados

Puede que uno de los errores que cometa el CIS en sus encuestas radique en la mala elección de encuestados en relación con su edad. Como se puede observar en el gráfico, resulta llamativo el bajo porcentaje de encuestas que se realizan a personas mayores de 68 años en relación al porcentaje de la población que estos representan. Pasa lo contrario con la población más joven a la que, proporcionalmente, se le hacen muchas más encuestas de las que les correspondería.

Esto, sumado a que es la población joven aquella que tiende a abstenerse con mayor frecuencia (un 25% de ellos) (Ley, 2019) hace que, a pesar de que, inicialmente, resulte trascendente hacer uso de esta variable, finalmente **se decida no emplearla** ya que puede generar problemas en los objetivos del estudio.

- **¿En qué situación laboral se encuentra actualmente? ¿En caso de trabajar, cuál es la ocupación o trabajo que tiene Ud. actualmente?**

Siguiendo con estudios realizados, se considera importante tomar en cuenta esta pregunta pues según estos estudios la situación laboral influye en el voto de los votantes, particularmente, cuando el paro aumenta la probabilidad de abstenerse (Caínzos & Voces, 2015).

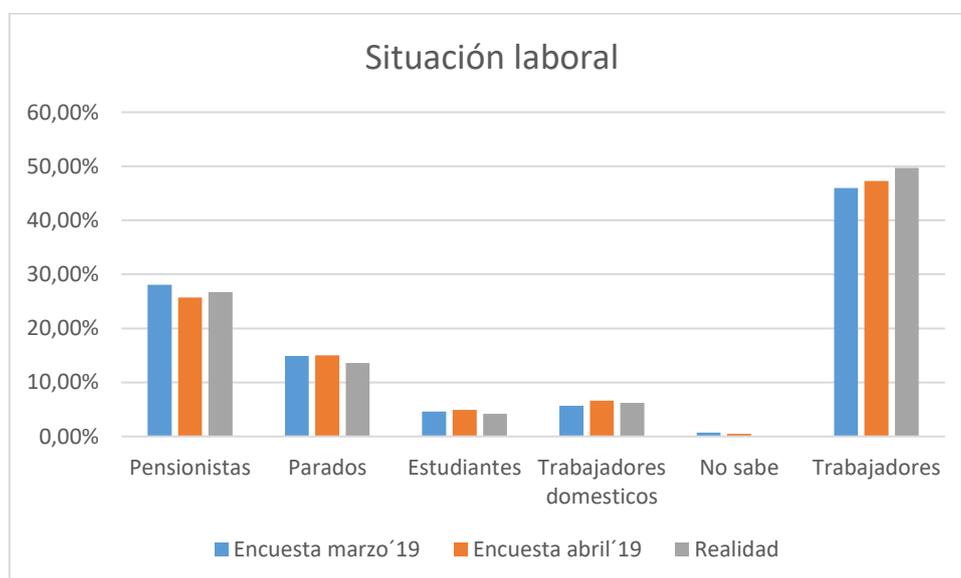


Ilustración 2.8. Comparación de la situación laboral de los encuestados

Aun así, se cree necesario comenzar comparando la realidad de la situación laboral con la de los encuestados. Se observa que, en este aspecto, se ha escogido una buena muestra por parte del CIS, siendo éstas muy similares a la realidad de España. También

se observa que, casi el 50% de los encuestados son trabajadores, seguidos de pensionistas y jubilados quienes forman un grupo de casi el 30%.

- **¿Cuáles son los estudios oficiales de más alto nivel que Ud. ha finalizado?**

Como se observa en el estudio “La influencia del nivel de estudios en el voto” (Electomanía, 2018) existen evidencias que demuestran que el nivel educativo de la población influye sustancialmente en su voto. Por este motivo, se considera absolutamente relevante la inclusión de esta pregunta dentro de nuestro estudio.

Además, siguiendo el estudio mencionado anteriormente, se lleva a cabo una recodificación de variables de modo que sea más sencillo llevar a cabo el estudio. Para ello, se agrupan los niveles educativos en tres: sin estudios o con estudios primarios, con estudios secundarios o formación profesional y finalmente, con estudios superiores.

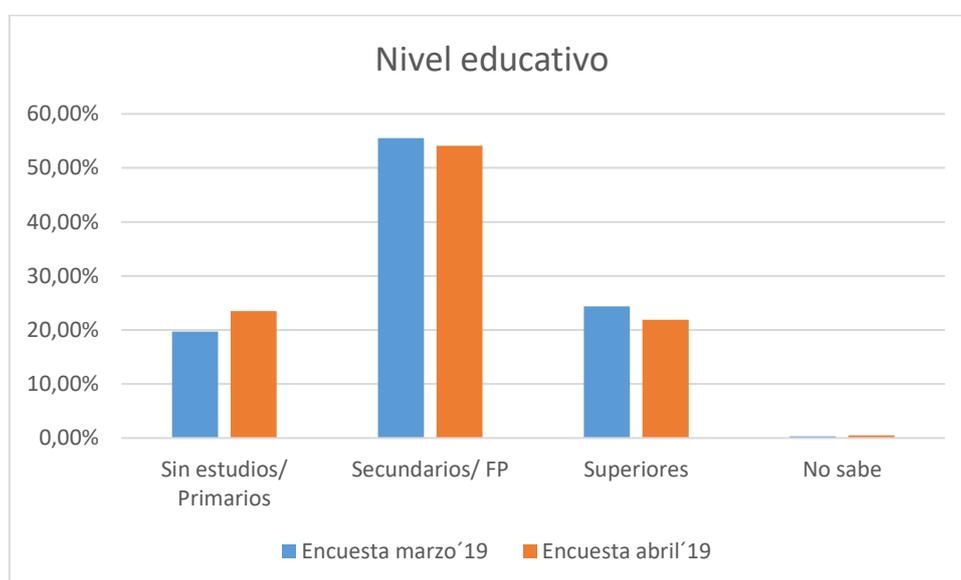


Ilustración 2.9. Comparación del nivel educativo de los encuestados

Se puede observar que más de la mitad de los encuestados tienen un nivel educativo medio y que el porcentaje de entrevistados sin estudios o estudios primarios y con estudios superiores es muy similar.

- **¿Cómo se definiría Ud. en política según la siguiente clasificación?**

Esta es una pregunta que tan solo se plantea en la encuesta de marzo para las elecciones generales. Sin embargo, ésta se puede considerar bastante determinante para predecir la intención de voto y por ello, será empleada en el estudio predictivo de los comicios al Parlamento español del 28 de abril.

En esta pregunta, se ofrecen hasta once opciones de ideología política a elegir. Estas son conservador, demócrata cristiano, liberal, progresista, socialdemócrata, socialista, comunista, nacionalista, feminista, ecologista y otra. También se baraja la opción de que el entrevistado indique que se considera apolítico, sin embargo, ésta no se lee para así evitar un posible sesgo que se pudiera producir en la respuesta.

Se considera oportuno llevar a cabo una recodificación de las respuestas, de modo que se rebaje el número de ellas, ya que, existen demasiadas opciones que podrían afectar en la significatividad de la variable y además, algunas de ellas son demasiado similares. Siguiendo con las explicaciones que ofrece Javier González Serrano, secretario del Colegio Nacional de Doctores y Licenciados en Ciencias Políticas y Sociología, se pasa a concentrar las opciones conservador y demócrata cristiano en una sola, también se concentra liberal y progresista y lo mismo ocurre con socialdemócrata y socialista (Calvo, 2017).

## 2.2. Datos faltantes

Para llevar a cabo este estudio, también resulta necesario analizar las preguntas que no responden algunos encuestados y que el CIS codifica como *NA* o bien, que contestan con respuestas como “no sabe” o “no contesta”.

También se deben de incluir como datos faltantes aquellos que no tienen sentido en relación a la pregunta como por ejemplo, indicar en nivel educativo “guardería” o indicar en ideología el valor 11 cuando la escala es del 1 al 10. Estos datos no suelen aparecer frecuentemente, sin embargo, hay que tomarlos en cuenta ya que, pueden desvirtuar el estudio.

Para comenzar con el análisis de datos faltantes se realiza un gráfico para cada una de las encuestas que se estudian. Estos gráficos harán que se puedan visualizar de forma sencilla los patrones de valores faltantes tomando en cuenta las variables de estudio.

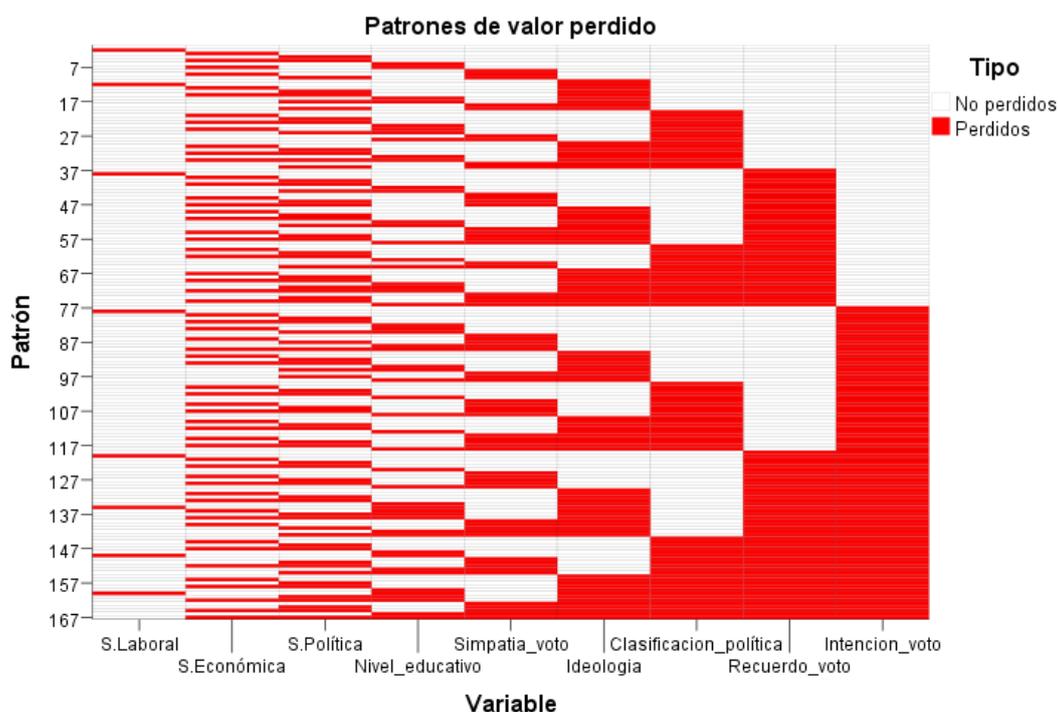


Ilustración 2.10. Patrones de valor perdido de la encuesta de marzo de 2019

Se observa la existencia de 167 patrones distintos y que la variable de situación laboral es la que menos datos faltantes tiene. Por el contrario, también se observa que la intención de voto es la variable que más datos faltantes tiene. Además, a ésta le sigue la variable de recuerdo de voto. Esto resulta bastante obvio ya que, son las preguntas que pueden resultar más comprometedoras para las personas entrevistadas mientras que, la situación laboral o la opinión acerca de la situación económica o política del país, probablemente, resulte más sencillo de responder.

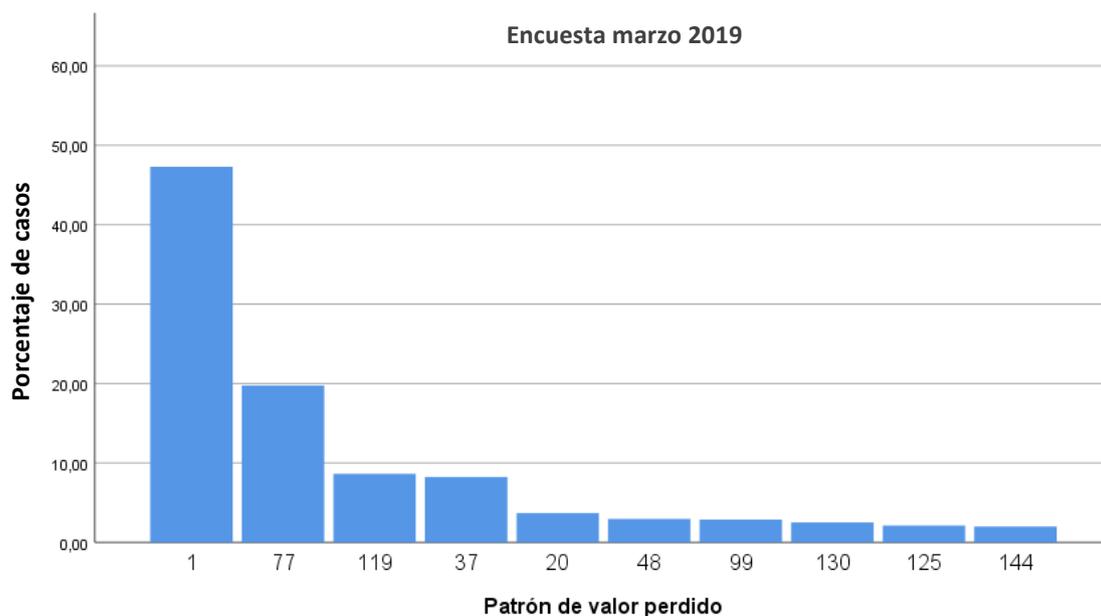


Ilustración 2.11. Patrones de valor perdido con mayor frecuencia en la encuesta de marzo de 2019

Por otro lado, en este otro gráfico se pueden observar cuáles son los patrones más habituales. Destacan por encima de todos cuatro patrones que serían el 1, el 77, el 119 y el 37. Estos representarían aproximadamente un 85% de las entrevistas realizadas.

El patrón 1 representa a casi un 50% de las personas entrevistadas y son aquellas que responden a todas las preguntas y por tanto, nos ofrecen toda la información que se necesita. El patrón 77 representa a un 20% de los entrevistados y lo forman aquellos que no desean responder acerca de su intención de voto. El patrón 119 representa a un 9% de los entrevistados, los que no responden a las preguntas sobre intención de voto y recuerdo de voto. Por último, el patrón 37 representa a un 8% de las personas entrevistadas, aquellas que no contestan al recuerdo de voto.

Se vuelve a realizar el mismo análisis para la encuesta de abril de 2019:

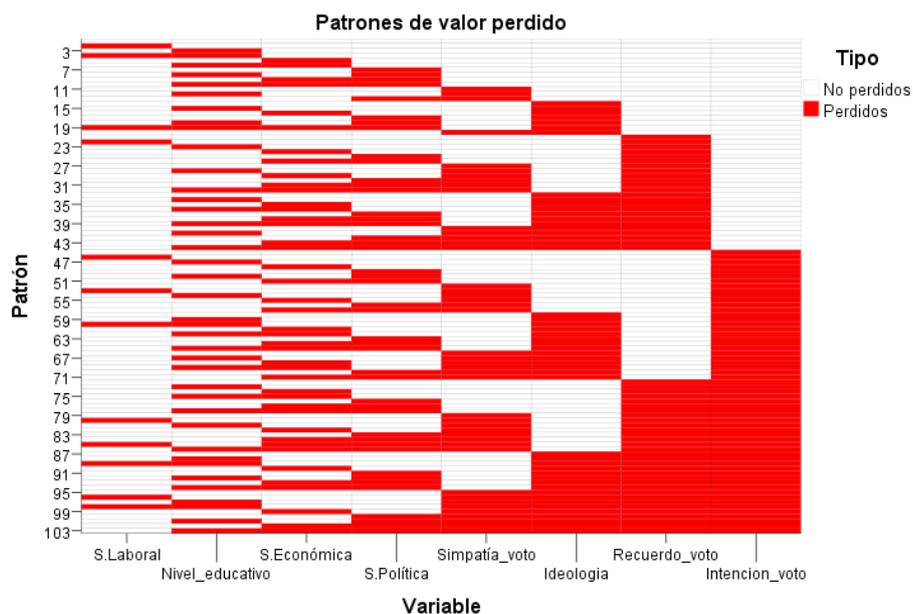


Ilustración 2.12. Patrones de valor perdido de la encuesta de abril de 2019

En este caso se diferencian 103 patrones distintos, valor bastante más bajo que en el caso de la encuesta de marzo, algo que, muy probablemente, se debe a que en esta ocasión se cuenta con una variable menos (clasificación política) debido a que esta pregunta no se formuló en dicha encuesta.

Como ya pasara con la entrevista de marzo, la variable situación laboral es la que menos datos faltantes tiene. Del mismo modo, intención de voto seguido de recuerdo de voto son las variables con más datos faltantes.

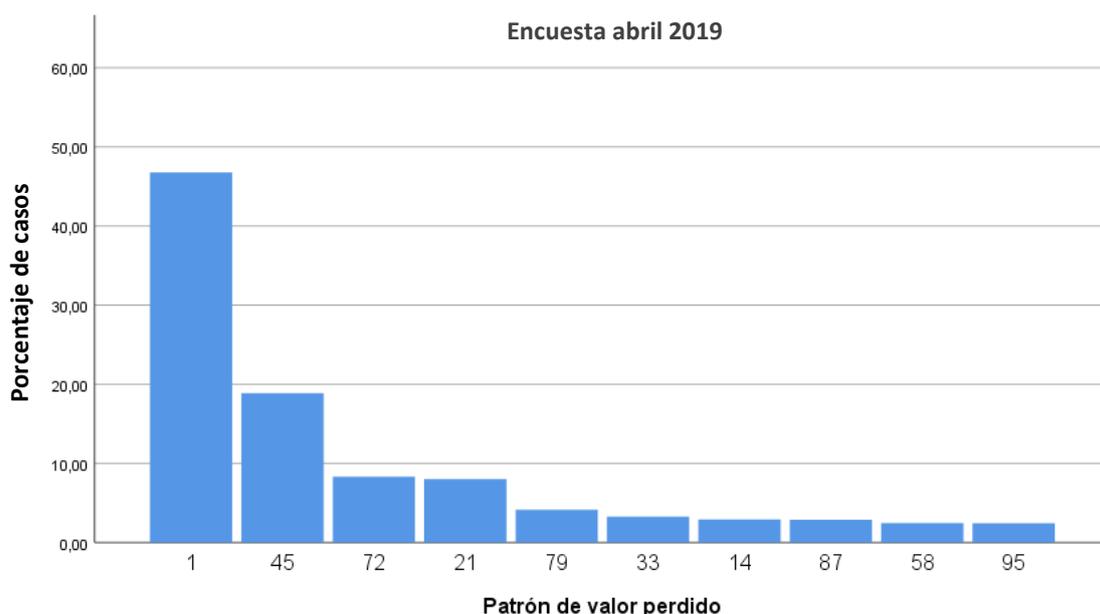


Ilustración 2.13. Patrones de valor perdido con mayor frecuencia en la encuesta de abril de 2019

En el gráfico de frecuencias se observan de nuevo cuatro patrones (1, 45, 72 y 21) que destacan por encima del resto representando aproximadamente un 83% de las respuestas.

Una vez más el patrón 1, representa más de un 45% de entrevistados, los que responden a todas las preguntas. El patrón 45 representa a casi un 20% de entrevistados, aquellos que no responden a su intención de voto. El patrón 72, lo forman, con un 8%, los que no dan información acerca de su intención de voto ni su recuerdo de voto. Finalmente, forma el patrón 21, un 8% de entrevistados que no responden a su recuerdo de voto.

### 3. METODOLOGÍA

El estudio que se desea realizar y su posterior predicción de datos faltantes se elaboran a través de árboles de decisión. Los árboles de decisión son un método analítico que a través de una representación esquemática de las alternativas disponible facilita la toma de mejores decisiones (GEO Tutoriales, 2016) desde un punto de vista probabilístico, ante un abanico de posibles decisiones (Berlanga, et al., 2013). Por medio de estos árboles se pueden diferenciar grupos y relaciones que puedan existir de una forma muy visual.

Para llevar a cabo la creación de los árboles de decisión se puede realizar a través de distintos métodos:

- **Detector automático de interacciones mediante Chi-cuadrado (CHAID):** Consiste en un rápido algoritmo de árbol estadístico y multidireccional que explora datos de forma rápida y eficaz, y crea segmentos y perfiles con respecto al resultado deseado. En cada paso, elige la variable independiente (predictora) que presenta la interacción más fuerte con la variable dependiente. Las categorías de cada predictor se funden si no son significativamente distintas respecto a la variable dependiente.
- **Detector automático de interacciones mediante Chi-cuadrado exhaustivo (CHID exhaustivo):** Supone una modificación de CHAID que examina todas las divisiones posibles para cada predictor y trata todas las variables por igual, independientemente del tipo y el número de categorías.
- **Árboles de clasificación y regresión (CART):** Utiliza un algoritmo de árbol binario completo que hace particiones de los datos y genera subconjuntos precisos y homogéneos.
- **Árbol estadístico, rápido, imparcial y eficaz (QUEST):** Emplea un algoritmo estadístico que selecciona variables sin sesgo y crea árboles binarios precisos de forma rápida y eficaz. (Berlanga, et al., 2013)

Se considera que el mejor método para llevar a cabo el estudio se debe de emplear el método CHAID. Esto se debe a que es necesario crear árboles de varias ramas ya que, las preguntas formuladas a las personas entrevistadas ofrecen multitud de respuestas. Por otro lado, el hecho de que existan bastantes variables hace necesario emplear este algoritmo, que funde las categorías de las variables predictoras respecto a la variable dependiente y que por tanto, no ofrecen demasiada información extra.

De igual forma, se considera necesario desechar la opción de emplear algoritmos CART o QUEST ya que, estos devuelven arboles binarios que generarían una mayor complicación a la hora de analizar los resultados.

### **3.1. Implementación**

Para crear el árbol de decisión siguiendo el algoritmo CHAID, mediante el cual se podrá obtener una predicción de la intención de voto, se comienza eliminando de la base de datos todas aquellas entrevistas que no disponen de respuesta para todas las preguntas que se usarán como regresoras. De este modo, y siguiendo con el gráfico de patrones de valores faltantes, la base de datos se reducirá a aquellas entrevistas que forman el patrón 1, aproximadamente un 45% del total de ellas.

Posteriormente, se asigna la variable intención de voto como variable dependiente o variable a predecir y todas las demás variables como variables independientes (simpatía de voto, recuerdo de voto, ideología, situación económica, situación política, situación laboral, nivel educativo y clasificación política para el caso de la encuesta de marzo).

Del mismo modo, se decide que el nivel de significación para la división de nodos y fusión de categorías sea de 0,05 y que el número máximo de niveles sea 10 para la encuesta de marzo y 9 para la de abril, coincidiendo con el número de variables que se tomarán en cuenta.

Una vez obtenido el árbol de decisión, éste se aplicará a aquellas entrevistas que sigan el patrón 77 de la encuesta de marzo y el patrón 45 de la encuesta de abril. Estos patrones son aquellos en los que los encuestados contestaron a todas las preguntas salvo a la de intención de voto. De este modo, se imputarán las respuestas de cada entrevistado al árbol creado, de modo que se predecirá estadísticamente la intención de voto de estas personas.

Se llevará a cabo la misma acción, de nuevo, para predecir la intención de voto de aquellos entrevistados que decidieron no responder acerca de su recuerdo de voto ni tampoco, obviamente, acerca de su intención de voto. Estas personas forman los patrones 119 en el caso de la encuesta de marzo y el 72 en la encuesta de abril.

Para ello, se comenzará creando, una vez más, el árbol de decisión a partir de las respuestas formuladas por los entrevistados que respondieron a todas las preguntas. Sin embargo, a diferencia del caso anterior, en este árbol no se tomará el recuerdo de voto como variable independiente o regresora ya que, las personas de las cuales se desea predecir la intención de voto, tampoco respondieron al recuerdo de voto. Además,

como en el caso anterior, la significación será de 0,05 y en este caso, el número máximo de niveles será 9 para la encuesta de marzo y 8 para la de abril. Esto se debe a que ese es el número de variables que se tomarán en el estudio.

Una vez obtenido el árbol de decisión, de nuevo, se realiza la imputación de las entrevistas que siguen los patrones 119 y 72 mencionados previamente para así obtener su intención de voto. De esta forma, se consigue disponer de aproximadamente el 85% de los datos con su correspondiente intención de voto.

Durante todo el proceso de obtención de los árboles de procesos mencionados y sus correspondientes predicciones se llevará a cabo un proceso de validación por división muestral de los árboles de decisión. Esto implica que cada uno de los árboles que se obtienen, se hará empleando una submuestra con un 70% de los datos de la muestra correspondientes mientras que, el 30% restante de la muestra se empleará para comprobar la validez de dicho árbol.

Este proceso se considera necesario cuando se desea emplear un árbol de decisión ante una población mayor ya que, permite evaluar la bondad de su estructura (López & González, 2007).

### **3.2. Árboles de decisión**

#### **Árbol para la predicción de los resultados de las elecciones de abril basado en las encuestas de marzo (Anexo 5)**

Este árbol se emplea para predecir la intención de voto de aquellas personas que responden a todas las preguntas exceptuando la propia intención de voto.

Para ello, el factor más influyente es la simpatía de voto. Aproximadamente, el 92% de las personas que declararon simpatía por Ciudadanos o Vox, votarían a dichos partidos. Además, según se observa en el árbol, no será necesario tomar en cuenta más variables para estas personas ya que, existe suficiente significatividad y su intención de voto es suficientemente evidente tomando en cuenta dicho factor.

Por el contrario, son aquellos que declaran su simpatía de voto a Unidas Podemos con los cuales hay que tomar en cuenta más variables para predecir con suficiente significatividad su intención de voto. En el caso de Unidas Podemos será necesario fijarse en la respuesta que den los entrevistados respecto a su ideología. De esta forma, si declaran sentirse en un punto superior al 3, siguiendo la escala 1 (izquierda) – 10 (derecha), tan solo el 68% de ellos, votarán a dicho partido. Por el contrario, si se declaran en un rango de 3 o menos, será necesario fijarse en el recuerdo de voto (69%

si votaron previamente a PSOE u otros partidos) e incluso en el nivel educativo del entrevistado.

Respecto a los que declaran su simpatía por el PSOE, el 94% de ellos les votan, sin embargo, en caso de que en las pasadas elecciones estas personas votasen al PP o a Ciudadanos, la intención de voto disminuirá a un 77%.

En el caso de los que indican tener simpatía por el PP les votan posteriormente el 88% de esas personas. Sin embargo, si tienen una ideología por debajo del valor 5, tan solo el 76,5% votarán al Partido Popular.

El 90% de las personas que tienen simpatía por otros partidos minoritarios decantarán su voto por estos partidos. Sin embargo, en caso de que hubieran votado en el pasado a alguno de los partidos mayoritarios, tan solo un 64% de ellos votarán, realmente, a los minoritarios.

Finalmente, entre aquellos que no tienen simpatía por ningún partido político existe un 24% ellos que votarán al PSOE y un 21,5% que lo harán a Ciudadanos, valores bastante elevados y a tomar en cuenta.

### **Árbol para la predicción de los resultados de las elecciones de abril basado en las encuestas de marzo sin tomar en cuenta el recuerdo de voto (Anexo 6)**

Este árbol se emplea para predecir la intención de voto de aquellas personas que responden a todas las preguntas exceptuando la propia intención de voto y la de recuerdo de voto.

Como pasaba con el anterior árbol, el factor simpatía de voto es el más significativo. El 92% de los que indiquen tener simpatía por Vox, votarán por este partido.

En el caso de los que indican tener simpatía por Ciudadanos también existe mucha significatividad, siendo el 91% de ellos votantes del propio partido. Sin embargo, si declaran ser pensionistas, serán un 10% menos de dicho valor.

En relación a los que declaran simpatía al PP, el 88% de ellos votan a dicho partido. Sin embargo, si su ideología es inferior o igual que 5 el dato desciende en un 15%.

Entre los que indican sentir simpatía por Unidas Podemos, un 81% de ellos votarán al partido morado. Sin embargo, en caso de que los entrevistados indiquen que su ideología está por encima del valor 3, tan solo el 68% votará por dicho partido. Sin

embargo, el porcentaje de personas que votarán al PSOE en esas condiciones aumentará hasta el 22,5%, valor muy reseñable y a tomar en cuenta.

Si consideran simpatía por el PSOE, el 94% de los ellos votarán a los socialistas, siendo este el porcentaje más elevado y que más seguridad ofrece. Sin embargo, este porcentaje se rebajará en 11 puntos porcentuales entre los entrevistados que se consideran liberales y con una ideología superior al valor 4.

Por último, el 21,5% de las personas que indican no tener simpatía por ningún partido, votarán a Ciudadanos y otro 20,5% votará al PSOE. Sin embargo, este valor ascenderá hasta el 42% para los que consideren su ideología por debajo del valor 4.

### **Árbol para la predicción de los resultados de las elecciones de mayo basado en las encuestas de abril (Anexo 7)**

Este árbol se emplea para predecir la intención de voto de aquellas personas que responden a todas las preguntas exceptuando la propia intención de voto.

Al igual que ocurre con los dos árboles analizados previamente, la simpatía de voto resulta ser el factor más determinante y significativo de entre las variables independientes, si bien, lo es en menor medida.

El 91% de las personas que declaran simpatía por el PSOE votarán a dicho partido. Sin embargo, esta cantidad disminuye en caso de que en anteriores elecciones hubieran votado a Unidas Podemos (hasta un 77%). De igual forma, tan solo un 58,5% de las personas votarían al PSOE en caso de que en elecciones previas no hubieran votado ni a Unidas Podemos ni al PSOE y además su ideología fuera superior al factor 4.

El 80% de votantes que declaran tener simpatía por Unidas Podemos, votan a este partido. Sin embargo, solo el 55,5% de ellos lo harían en caso de que en el pasado no se hubieran decantado por votar a este partido. Sin embargo, un 74% de ellos votarían por Unidas Podemos en caso de que su ideología fuera superior al nivel 3.

Por otro lado, el 76% de personas que declararon simpatía por Ciudadanos, se decantaron por esta opción de voto, dato muy inferior respecto a los árboles creados para las elecciones generales de abril. Además, este dato disminuye aún más (hasta un

64% - 66%) entre los entrevistados que declarasen haber votado a otro partido que no fuera Ciudadanos en las elecciones pasadas.

El 88% de personas que declararon tener simpatía por el PP, eligieron a los populares en su voto. De nuevo, este dato desciende en 19 puntos entre los que en las pasadas elecciones no votaron al PP o en blanco. Además, también aumenta hasta el 92% entre aquellos que votaron en el pasado al PP o en blanco y que no se encuentran parados y no tienen estudios o son primarios.

Tan solo un 67% de simpatizantes de Vox, votaron a dicho partido, dato muy inferior también con respecto a los árboles creados para las elecciones generales. De igual forma, si en el pasado votaron a Ciudadanos o PSOE este porcentaje disminuye en 5 puntos más.

Si declaran simpatía por otros partidos minoritarios, el 78,5% de ellos se decantan por estos partidos, dato que disminuye entre un 35,5% y un 29% si hubieran votado a partidos mayoritarios en pasadas elecciones. Por el contrario, un 94,5% de aquellos que declarasen haber votado a otros partidos en elecciones pasadas y considerasen que la situación política es buena o muy buena seguirán votando a dichos partidos minoritarios.

Finalmente, el 32% de las personas que declaran no tener simpatía por ningún partido político, no votará. Sin embargo, si en anteriores elecciones votaron al PSOE, volverán a hacerlo el 54% de ellos y si votaron a Ciudadanos o en blanco, volverán a votar a Ciudadanos el 44,5%.

## **Árbol para la predicción de los resultados de las elecciones de mayo basado en las encuestas de abril marzo sin tomar en cuenta el recuerdo de voto (Anexo 8)**

Finalmente, este último árbol se emplea para predecir la intención de voto de aquellas personas que responden a todas las preguntas exceptuando la propia intención de voto y su recuerdo de voto.

Una vez más, la simpatía de voto resulta ser el factor más determinante y significativo de entre las variables independientes.

Entre aquellos que declaran tener simpatía por el PP, el 88% se decantarán por votar a dicho partido. Sin embargo, si su ideología se encuentra en el nivel 5 o por debajo tan solo lo hará el 77%. En caso de que se encuentre por encima del 5, si los entrevistados no tienen estudios, votarán al PP el 92,5%. Sin embargo, si tiene estudios de secundaria o formación profesional y consideran que la situación política no es muy buena desciende hasta el 81%.

Si indican tener simpatía por el PSOE, votarán a este partido el 90,5% de las personas. Este valor disminuirá al 85% para aquellos que declaren su ideología por encima del valor 4 y por el contrario, alcanzará su cota más alta (93,5%) entre aquellos que se declaren entre el 2 y el 3 y no tengan estudios superiores.

El 81% de simpatizantes de Unidas Podemos votarán a dicho partido, llegando al 87% entre aquellos que tienen estudios superiores.

El 76% de entrevistados que indican tener simpatía por Ciudadanos votarán a este partido y el 69% de los que tengan simpatía por Vox se decantarán por el partido verde.

El 76,5% de las personas que declaran tener simpatía por otros partidos minoritarios votarán por ellos. Sin embargo, este porcentaje alcanzará su mínimo (60%) si indican encontrarse ideológicamente por debajo del nivel 4 y consideran que la situación política del país es regular, mala o muy mala. Por el contrario, alcanzará su nivel máximo si ideológicamente se declaran por debajo del nivel 4 pero valoran la situación política del país como muy buena.

Finalmente, entre aquellos que no declaren simpatía por ninguno de los partidos, el 29% de ellos no votará. Sin embargo, si indican que su ideología se encuentra por debajo del nivel 5 votarán el 39,5% al PSOE y si se encuentra por encima de dicho nivel votarán al PP el 34,5%.

## 4. RESULTADOS

Tras eliminar los registros de las personas que indican o que se prevé (según el estudio realizado) que no acudan a votar los resultados obtenidos para las elecciones generales de abril de 2019 son los siguientes:

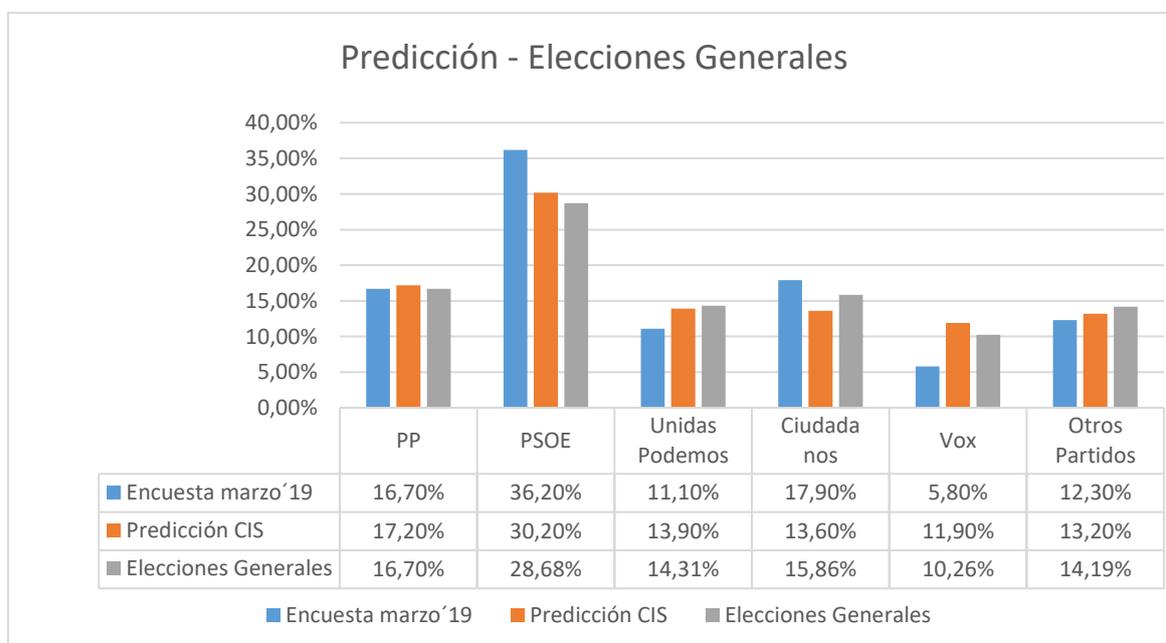


Ilustración 4.1. Comparación de los resultados de la predicción respecto a la predicción del CIS y a los resultados de las elecciones del 28 de abril

A niveles generales se puede observar que las tendencias son muy similares, si bien, en algunos casos existen más diferencias que en otros. Cabe destacar como positiva la predicción para el Partido Popular, que coincide absolutamente con los resultados electorales.

Por otro lado, la peor predicción ocurre con el PSOE. A pesar de que los resultados del estudio le declaran ganador de las elecciones, algo que ocurrió, lo cierto es que el porcentaje de voto fue un 7,5% inferior a lo que se podría prever según el método empleado.

Respecto a Unidas Podemos, Ciudadanos y otros partidos, se puede observar como la predicción también se ajusta mucho a la realidad difiriéndose en un 3% y un 2% respectivamente.

Finalmente, para Vox la predicción es inferior a los resultados que se dieron en las elecciones, existiendo una diferencia del 4,5%.

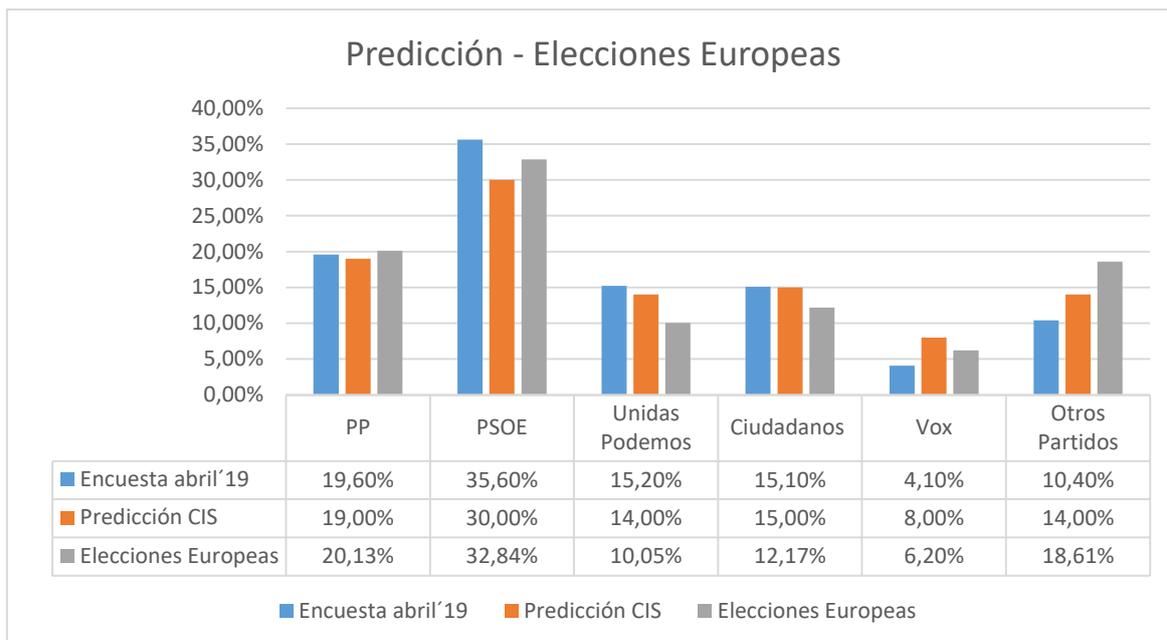


Ilustración 4.2. Comparación de los resultados de la predicción respecto a la predicción del CIS y a los resultados de las elecciones del 26 de mayo

Respecto a la predicción llevada a cabo de cara a las elecciones europeas con ayuda de la encuesta de abril, se puede observar de nuevo que los resultados del estudio también son bastante cercanos a los que se dieron en los comicios.

De nuevo, se observa que la predicción para el PP es muy buena, difiriendo sus resultados tan solo en un 0,5%. En el caso del PSOE se puede observar que la predicción es bastante ajustada, existiendo una diferencia de un 2,8%, dato muy alejado de los 7,5 puntos de diferencia que se observaban para el caso de las elecciones generales. Para Ciudadanos y Vox se observa que, de nuevo, sus resultados son bien predichos, encontrando una diferencia de un 3% y un 2%, respectivamente. Sin embargo, estos resultados no son tan buenos para el caso de Unidas Podemos y otros partidos en los cuales se observan diferencias del 5% y el 8%, valores, sin duda, más abultados.

En líneas generales, se podrían observar como el bloque de derechas (PP, Ciudadanos y Vox) y el bloque de izquierdas y partidos regionalistas, nacionalistas e independentistas (PSOE, Unidas Podemos y otros partidos) se autoajustan bastante bien. Esto significa que las variaciones que existen entre los resultados del estudio están bastante ajustados a la realidad electoral y por tanto, se puede confirmar la influencia de las variables empleadas en el estudio para predecir la intención de voto.

## 5. CONCLUSIÓN

Con este análisis, se ha podido observar cómo, a través de las preguntas que el CIS realiza, se puede predecir la intención de voto de la población a pesar de que en muchas ocasiones, los entrevistados no quieren responder sobre este asunto. Para ello, sin duda alguna, saber la simpatía de voto de las personas puede resultar esencial. Sin embargo, no solo esta cuestión es importante, sino que el recuerdo de voto, la ideología o el nivel educativo, entre otras, pueden resultar indicativos para llevar a cabo esta predicción. Por el contrario, también es necesario indicar que la clasificación política es la única pregunta que se ha tenido en cuenta en el estudio y que, sin embargo, no se ha tomado en cuenta según los árboles de decisión creados. Esto se debe a que tiene una significatividad inferior. También puede deberse a que se han tomado en cuenta otras variables muy similares como ideología que habrían podido explicar de mejor forma lo que el estudio necesitaba.

Se ha podido visualizar que los votantes de algunos partidos son más reacios a contestar su intención de voto que los de otros. De esta forma, aproximadamente la mitad de los votantes de Vox no dicen serlo y por el contrario, existe una alta proporción de votantes que dicen votar al PSOE o a Ciudadanos y que finalmente, no lo realizan. Este es un factor que resulta absolutamente esencial a la hora de predecir los resultados electorales y por tanto, demuestra la necesidad de hacer uso de la famosa cocina electoral que actualmente, supone tantas disputas y discusiones.

Además, resulta llamativo el hecho de que aquellos que no sienten simpatía por ninguno de los partidos tienen una probabilidad mayor de no votar ya que, parece que no se sienten identificados con ninguno de los partidos del espectro político. Sin embargo, hay una gran proporción de estas personas que, finalmente, sí que votan y normalmente lo hacen por Ciudadanos o PSOE.

Por otro lado, también resulta muy necesario mencionar el error que comete el CIS al no tomar en cuenta la edad de los entrevistados para llevar a cabo el estudio. Este organismo entrevista a una cantidad demasiado elevada de personas jóvenes teniendo en cuenta que según los rangos de edades, representan a una minoría de la población española y que, además, son el rango de población más abstencionista. Por el contrario, el CIS toma en cuenta mínimamente a las personas de tercera edad cuando representan un alto porcentaje de la sociedad española y además, tienden a votar (Noriega, 2019). Esto puede suponer un problema grave a la hora de llevar a cabo predicciones como las elaboradas en este estudio. Por este motivo, en caso de considerar demasiado costoso el hecho de tomar en cuenta los rangos de edad para hacer estudios

proporcionales a la edad, deben de considerar el empleo de ponderaciones según dichos rangos además de los que ya emplean según Comunidades Autónomas para paliar los problemas que genera la Ley d'Hont en las predicciones.

Finalmente, cabe destacar que el estudio ha cumplido con los objetivos propuestos y por tanto, se ha obtenido una predicción que se puede considerar bastante acertada de los resultados que se dieron en las pasadas elecciones generales del 28 de abril de 2019 y en las pasadas elecciones europeas del 26 de mayo del mismo año. Esto hace, que pueda resultar interesante seguir ahondando en los árboles de decisión como método para llevar a cabo la predicción de resultados electorales. Para ello, sería muy conveniente que en el futuro se tomasen en cuenta ciertas circunstancias sociológicas que afectan a los resultados. Una de estas podría ser el hecho de que algunos tipos de votantes sean más propensos a no decir la verdad acerca de su intención de voto que otros por factores como la influencia de la opinión pública y actitudes generalizadas en la sociedad.

## 6. REFERENCIAS

Berlanga, V., Rubio, M. J. & Vilá, R., 2013. Cómo aplicar árboles de decisión en SPSS. *Revista d'Innovació i Reserca en Educació*, 8 Enero, 6(1), pp. 65-68.

Caínzos, M. & Voces, C., 2015. Paro y voto: ¿Afecta al voto la experiencia de desempleo?. *Revista de Estudios Políticos*, Issue 168, pp. 115-150.

Calvo, E., 2017. ABC. [En línea]

Available at: [https://www.abc.es/espana/abci-ideologias-politicas-espana-progresista-201702091928\\_noticia.html](https://www.abc.es/espana/abci-ideologias-politicas-espana-progresista-201702091928_noticia.html)

[Último acceso: 18 Julio 2019].

Carranza, O., 2012. *Teoría de la espiral del silencio*, El Salvador: Ideaworks.

CIS, 2017. *Centro de Investigaciones Sociológicas*. [En línea]

Available at:

[http://www.cis.es/cis/opencms/ES/1\\_encuestas/ComoSeHacen/pasosencuesta.html#pasos2](http://www.cis.es/cis/opencms/ES/1_encuestas/ComoSeHacen/pasosencuesta.html#pasos2)

[Último acceso: 28 Junio 2019].

CIS, 2019. *Centro de Investigaciones Sociológicas*. [En línea]

Available at: [http://www.cis.es/cis/opencms/ES/11\\_barometros/metodologia.html](http://www.cis.es/cis/opencms/ES/11_barometros/metodologia.html)

[Último acceso: 3 Julio 2019].

El País, 2019. *El País*. [En línea]

Available at: <https://resultados.elpais.com/elecciones/2016/generales/congreso/>

[Último acceso: 30 Junio 2019].

Electomanía, 2018. *electomania*. [En línea]

Available at: <https://electomania.es/cisestudios/>

[Último acceso: 9 Julio 2019].

Ferrero, R., s.f. *Máxima Formación*. [En línea]

[Último acceso: 5 Julio 2019].

GEO Tutoriales, 2016. *Gestión de Operaciones*. [En línea]

Available at: <https://www.gestiondeoperaciones.net/procesos/arbol-de-decision/>

[Último acceso: 9 Julio 2019].

Guisado, P. & Ley, M., 2015. *El Mundo*. [En línea]

Available at:

<https://www.elmundo.es/grafico/espana/2015/12/20/5669a90346163f741f8b4587.html>

[Último acceso: 4 Julio 2019].

Ley, M., 2019. *El Mundo*. [En línea]

Available at: <https://www.elmundo.es/espana/2019/04/27/5cc41a50fdddffe0318b4620.html>

[Último acceso: 5 Julio 2019].

López, C. P. & González, D. S., 2007. Capítulo 18: Clústers y árboles de decisión con SPSS y SAS.

En: *Minería de datos. Técnicas y herramientas*. Madrid: Thomson, p. 691.

Noriega, D., 2019. *El Diario*. [En línea]

Available at: [https://www.eldiario.es/politica/tendencia-universal-Espana-participacion-llegaria\\_0\\_876662426.html](https://www.eldiario.es/politica/tendencia-universal-Espana-participacion-llegaria_0_876662426.html)

[Último acceso: 20 Julio 2019].

## 7. ANEXOS

### Anexo 1. Recodificación encuesta marzo

P.10 (Intención de voto)		Recodificación intención de voto	
PP	1	PP	1
PSOE	2	PSOE	2
Podemos	3	Unidas Podemos <sup>1</sup>	3
Ciudadanos	4	Ciudadanos	4
IU	5	Vox	5
VOX	18	Otros Partidos	6
PACMA	17	No votará	97
En Comú Podem	6		
Compromís	7		
ERC	8		
PDeCAT (JxCat)	9		
En Marea	10		
EAJ-PNV	11		
EH Bildu	12		
CC	13		
UPN	14		
FAC (Foro Asturias)	15		
Nueva Canarias	16		
CUP	19		
Los Verdes	20		
Unidos Podemos	21		
EQUO	22		
Voto blanco	96		
Voto nulo	77		
No votará	97		
No lo tiene decidido aún	94		
N.S.	98		
N.C.	99		

<sup>1</sup> Unidas Podemos está formada por la coalición de Podemos, EQUO, IU y En Comú Podem.

P.14 (Simpatía de voto)		Recodificación simpatía de voto	
PP	1	PP	1
PSOE	2	PSOE	2
Podemos	3	Unidas Podemos <sup>2</sup>	3
Ciudadanos	4	Ciudadanos	4
IU	5	Vox	5
VOX	18	Otros Partidos	6
PACMA	17	Ninguno	97
En Comú Podem	6		
Compromís	7		
ERC	8		
PDeCAT (JxCat)	9		
En Marea	10		
EAJ-PNV	11		
EH Bildu	12		
CC	13		
UPN	14		
FAC (Foro Asturias)	15		
Nueva Canarias	16		
CUP	19		
Los Verdes	20		
Unidos Podemos	21		
EQUO	22		
Ninguno	97		
N.S.	98		
N.C.	99		

<sup>2</sup> Unidas Podemos está formada por la coalición de Podemos, EQUO, IU y En Comú Podem.

P.18AR (Reuerdo de voto)		Recodificación reuerdo de voto	
PP	1	PP	1
PSOE	2	PSOE	2
Unidos Podemos	21	Unidos Podemos <sup>3</sup>	3
Ciudadanos	4	Ciudadanos	4
En Comú Podem	6	Vox	5
Compromís-Podemos-EUPV	37	Otros Partidos	6
ERC	8	En blanco/ nulo	96
Convergència (CDC)	36		
En Marea	10		
EAJ-PNV	11		
EH Bildu	12		
CC-PNC	13		
VOX	18		
En blanco	96		
Voto nulo	77		
No recuerda	98		
N.C.	99		

P.3 y P.4 (Situación económica y política)		Recodificación situación económica y política	
Muy buena	1	Muy buena	1
Buena	2	Buena	2
Regular	3	Regular	3
Mala	4	Mala	4
Muy mala	5	Muy mala	5
N.S.	8		
N.C.	9		

P.25 (Situación laboral)		Recodificación situación laboral	
Trabaja	1	Trabaja	1
Jubilado/a o pensionista (Anteriormente ha trabajado)	2	Pensionista	2
Pensionista (Anteriormente no ha trabajado, sus labores, etc.)	3	Parado	3
Parado/a y ha trabajado antes	4	Estudiante	4
Parado/a y busca si primer empleo	5	Trabajo doméstico no remunerado	5
Estudiante	6	Otra situación	6
Trabajo doméstico no remunerado	7		
Otra situación	8		
N.C.	9		

<sup>3</sup> Unidos Podemos estuvo formado por la coalición de Podemos, EQUO, IU, En Comú Podem, Compromís y En Marea.

P.24 (Nivel educativo)		Recodificación nivel educativo	
N.P.	0	Sin estudios/ Primarios <sup>4</sup>	1
Menos de 5 años de escolarización	1	Secundarios/ Formación Profesional <sup>5</sup>	2
Primaria (enseñanza primaria o hasta 5º de EGB)	2	Superiores <sup>6</sup>	3
Formación Profesional inicial	3	No sabe	4
Secundaria (Bachillerato Elemental, EGB completa, ESO completa)	4		
Formación Profesional de grado medio (FP I)	5		
Bachillerato/ COU/ PREU (Bachillerato superior, BUP, bachillerato)	6		
Formación Profesional de grado superior (FP II)	7		
Universitarios medios (diplomatura, arquitectura o ingeniería técnica)	8		
Universitarios superiores (licenciatura, grado, máster oficial, doctorado)	9		
Otros estudios no reglados	10		
N.C.	99		

<sup>4</sup> Formada por respuestas N.P., Menos de 5 años de escolarización y Primaria (enseñanza primaria o hasta 5º de EGB).

<sup>5</sup> Formada por respuestas Formación Profesional inicial, Secundaria (Bachillerato Elemental, EGB completa, ESO completa), Formación Profesional de grado medio (FP I), Bachillerato/ COU/ PREU (Bachillerato superior, BUP, bachillerato) y Formación Profesional de grado superior (FP II).

<sup>6</sup> Formada por respuestas Universitarios medios (diplomatura, arquitectura o ingeniería técnica) y Universitarios superiores (licenciatura, grado, máster oficial, doctorado).

P.21 (Clasificación política)		Recodificación clasificación política	
Conservador/a	1	Derechas	1
Demócrata cristiano/a	2	Centro	2
Liberal	3	Izquierdas	3
Progresista	4	Comunista	4
Socialdemócrata	5	Nacionalista	5
Socialista	6	Apolítico	97
Comunista	7	Otras	6
Nacionalista	8		
Feminista	9		
Ecologista	10		
Otras respuestas	11		
(NO LEER) Apolítico	97		
N.S.	98		
N.C.	99		

## Anexo 2. Recodificación encuesta abril

P.10 (Intención de voto)		Recodificación intención de voto	
PP	1	PP	1
PSOE	2	PSOE	2
Podemos	3	Unidas Podemos <sup>7</sup>	3
Ciudadanos	4	Ciudadanos	4
IU	5	Vox	5
VOX	18	Otros Partidos	6
PACMA	17	No votará	97
En Comú Podem	6		
Compromís	7		
ERC	8		
PDeCAT (JxCat)	9		
En Marea	10		
EAJ-PNV	11		
EH Bildu	12		
CC	13		
UPN	14		
FAC (Foro Asturias)	15		
Nueva Canarias	16		
CUP	19		
Los Verdes	20		
Unidos Podemos	21		
EQUO	22		
Voto blanco	96		
Voto nulo	77		
No votará	97		
No lo tiene decidido aún	94		
N.S.	98		
N.C.	99		

<sup>7</sup> Unidas Podemos está formada por la coalición de Podemos, EQUO, IU y En Comú Podem.

P.10 (Simpatía de voto)		Recodificación simpatía de voto	
PP	1	PP	1
PSOE	2	PSOE	2
Podemos	3	Unidas Podemos <sup>8</sup>	3
Ciudadanos	4	Ciudadanos	4
IU	5	Vox	5
VOX	18	Otros Partidos	6
PACMA	17	Ninguno	97
En Comú Podem	6		
Compromís	7		
ERC	8		
PDeCAT (JxCat)	9		
En Marea	10		
EAJ-PNV	11		
EH Bildu	12		
CC	13		
UPN	14		
FAC (Foro Asturias)	15		
Nueva Canarias	16		
CUP	19		
Los Verdes	20		
Unidos Podemos	21		
EQUO	22		
Ninguno	97		
N.S.	98		
N.C.	99		

<sup>8</sup> Unidas Podemos está formada por la coalición de Podemos, EQUO, IU y En Comú Podem.

P.14AR (Reuerdo de voto)		Recodificación reuerdo de voto	
PP	1	PP	1
PSOE	2	PSOE	2
Unidos Podemos	21	Unidos Podemos <sup>9</sup>	3
Ciudadanos	4	Ciudadanos	4
En Comú Podem	6	Vox	5
Compromís-Podemos-EUPV	37	Otros Partidos	6
ERC	8	En blanco/ nulo	96
Convergència (CDC)	36		
En Marea	10		
EAJ-PNV	11		
EH Bildu	12		
CC-PNC	13		
VOX	18		
En blanco	96		
Voto nulo	77		
No recuerda	98		
N.C.	99		

P.2 y P.3 (Situación económica y política)		Recodificación situación económica y política	
Muy buena	1	Muy buena	1
Buena	2	Buena	2
Regular	3	Regular	3
Mala	4	Mala	4
Muy mala	5	Muy mala	5
N.S.	8		
N.C.	9		

P.33 (Situación laboral)		Recodificación situación laboral	
Trabaja	1	Trabaja	1
Jubilado/a o pensionista (Anteriormente ha trabajado)	2	Pensionista	2
Pensionista (Anteriormente no ha trabajado, sus labores, etc.)	3	Parado	3
Parado/a y ha trabajado antes	4	Estudiante	4
Parado/a y busca si primer empleo	5	Trabajo doméstico no remunerado	5
Estudiante	6	Otra situación	6
Trabajo doméstico no remunerado	7		
Otra situación	8		
N.C.	9		

<sup>9</sup> Unidos Podemos estuvo formado por la coalición de Podemos, EQUO, IU, En Comú Podem, Compromís y En Marea.

P.32A (Nivel educativo)		Recodificación nivel educativo	
N.P.	0	Sin estudios/ Primarios <sup>10</sup>	1
Menos de 5 años de escolarización	1	Secundarios/ Formación Profesional <sup>11</sup>	2
Primaria (enseñanza primaria o hasta 5º de EGB)	2	Superiores <sup>12</sup>	3
Formación Profesional inicial	3	No sabe	4
Secundaria (Bachillerato Elemental, EGB completa, ESO completa)	4		
Formación Profesional de grado medio (FP I)	5		
Bachillerato/ COU/ PREU (Bachillerato superior, BUP, bachillerato)	6		
Formación Profesional de grado superior (FP II)	7		
Universitarios medios (diplomatura, arquitectura o ingeniería técnica)	8		
Universitarios superiores (licenciatura, grado, máster oficial, doctorado)	9		
Otros estudios no reglados	10		
N.C.	99		

<sup>10</sup> Formada por respuestas N.P., Menos de 5 años de escolarización y Primaria (enseñanza primaria o hasta 5º de EGB).

<sup>11</sup> Formada por respuestas Formación Profesional inicial, Secundaria (Bachillerato Elemental, EGB completa, ESO completa), Formación Profesional de grado medio (FP I), Bachillerato/ COU/ PREU (Bachillerato superior, BUP, bachillerato) y Formación Profesional de grado superior (FP II).

<sup>12</sup> Formada por respuestas Universitarios medios (diplomatura, arquitectura o ingeniería técnica) y Universitarios superiores (licenciatura, grado, máster oficial, doctorado).

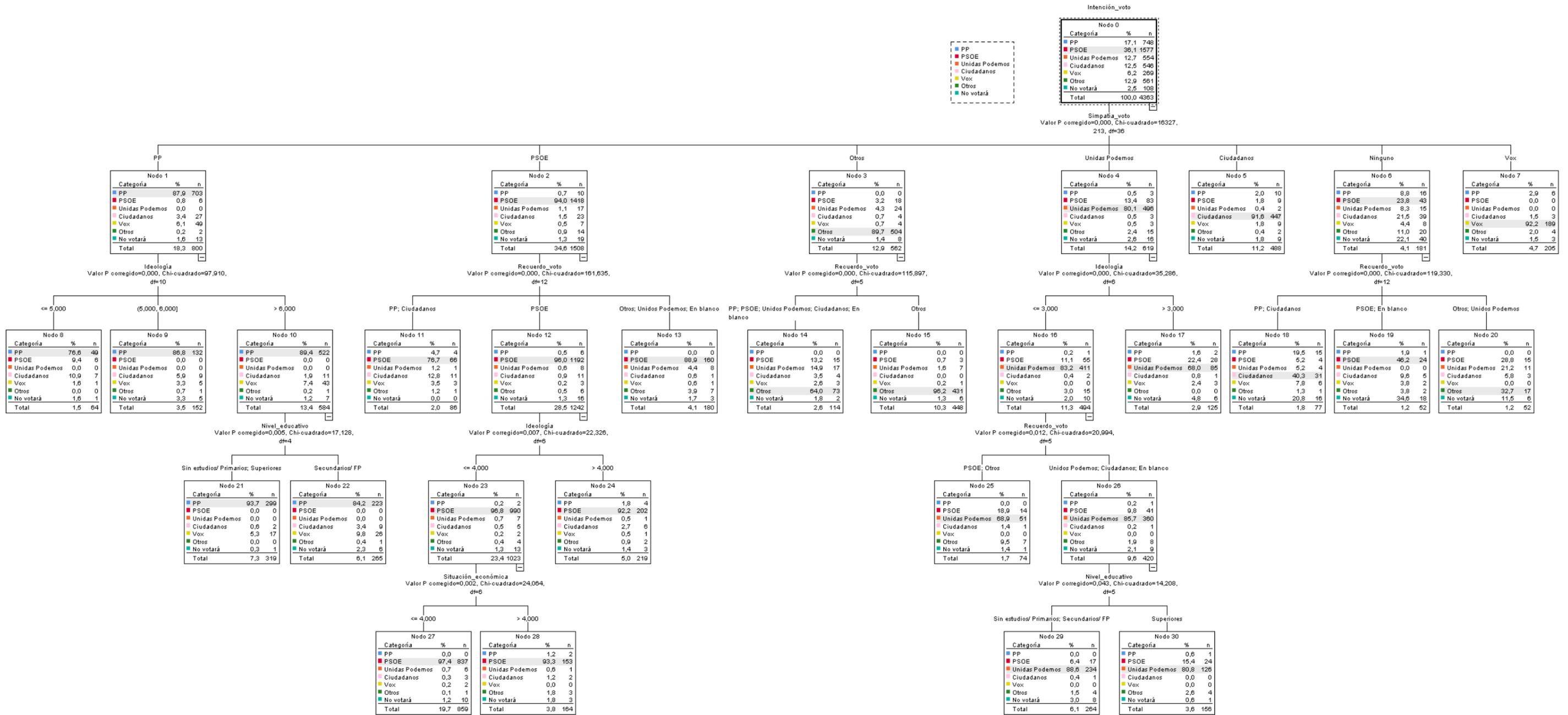
**Anexo 3. Resultados de las Elecciones Generales celebradas el 26 de abril de 2019.**

Resultados Elecciones Generales 28 de abril	
PSOE	28,68%
PP	16,70%
Ciudadanos	15,86%
Unidas Podemos	14,31%
Vox	10,26%
Otros	14,19%

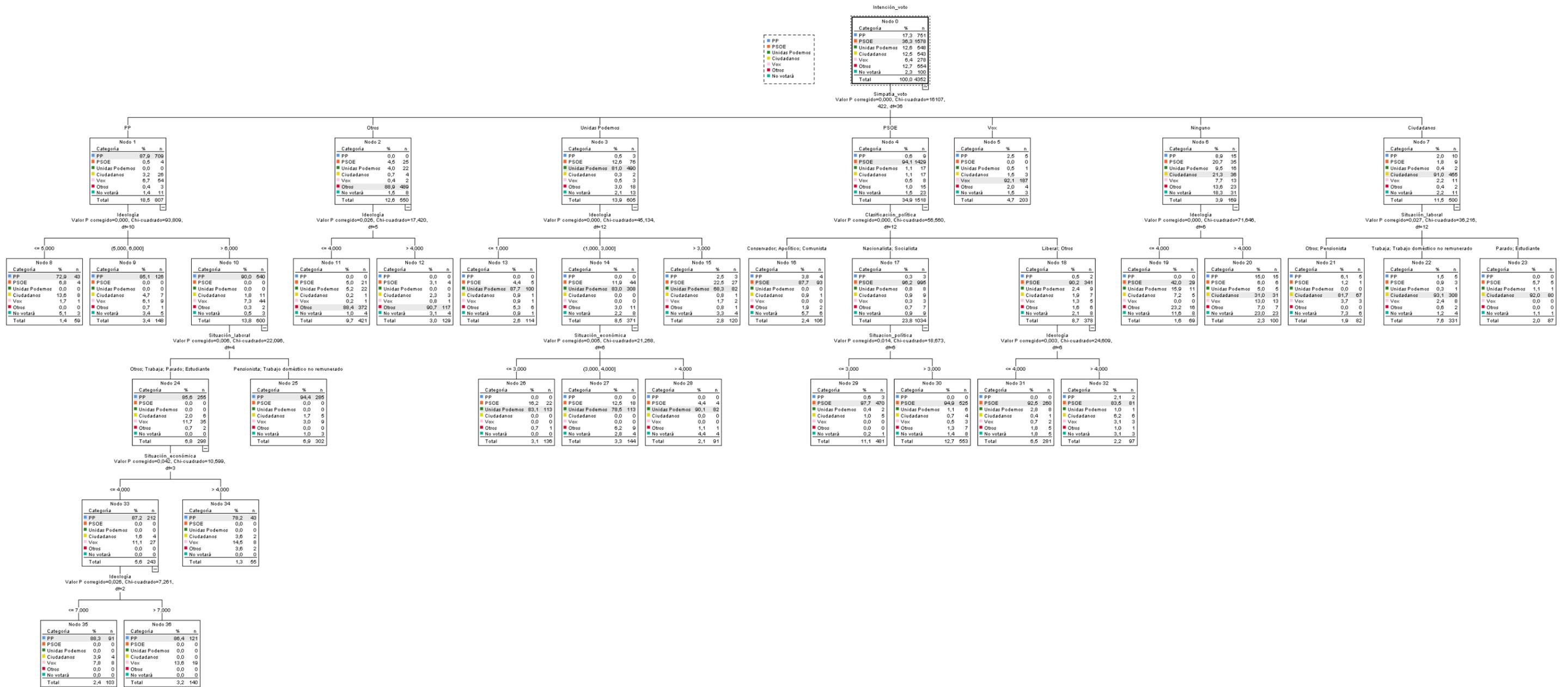
**Anexo 4. Resultados de las Elecciones Europeas celebradas el 28 de mayo de 2019.**

Resultados Elecciones Generales 28 de abril	
PSOE	32,84%
PP	20,13%
Ciudadanos	12,17%
Unidas Podemos	10,05%
Vox	6,20%
Otros	18,61%

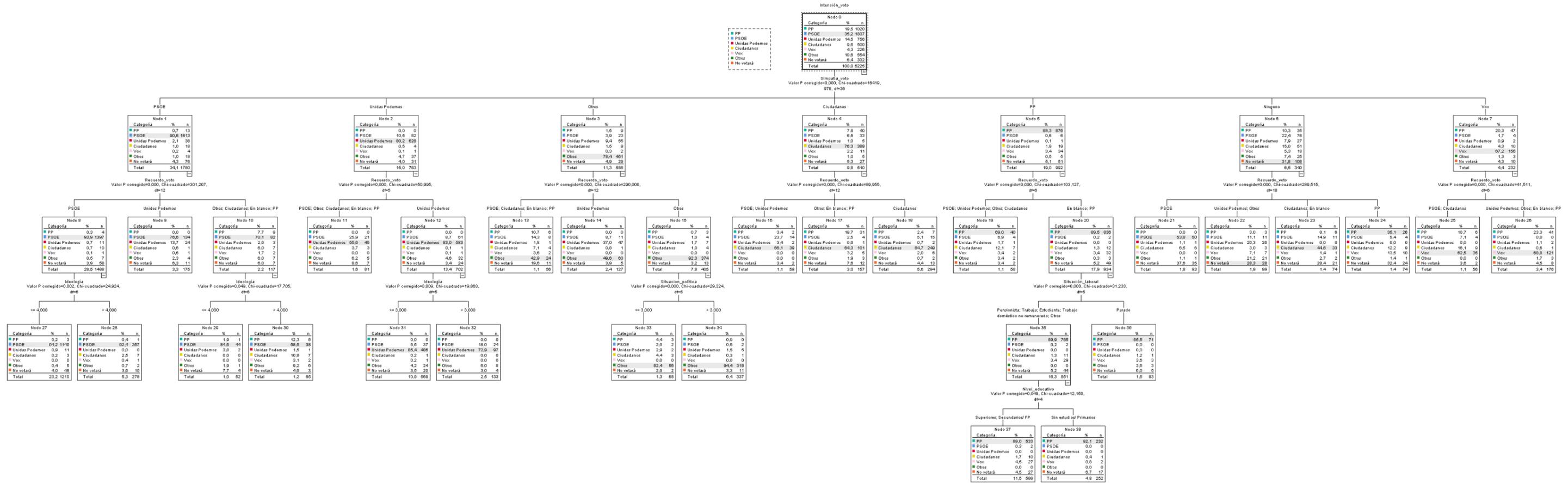
## Anexo 5. Árbol para la predicción de los resultados de las elecciones de abril basado en las encuestas de marzo



## Anexo 6. Árbol para la predicción de los resultados de las elecciones de abril basado en encuesta de marzo sin tomar en cuenta el recuerdo de voto



# Anexo 7. Árbol para la predicción de los resultados de las elecciones de mayo basado en las encuestas de abril



# Anexo 8. Árbol para la predicción de los resultados de las elecciones de mayo basado en encuesta de abril sin tomar en cuenta el recuerdo de voto

