



**Ciencia Latina**  
Internacional

Ciencia Latina Revista Científica Multidisciplinaria, Ciudad de México, México.

ISSN 2707-2207 / ISSN 2707-2215 (en línea), mayo-junio 2024,  
Volumen 8, Número 3.

[https://doi.org/10.37811/cl\\_rcm.v8i3](https://doi.org/10.37811/cl_rcm.v8i3)

## **CONTEO RÁPIDO AUTOMATIZADO: ELECCIONES EN MÉXICO 2024**

**AUTOMATED QUICK COUNT:  
ELECTIONS IN MEXICO 2024**

**Jesús Salazar Ibarra**  
Investigador Independiente, México

**Luz Judith Rodríguez Esparza**  
Universidad Autónoma de Aguascalientes, México

**Rafael G. Vargas Pasaye**  
Investigador Independiente, México

**Lorena Yolanda Carrillo Cerrillo**  
Investigador Independiente, México

DOI: [https://doi.org/10.37811/cl\\_rcm.v8i3.11960](https://doi.org/10.37811/cl_rcm.v8i3.11960)

## Conteo Rápido Automatizado: Elecciones en México 2024

**Jesús Salazar Ibarra<sup>1</sup>**

[ojala1308@hotmail.com](mailto:ojala1308@hotmail.com)

<https://orcid.org/0000-0001-8864-2662>

Investigador Independiente

México

**Rafael G. Vargas Pasaye**

[rafaelvargaspasaye@gmail.com](mailto:rafaelvargaspasaye@gmail.com)

<https://orcid.org/0000-0002-2154-6083>

Investigador Independiente

México

**Luz Judith Rodríguez Esparza**

[luz.rodriguez@edu.uaa.mx](mailto:luz.rodriguez@edu.uaa.mx)

<https://orcid.org/0000-0003-2241-1102>

Universidad Autónoma de Aguascalientes

México

**Lorena Yolanda Carrillo Cerrillo**

[lorenacarrillo@live.com](mailto:lorenacarrillo@live.com)

<https://orcid.org/0009-0002-5962-6027>

Investigador Independiente

México

### RESUMEN

En este estudio se presenta una propuesta de Conteo Rápido Automatizado (CRA) utilizando un enfoque de *machine learning* para estimar los resultados de la votación tanto para la elección a la presidencia de México como para las gubernaturas 2024. El CRA aprovecha los datos en tiempo real proporcionados por el Instituto Nacional Electoral (INE) para determinar, para cada candidato, el modelo de *machine learning* que mejor estima el porcentaje obtenido, utilizando a su vez métodos de remuestreo así como el error cuadrático medio para la selección del mejor modelo y la evaluación de su rendimiento. Se realizó una calibración del CRA utilizando datos de elecciones anteriores en México. Posteriormente, se aplicó el CRA a los datos de las elecciones de 2024, obteniendo estimaciones que fueron consistentes con las presentadas por el Comité Técnico Asesor del INE. Las principales diferencias entre la metodología de Conteo Rápido presentada por el INE y la propuesta radica en la muestra considerada y los algoritmos utilizados. Mientras el INE espera a que se cumpla un porcentaje determinado de la muestra, nuestra metodología toma en cuenta únicamente los datos que se van reportando.

**Palabras clave:** conteo rápido, elecciones, machine learning, México

<sup>1</sup> Autor principal.

Correspondencia: [luz.rodriguez@edu.uaa.mx](mailto:luz.rodriguez@edu.uaa.mx)

# Automated Quick Count: Elections in Mexico 2024

## ABSTRACT

This work presents an Automated Quick Count (CRA) proposal using a machine learning approach to estimate voting results for both the Mexican presidential election and the governorships in 2024. The CRA takes advantage of real-time data provided by the National Electoral Institute (INE) to determine, for each candidate, the machine learning model that best estimates the percentage obtained, using resampling methods as well as the Mean Square Error for the selection of the best model and the evaluation of its performance. A calibration of the CRA was performed using data from past elections in Mexico. Subsequently, the CRA was applied to the data from the 2024 elections, obtaining estimates that were consistent with those presented by the Technical Advisory Committee of the INE. The main differences between the methodology of Quick Count presented by the INE and the proposal lie in the sample considered and the algorithms used. While the INE waits for a certain percentage of the sample to be fulfilled, our methodology takes into account only the data that is reported.

**Keywords:** quick count, elections, machine learning, Mexico

*Artículo recibido 22 mayo 2024  
Aceptado para publicación: 25 junio 2024*



## INTRODUCCIÓN

La importancia de las elecciones en cualquier país es innegable, ya que marcan el curso futuro de la nación (Hartlyn, McCoy & Mustillo, 2009). En la actualidad, cada país tiene su propio proceso para llevar a cabo las elecciones. Por ejemplo, en México<sup>2</sup> se inicia con la preparación de la lista nominal, que es la relación de todos los ciudadanos habilitados para votar. Luego, se seleccionan y acondicionan las casillas electorales, lugares donde se llevará a cabo el proceso de votación. El día de las elecciones, los ciudadanos acuden a las casillas asignadas, presentan su identificación oficial y reciben una boleta para emitir su voto. En la intimidad de la casilla, marcan su preferencia en la boleta de manera secreta y depositan su voto en una urna. Alrededor de las 6pm se cierran las casillas y se procede al conteo de votos en presencia de representantes de los partidos políticos. Este conteo se realiza para determinar los resultados preliminares de la elección en cada casilla, y puede ser supervisado por funcionarios electorales y representantes partidistas. Los resultados de las casillas se transmiten a los centros de cómputo del Instituto Nacional Electoral (INE) para su recopilación y procesamiento.

El INE define el Conteo Rápido<sup>3</sup> como sigue:

*“El Conteo Rápido es un procedimiento estadístico para estimar las tendencias de los resultados finales de una elección. Se selecciona una muestra aleatoria de todas las casillas instaladas el día de la Jornada Electoral, a través de un proceso matemático que se realiza con diferentes métodos de estimación y es aprobado previamente por la autoridad electoral.”*

Desde 1994, los conteos rápidos se han integrado en las elecciones presidenciales de México (Woldenberg, 2012). Estos conteos permiten estimar los resultados electorales mediante una muestra representativa de actas de votación, lo que ofrece resultados preliminares poco después del cierre de las casillas. Esta metodología se implementó con el propósito de garantizar transparencia y confianza en el proceso electoral, al proporcionar resultados preliminares rápidos y confiables.

---

<sup>2</sup> <https://www.ine.mx/>

<sup>3</sup> <https://www.ine.mx/voto-y-elecciones/conteos-rapidos-ine/>



Desde entonces, los conteos rápidos se han convertido en una práctica habitual en las elecciones federales y estatales en México.

El INE establece un Comité Técnico Asesor para el Conteo Rápido el cual es responsable de generar las inferencias del mismo, presentando además intervalos de confianza de al menos el 95%. La rapidez y precisión de este conteo promueven la confianza entre la población y actúan como salvaguarda contra el fraude electoral<sup>4</sup>.

Después del cierre de las casillas electorales, generalmente, cada cinco minutos, los miembros del Comité Técnico reciben actualizaciones sobre los resultados de votos en las casillas. Con el transcurso del tiempo, se acumula el número de casillas utilizadas para el análisis. Sin embargo, las muestras diseñadas no se completan durante esa misma noche, por lo que los resultados se presentan utilizando solo una parte de esas muestras (Anzarut et al., 2019).

En México, además del Conteo Rápido se implementa el Programa de Resultados Electorales Preliminares (PREP)<sup>5</sup>, un sistema informático diseñado por las autoridades electorales para recopilar, procesar y publicar de forma preliminar los resultados de la votación. Mediante el uso de tecnología avanzada, las actas son digitalizadas en una plataforma informática centralizada. Los resultados preliminares del PREP se van presentando a medida que se reciben las actas de las casillas, brindando una visión en tiempo real del avance del conteo de votos. Es importante destacar que los resultados del PREP no tienen validez oficial. Así, el PREP proporciona resultados directamente basados en el conteo de actas de casillas, mientras que el Conteo Rápido utiliza un enfoque de muestreo para obtener estimaciones más ágiles. En conjunto, ambos sistemas se complementan para fortalecer la integridad del proceso electoral.

Anteriormente, se requería una muestra de todos los distritos federales para garantizar la representatividad y la estabilidad de las estimaciones a medida que más datos llegaban. Con el objetivo de agilizar los resultados de los conteos rápidos, Anzarut et al. (2019) propusieron utilizar un modelo jerárquico bayesiano que emplea una regresión para modelar las respuestas de casillas individuales en función de varios atributos, incluidos los factores conocidos por influir en la hora de llegada de los

---

<sup>4</sup> <https://www.ine.mx/voto-y-elecciones/conteos-rapidos-ine/>

<sup>5</sup> <https://www.ine.mx/voto-y-elecciones/prep/>



datos. Cuando falta información de una casilla, ese modelo atrae los parámetros de la casilla hacia otras con atributos similares, lo que mejora los resultados en caso de sesgo en los datos de la submuestra. Además, realizan un ajuste basado en el porcentaje de la muestra observada, lo que aumenta el tamaño de los intervalos de confianza a medida que aumenta el número de casillas faltantes (reflejando la incertidumbre cuando falta más muestra por llegar). Para modelar el tiempo hasta que se reporta una casilla, utilizan un modelo de supervivencia<sup>6</sup>.

En la actualidad, la construcción de modelos bayesianos o frecuentistas con suposiciones adecuadas es crucial para garantizar la precisión y la confiabilidad de las proyecciones electorales. Sin embargo, su implementación, aunque pública en sitios como <https://github.com/tereom/quickcountmx>, no resulta ser muy accesible para aquellos que no tienen experiencia en matemáticas o estadística.

La selección de un tamaño de muestra adecuado es fundamental para obtener estimaciones precisas. Sin embargo, determinar el tamaño óptimo de la muestra es un desafío en sí mismo, ya que implica equilibrar la necesidad de representatividad con consideraciones logísticas y de recursos. Por otro lado, la eficacia del Conteo Rápido depende en gran medida de la recopilación oportuna de datos de la muestra seleccionada por el Comité Técnico. Sin embargo, existen diversos factores que pueden obstaculizar esta tarea, como la infraestructura de comunicaciones deficiente, errores humanos en la transmisión de datos o situaciones de conflicto que afecten el acceso a las casillas. Además, el tiempo necesario para integrar la muestra total es otro desafío importante. La rapidez con la que se puedan recopilar y procesar los datos de todas las casillas seleccionadas determina la agilidad y la fiabilidad de las proyecciones electorales.

Abordar estos problemas de manera efectiva es crucial para desarrollar un nuevo Conteo Rápido basado en algoritmos de *machine learning* que sea confiable, eficiente y capaz de proporcionar proyecciones electorales precisas durante las elecciones en México 2024. Los modelos de *machine learning* aprenden de las características sociodemográficas de los distritos, lo que permite que las estimaciones sean más precisas y estén mejor adaptadas a las particularidades de cada región.

El objetivo principal de esta investigación es presentar una nueva propuesta metodológica para estimar

---

<sup>6</sup> <https://datos.nexos.com.mx/conteos-rapidos-2022-el-desafio-de-pronosticar-elecciones-con-muestras/>



los resultados de la votación para la elección del presidente de México así como varias gubernaturas, utilizando exclusivamente los datos conforme llegan al INE. Para lograr este propósito, se emplearán algoritmos de *machine learning* y diversos métodos de remuestreo. A este nuevo Conteo Rápido lo llamaremos Conteo Rápido Automatizado (CRA).

Para cada candidato, se obtendrá un mejor modelo de *machine learning* robusto y preciso, permitiendo así estimar la proporción de votos con base en los datos proporcionados por el INE en tiempo real y proporcionando intervalos de confianza para los resultados obtenidos.

Se espera que esta investigación contribuya significativamente a la mejora del proceso de conteo rápido en las elecciones presidenciales y estatales, proporcionando estimaciones preliminares que sean cercanas y confiables en comparación con los resultados finales de los cómputos distritales, lo que contribuirá a fortalecer la credibilidad y la transparencia de los procesos electorales en México.

La pregunta de investigación de este trabajo es: ¿Los resultados del Conteo Rápido Automatizado en las elecciones de México 2024 son consistentes con los resultados presentados por el INE?

## **ANTECEDENTES**

En México poco se ha escrito en el ámbito científico respecto a las metodologías estadísticas utilizadas y los resultados obtenidos en las elecciones. Sin embargo, para 2006, debido a la cerrada competencia, sí existen varios estudios Eslava (2009), Mendoza y Nieto-Barajas (2016), Aparicio (2009) y Erdely-Ruiz (2018, 2019).

Predecir los resultados de cualquier elección es de suma importancia. Antes de la elección, éstos se pueden predecir mediante encuestas que miden la intención del voto (ver del Tronco Paganelli, Flores Ivich & Madrigal Ramírez (2016) y Kavanagh (2011)) y durante la elección a través de una muestra aleatoria de personas que son entrevistadas y dicen por qué partido votaron. Sin embargo, estas encuestas de opinión han tenido sus críticas por la forma en que se conducen y sus implicaciones (ver Brown et al (1999), Barreto et al (2006) y Curtice and Firth (2008)).

Otra forma de predecir los resultados finales de la elección es utilizando los resultados que van llegando al INE. Esta información no es el resultado de una selección aleatoria, sin embargo, algunos modelos estadísticos se pueden utilizar, relacionando estos conteos con elecciones previas en las mismas casillas.

Bernando y Girón (1992) usaron un método bayesiano mientras que Pavia-Miralles (2005) un modelo



de regresión multivariado para este tipo de análisis. Los autores usaron la información y obtuvieron estimadores significativos. Estimar el total de votos de cada candidato utilizando la metodología de conteo casilla-por-casilla (registro de casilla que llega al INE) podría producir resultados engañosos (Mendoza & Nieto-Barajas, 2016), por lo que Sedransk & Clyde (1966) proponen usar el tamaño de la población y elecciones pasadas como información valiosa disponible para mejorar la estimación.

Contar los votos finales a través de un Conteo Rápido resulta en una herramienta de gran poder (Estok et al 2002, Cochran, 2001). En este caso, el principal problema es estimar el número total de votos válidos y el total de votos a favor a cada candidato. Generalmente se usan modelos normales para aproximar la distribución de muestreo de los estimadores usando teoría asintótica.

Hacer predicciones basadas en resultados anteriores podría ser técnicamente correcto, pero políticamente incorrecto. Así, Mendoza & Nieto-Barajas (2016) propusieron un modelo bayesiano paramétrico usando datos finales de casillas que fueron seleccionadas a través de un muestreo aleatorio. Este modelo fue utilizado en México para la elección presidencial del 2006 y del 2012.

Recientemente, Argandoña-Mamani, Ormeño-Alarcón, Iparraguirre-Villanueva, Paulino-Moreno & Cabanillas-Carbonell (2024) han presentado una revisión de la literatura (1638 manuscritos) para mostrar el tipo de modelos de aprendizaje automático que mejor funcionaron en la predicción de resultados electorales.

El análisis de sentimientos es uno de los modelos más utilizados; sin embargo, tanto bosques aleatorios (*Random Forest*) como las redes neuronales recurrentes tuvieron una tasa de precisión superior al 91%; estas últimas requirieron un alto esfuerzo computacional, concluyendo que el modelo de bosques aleatorios es el más adecuado para la predicción de resultados electorales.

## METODOLOGÍA

En esta sección se introduce el modelo de regresión general que se empleará para estimar el porcentaje de votos obtenido por cada candidato en la elección. Se proporciona una breve explicación de los algoritmos de *machine learning* utilizados, así como de los métodos de remuestreo que ayudan a optimizar los modelos.

### Modelo estadístico

Consideremos el siguiente modelo:



$$Y = f(X) + \varepsilon$$

donde  $Y$  es la variable de respuesta, tipo numérica (total de votos para cada partido/candidato); así pues, el procedimiento estadístico a realizar es una regresión;  $X$  es un vector de variables independientes o predictoras dadas por: Distritación Federal (DF), Distritación Local (DL), Tipo de casilla electoral y Lista Nominal (que llamaremos en este documento solamente por Lista), para el caso de las gubernaturas y por Tipo, Entidad Federativa (que denotaremos como EDO) y Lista para la elección presidencial;  $f$  es cierta función, dependiendo del algoritmo utilizado y  $\varepsilon$  es el error.

En 2021, el INE comenzó los trabajos para la distritación electoral local y federal<sup>7</sup>, que se realiza periódicamente por mandato constitucional cuando hay un nuevo censo de población, quedando un total de 300 distritos electorales federales, además se crearon 44 distritos electorales federales indígenas y afromexicanos. Además, quedaron 679 distritos electorales locales y se crearon 115 distritos electorales locales indígenas y afromexicanos. Cada Entidad Federativa tiene un número específico de distritos federales y locales.

La variable Tipo se refiere a la tipología de las secciones electorales<sup>8</sup>: Urbana (mayor a 2500 habitantes), Urbana menor (de 1500-2499 habitantes), Rural con amanzanamiento definido (de 100-1499 habitantes) y Rural (de 1-99 habitantes).

La Lista Nominal contiene nombre y foto de la ciudadanía que cuenta con credencial para votar vigente. Es decir, la variable Lista contiene a los habitantes que podrán emitir su voto en la jornada electoral. Debido a que nuestra metodología propone estimar el porcentaje de votos de cada partido/candidato conforme llegan las actas al INE, podría ocurrir que a un determinado tiempo no se hayan capturado aún las actas con todos los registros de las variables DF ó DL del lugar donde se desea hacer la estimación, o bien EDO para la elección presidencial, en dado caso, se consideran solamente las variables Tipo y Lista para la estimación.

### **Algoritmos de *Machine Learning***

La predicción de los resultados de las elecciones a través de modelos de *machine learning* es un tópico que ha ido recobrando importancia en los últimos años (Argandoña-Mamani, et al., 2024). En este

<sup>7</sup><https://www.ine.mx/sobre-el-ine/distritacion-electoral-2021/>

<sup>8</sup>[https://transparencia.ine.mx/obligaciones/rsc/documentos/Articulo74/Formato3a/TIPOLOGIA\\_SECCIONES.pdf](https://transparencia.ine.mx/obligaciones/rsc/documentos/Articulo74/Formato3a/TIPOLOGIA_SECCIONES.pdf)



trabajo se van a considerar los siguientes modelos:

**Redes neuronales (Neural Networks —nnet—):** Una red neuronal es un modelo computacional inspirado en la estructura y funcionamiento del cerebro humano. Consiste en una colección de nodos (neuronas) interconectados, organizados en capas. Cada neurona toma entradas, las procesa mediante una función de activación y produce una salida. En una red neuronal, las capas pueden incluir una capa de entrada, una o más capas ocultas y una capa de salida.

**Vecinos más cercanos (k-Nearest-Neighbor Regression Learner —kknn—):** dado un valor para  $k$  y un punto de predicción  $x_0$ , la regresión kknn primero identifica el conjunto de  $k$  observaciones de entrenamiento más cercanas a  $x_0$ , representado por  $N_0$ , luego estima  $f(x_0)$  usando el promedio de todas las respuestas de entrenamiento en  $N_0$ . El valor óptimo para  $k$  dependerá de la compensación sesgo-varianza. Para  $k$  pequeño el ajuste es más flexible, lo que implica bajo sesgo, pero alta varianza.

**Modelo de regresión lineal (Linear Model —lm—):** es un enfoque muy sencillo para predecir una respuesta cuantitativa teniendo varias variables predictoras, el ajuste se hace a través de una función lineal.

**Árboles de decisión (Classification And Regression Trees —rpart—):** Esta es una técnica de aprendizaje supervisado. Tenemos una variable objetivo (dependiente) y nuestra meta es obtener una función que nos permita predecir, a partir de variables predictoras (independientes), el valor de la variable objetivo para casos desconocidos. Lo que hace este algoritmo es encontrar la variable independiente que mejor separa nuestros datos en grupos, que corresponden con las categorías de la variable objetivo. Esta mejor separación es expresada con una regla. A cada regla corresponde un nodo.

**Bosques Aleatorios (Random Forest —ranger—):** es un tipo de modelo de aprendizaje automático que utiliza múltiples árboles de decisión para realizar predicciones. Cada árbol de decisión se entrena de forma independiente utilizando una parte aleatoria del conjunto de datos de entrenamiento y puede tener diferentes características. Luego, las predicciones de cada árbol se combinan para producir una predicción final.

Para más detalles de los modelos se recomienda Probst, Boulesteix & Bischl (2019) y Fernández-Delgado, Sírsat, Cernadas, Alawadi, Barro & Febrero-Bande (2019).



## Métodos de remuestreo

Los métodos de remuestreo son muy importantes en estadística, pues permiten realizar estimaciones de precisión (James, Witten, Hastie & Tibshirani, 2013). Estos métodos implican extraer muestras aleatorias de un conjunto de entrenamiento y proponer un modelo de interés en cada muestra, para obtener información del modelo ajustado.

**Validación Cruzada (CV por sus siglas en inglés *Cross Validation*):** La CV es uno de los métodos de remuestreo más utilizados y es efectiva para estimar el rendimiento del modelo y seleccionar hiperparámetros. La CV implica dividir aleatoriamente el conjunto disponible de observaciones en dos partes, un conjunto de entrenamiento y un conjunto de validación. El modelo se ajusta al conjunto de entrenamiento y el modelo ajustado se utiliza para predecir las respuestas de las observaciones en el conjunto de prueba o validación.

**Validación Cruzada Repetida (repeated\_cv):** Es una extensión de la CV que repite el proceso de validación cruzada varias veces con diferentes divisiones aleatorias del conjunto de datos. Es computacionalmente más costosa que la CV estándar.

**Holdout:** Este método es el más simple de implementar y comprender. Consiste en dividir el conjunto de datos en dos partes: una para entrenar el modelo y otra para evaluarlo. Este método puede ser útil cuando tienes suficientes datos y deseas una evaluación rápida del rendimiento del modelo. Sin embargo, el método de *holdout* puede producir estimaciones sesgadas del rendimiento del modelo, especialmente con conjuntos de datos pequeños.

**Bootstrap:** Este método es útil cuando tienes un conjunto de datos pequeño y deseas estimar la variabilidad de una estadística de interés (como el error de generalización) sin hacer suposiciones sobre la distribución de los datos. Toma muestras aleatorias con reemplazo de los datos originales y calcula el estadístico de interés en cada muestra.

**Subsampling:** El submuestreo es un método alternativo para aproximar la distribución muestral de un estimado. Divide el conjunto de datos en subconjuntos más pequeños y luego realiza entrenamiento y evalúa el modelo en cada uno de estos subconjuntos. Es útil cuando se trabaja con una gran cantidad de datos. A diferencia del *bootstrap*, el remuestreo se realiza sin reemplazo.

Estos métodos de remuestreo, cada uno con ciertos parámetros específicos, nos permiten dividir los



datos, para así evaluar la variabilidad, comparar el rendimiento y la estabilidad de los modelos de manera robusta.

### **Elección del mejor modelo**

Sea  $\hat{V}_i$  el estimador del total de votos que obtuvo el partido/candidato  $i$  (los valores que toma el índice  $i$  dependen para cada elección, supongamos por ejemplo que  $i \in \{1, \dots, m\}$ ) y sea  $\hat{V} = \sum_i \hat{V}_i$  el total de votos. Entonces la estimación de la proporción que obtuvo el partido/candidato  $i$  está dado por  $\hat{p}_i = \frac{\hat{V}_i}{\hat{V}}$ , para  $i \in \{1, \dots, m\}$ , y por tanto el porcentaje será  $\hat{p}_i \times 100$ .

Para un intervalo de confianza (IC), por ejemplo del 95%, se utiliza el valor crítico  $z_{\alpha/2}$ , donde  $\alpha$  es el nivel de significancia (1 - nivel de confianza). Sea  $d\hat{p}_i = z_{\alpha/2} \sqrt{\left(1 - \frac{n}{N}\right) \frac{\hat{p}_i(1-\hat{p}_i)}{n}}$  donde  $n$  es el número de actas consideradas en el cómputo y  $N$  el total de actas. Luego, el intervalo de confianza para la proporción obtenida por cada partido/candidato  $i$  está dada por:  $\hat{p}_i \pm d\hat{p}_i$ . Por otro lado, si  $d\hat{V}_i = \hat{V} \times d\hat{p}_i$ , entonces el intervalo de confianza para el total de votos que obtuvo el partido/candidato  $i$  está dado por:  $\hat{V}_i \pm d\hat{V}_i$ .

Para elegir el mejor modelo de *machine learning*, es importante considerar varias métricas de evaluación que proporcionen una visión completa del rendimiento del modelo en los datos. En este trabajo consideraremos el Error cuadrático medio (ECM), que es la media de los errores al cuadrado entre las predicciones del modelo y los valores reales. Es útil para evaluar qué tan cerca están las predicciones del modelo de los datos reales.

## **RESULTADOS Y DISCUSIÓN**

### **Ejemplos numéricos: Calibración**

Esta sección tiene como objetivo presentar los resultados de estimaciones de elecciones anteriores para verificar el desempeño óptimo de nuestro modelo. Para lograr este propósito, primero, se tomaron en cuenta cuatro elecciones para gobernador: Aguascalientes 2022, Durango 2022, Estado de México (Edo Mex) 2023 y Oaxaca 2022, se consideraron las primeras 500 actas, i.e.,  $n = 500$ . Después, se consideró la elección presidencial de la República Mexicana en 2018, en este caso  $n = 6,545$ .

Respecto al manejo de los métodos de remuestreo, se consideró lo siguiente. El método de validación



cruzada dividió el conjunto de datos en 3 particiones. La validación cruzada repetida, se repitió 3 veces y además el conjunto de datos también se dividió en 3 particiones. Respecto al método de *holdout*, el 50% del conjunto de datos se utilizó para entrenamiento y el otro 50% para el conjunto de prueba. El *bootstrap* se repitió 3 veces, con tamaño de muestra del 95% del tamaño del conjunto de datos original. Finalmente, el *subsampling* se repitió también 3 veces, utilizando el 95% como tamaño de muestra del conjunto de datos original.

Comenzamos mostrando en la Tabla 1 los porcentajes reales que el INE obtuvo para cada una de las elecciones para gobernador de las Entidades antes mencionadas.

**Tabla 1.**

Resultados reales de los candidatos para distintas gubernaturas de Entidades Federativas de México

Entidad	Candidato	Porcentaje	Entidad	Candidato	Porcentaje
<b>Aguascalientes</b> <b>2022</b>	Teresa Jiménez	53.72	<b>Edo Mex 2023</b>	Alejandra del Moral	44.53
	Cecilia Márquez	1.57		Delfina Gómez	52.71
	Anayeli Muñoz	6.96		No registrados	0.14
	Nora Ruvalcaba	33.70		Nulos	2.62
<b>Durango</b> <b>2022</b>	Teresita Rodríguez	1.34	<b>Oaxaca 2022</b>	Antonia Natividad	3.76
	No registrados	0.05		Alejandra García	3.27
	Nulos	2.66		Bersahin Asael	1.62
	Esteban Villegas	53.78		Mauricio Cruz	1.84
	Marina Vitela	39.29		Jesús López	0.76
	Patricia Flores	4.31		Alejandro Avilés	25.04
	No registrados	0.50		Salomón Jara	60.57
	Nulos	2.13		No registrados	0.08
				Nulos	3.06

Fuente: Elaboración propia con datos de la página del INE<sup>9</sup>.

A continuación se muestran los resultados que se obtuvieron mediante el CRA de cada Entidad Federativa, incluyendo la estimación del total de votos y porcentajes, así como sus intervalos de confianza al 95%.

<sup>9</sup> <https://siceen21.ine.mx/home>



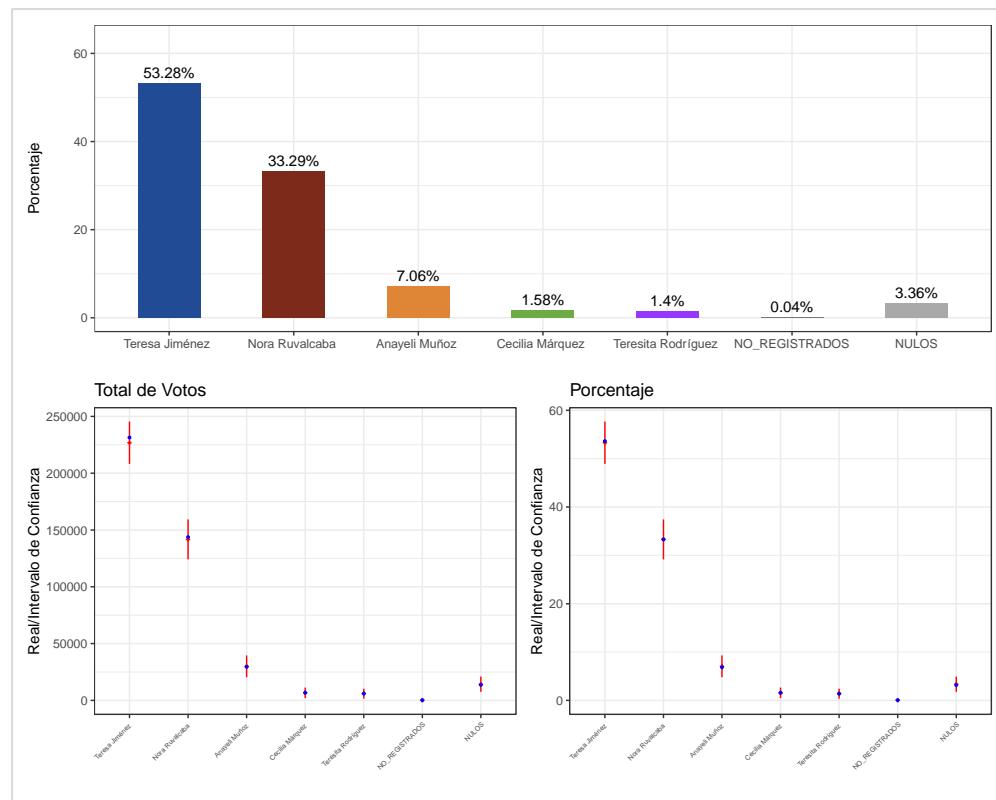
## Aguascalientes 2022<sup>10</sup>

En la Figura 1 se presentan las estimaciones del porcentaje obtenido en las elecciones para gobernador de Aguascalientes en 2022. Notamos que Teresa Jiménez obtuvo una estimación del 53.28% de los votos, mientras que Nora Ruvalcaba el 33.29%. Se muestran además, en color rojo los intervalos de confianza del 95% para el total y el porcentaje, mientras que en color azul se muestra el valor real; notando que todos los valores reales están dentro de los intervalos de confianza obtenidos por el CRA.

## Durango 2022

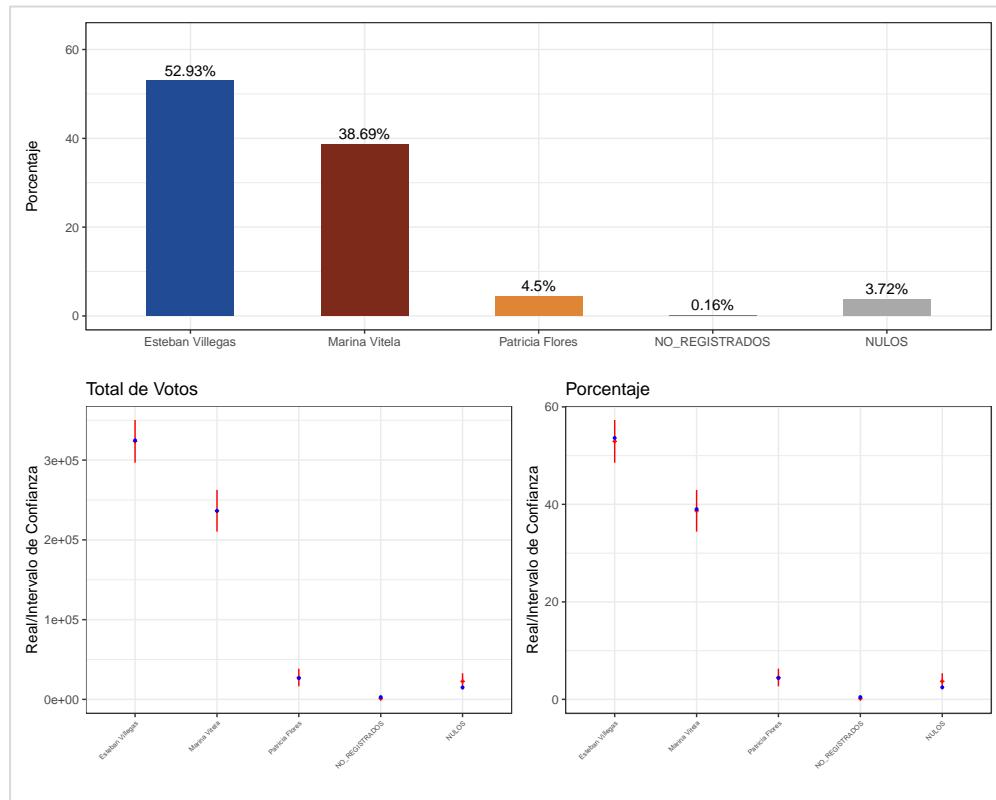
En la Figura 2 se muestran los resultados de las estimaciones de las elecciones de Durango. Esteban Villegas obtuvo una estimación de 52.93%, mientras que Marina Vitela un 38.69%. Todas las estimaciones (total y porcentaje) estuvieron dentro de los intervalos de confianza obtenidos por la estimación del CRA.

**Figura 1.** Estimación de las elecciones para la gubernatura de Aguascalientes en 2022.



<sup>10</sup> <https://siceen21.ine.mx/busqueda/Gobernatura/7/2/2022/2>

**Figura 2.** Estimación de las elecciones para la gubernatura de Durango en 2022.



### Edo Mex 2023

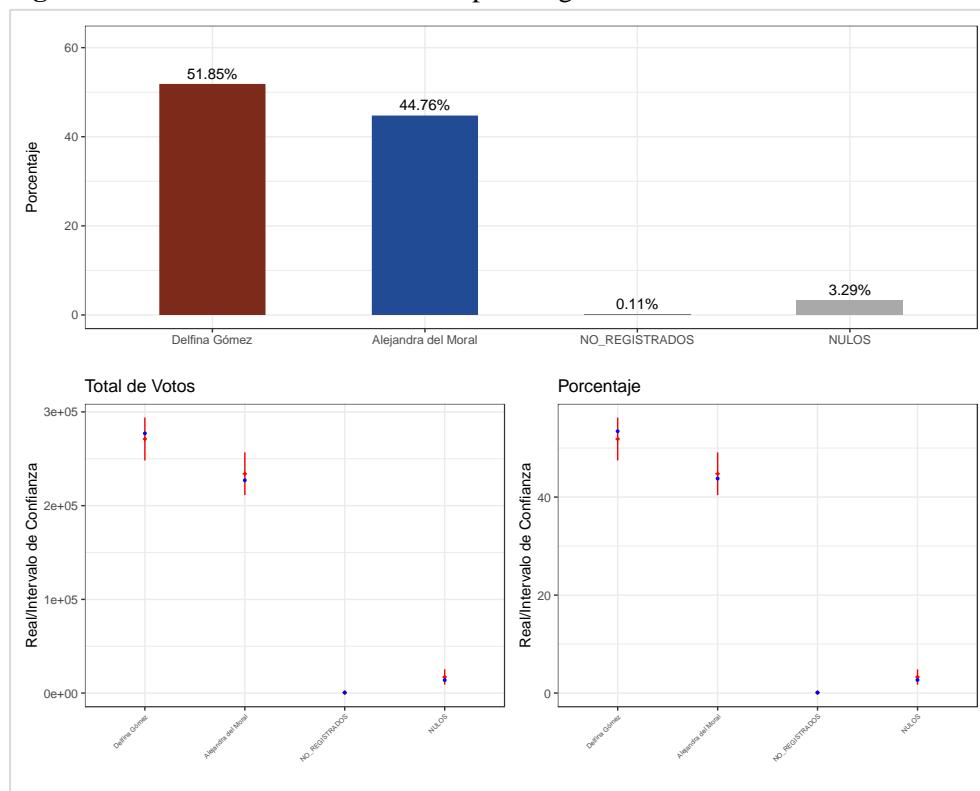
En la Figura 3 se muestran los resultados de las estimaciones de las elecciones del Edo Mex, Delfina Gómez obtuvo una estimación de 51.85%, mientras que Alejandra del Moral un 44.76%.

### Oaxaca 2022

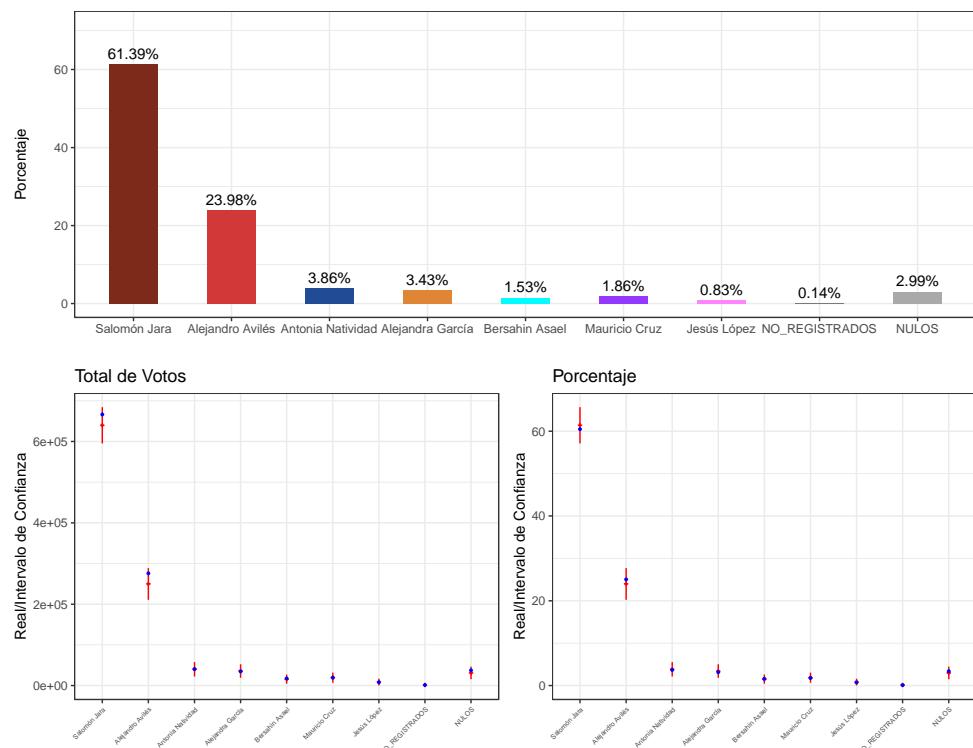
En la Figura 4 se muestran los resultados de las estimaciones de las elecciones de Oaxaca. Salomón Jara obtuvo una estimación de 61.39%, mientras que Alejandro Avilés un 23.98%.

La Tabla 2 muestra los resultados de los algoritmos y métodos de remuestreo más destacados para cada candidato en las cuatro elecciones para gobernador, junto con los correspondientes ECM. Observamos que en el 44% de los casos, bosques aleatorios (ranger) fue el algoritmo más eficaz, seguidos por el modelo de regresión lineal (20%), árboles de decisión (16%), redes neuronales (12%) y, finalmente, vecinos más cercanos (8%). Mientras que el método de remuestreo más utilizado fue el *holdout* (84%), seguido del *subsampling* (12%) y el *bootstrap* (4%).

**Figura 3.** Estimación de las elecciones para la gubernatura del Estado de México en 2023.



**Figura 4.** Estimación de las elecciones para la gubernatura de Oaxaca en 2022.



**Tabla 2.** Resultados de los algoritmos utilizados para la estimación del porcentaje de cada candidato así como el tipo del método de remuestreo que nos proporcionó el menor ECM.

Candidato	Algoritmo	Remuestreo	ECM	Candidato	Algoritmo	Remuestreo	ECM		
<b>Aguascalientes</b>						<b>Edo Mex</b>			
Teresa Jiménez	ranger	holdout	722.049976	Delfina Gómez	ranger	holdout	1000.50708		
Nora Ruvalcaba	ranger	bootstrap	285.426775	Alejandra del Moral	kknn	holdout	2325.30825		
Anayeli Muñoz	ranger	holdout	34.2669996	No registrados	ranger	holdout	0.2318771		
Cecilia Márquez	lm	holdout	3.60832872	Nulos	ranger	holdout	11.7869375		
Teresita Rodríguez	nnet	holdout	4.4007056	<b>Oaxaca</b>					
No registrados	lm	holdout	0.06097069	Salomón Jara	rpart	holdout	1798.71117		
Nulos	nnet	subsampling	13.5253044	Alejandro Avilés	rpart	holdout	618.368218		
				Antonia Natividad	rpart	holdout	33.899487		
<b>Durango</b>						Alejandra García	subsampling	28.7297415	
Esteban Villegas	ranger	holdout	974.667669	Bersahin Asael	lm	subsampling	6.87686634		
Marina Vitela	ranger	holdout	605.920378	Mauricio Cruz	nnet	holdout	10.7871973		
Patricia Flores	ranger	holdout	26.7755123	Jesús López	ranger	holdout	3.44790252		
No registrados	rpart	holdout	0.44747505	No registrados	ranger	holdout	0.16094363		
Nulos	kknn	holdout	10.3268119	Nulos	lm	holdout	21.9526943		

## Elección 2018

En la Tabla 3 se presentan las estimaciones del porcentaje estimado para los candidatos a la presidencia de México en 2018. Las estimaciones de todos los candidatos fueron satisfactoriamente estimadas por la metodología planteada por el CRA.



**Tabla 3.** Resultados de las estimaciones de la elección a presidente de México 2018.

Candidato	Algoritmo	Remuestreo	ECM	Porcentaje Real	IC Inferior	Porcentaje estimado	IC Superior
Ricardo Anaya	lm	subsampling	1528.8	22.27	21.77	22.79	23.81
			3				
José Antonio Meade	lm	holdout	615.08	16.43	15.20	16.09	16.98
Andrés M. López Obrador	kknn	cv	2182.7	53.17	51.50	52.71	53.92
			1				
Jaime Rodríguez	ranger	subsampling	1.43	0.06	0.04	0.13	0.22
Margarita Zavala	kknn	holdout	126.48	5.24	4.79	5.33	5.88
No registrados	rpart	holdout	6.52	0.06	0.06	0.16	0.26
Nulos	lm	subsampling	44.72	2.78	2.38	2.78	3.18

La calibración de nuestra propuesta metodológica del CRA ha demostrado resultados altamente satisfactorios. Las estimaciones de los porcentajes de votos se asemejan considerablemente a los resultados reales, y nuestros intervalos de confianza siempre incluyen el valor real.

### Resultados de las elecciones en 2024

En esta sección se presentan los resultados de la elección presidencial en 2024 así como de 9 gubernaturas (Jalisco, Guanajuato, Yucatán, Veracruz, Tabasco, Morelos, Puebla, Ciudad de México, Chiapas) que presentó el Comité Técnico Asesor del Conteo Rápido del INE, así como de los diferentes PREP de cada Entidad Federativa.

#### Nacional

Los candidatos a la presidencia de México en 2024 fueron Bertha Xóchitl Gálvez Ruiz de la coalición Fuerza y Corazón por México (PAN-PRI-PRD), Claudia Sheinbaum Pardo de la coalición Sigamos Haciendo Historia (PT-PVEM-MORENA) y Jorge Álvarez Mázquez de Movimiento Ciudadano. El Comité Técnico Asesor del Conteo Rápido del INE reportó el 2 de junio de 2024 a las 22:50 hrs que tenía información de 5,651 casillas de un total de 7,602, i.e., el 74.3% de la muestra total, el límite inferior y superior de los porcentajes de cada candidato se presentan en la Tabla 4. Además, dicha tabla



presenta el resultado del PREP reportado a las 17:14 hrs del 4 de junio de 2024<sup>11</sup> y considerando 162,517 de 170,648 actas (i.e, 95.2352%). El CRA tomó una muestra de  $n = 7,097$  actas, es decir el 4% de los datos, teniendo información (actas) de las 32 Entidades Federativas. Los resultados del CRA se dieron a conocer vía medios de comunicación<sup>12</sup>, así como redes sociales a las 21:50hrs<sup>13</sup> del 2 de junio de 2024.

**Tabla 4.** Resultados (porcentajes) de las estimaciones de la elección a presidente de México 2024.

Candidato	Intervalos INE		Intervalos CRA		PREP	CRA
	Inferior	Superior	Inferior	Superior		
Xóchitl Gálvez	26.6	28.6	26.6	28.5	27.9056	27.21
Claudia Sheinbaum	58.3	60.7	58.1	61.0	59.3577	59.82
Jorge Álvarez	9.9	10.8	9.8	10.4	10.4187	10.32

Como se puede observar en la Tabla 4, nuestros resultados son muy similares a los presentados por el PREP, y ambos se encuentran dentro de los intervalos de confianza establecidos por el Comité Técnico Asesor del INE.

En la Tabla 5 se presentan los algoritmos de *machine learning* y métodos de remuestreo que mejor estimaron los porcentajes de los candidatos a la presidencia de México en 2024.

**Tabla 5.** Algoritmos de *machine learning* y métodos de remuestreo que mejor estimaron las elecciones para presidente de México 2024.

Candidato	Algoritmo	Remuestreo	ECM
Xóchitl Gálvez	nnet	subsampling	2645.512
Claudia Sheinbaum	lm	bootstrap	2066.171
Jorge Álvarez	kknn	subsampling	148.3534
No registrados	nnet	subsampling	0.588789
Nulos	ranger	subsampling	18.58745

A continuación se presentan los resultados de las elecciones a gobernadores.

## Jalisco

A las 00:30 hrs del 3 de junio de 2024, el Comité Técnico Asesor reportó información de 347 casillas de 500 en total, informando que el rango del porcentaje obtenido por Laura Lorena Haro Ramírez es de

<sup>11</sup> <https://prep2024.ine.mx/publicacion/nacional/presidencia/nacional/candidatura>

<sup>12</sup> <https://consentidocomun.mx/sheinbaum-ganaria-eleccion-presidencial-codigo-27/>

<sup>13</sup> <https://x.com/MxSentidoComun/status/1797475917931147557>



(15.2, 17.0), mientras que el rango de Claudia Delgadillo González es de (36.4, 39.4) y el de Jesús Pablo Lemus Navarro de (42.5, 45.1). Considerando 6,588 de 10,917 actas, i.e., 60.3462%<sup>14</sup>, el PREP reportó que Lorena Haro tenía 17.5655%, Claudia Delgadillo 38.7341%, mientras que Pablo Lemus 40.7216%. El CRA consideró  $n = 1,808$  actas y las estimaciones con esta metodología se presentan en la Figura 5. Claudia Delgadillo obtuvo el 37.54%, Laura Haro 17.07%, Pablo Lemus 42.65%, 0.17% No registrados y 2.57% Nulos.

### **Guanajuato**

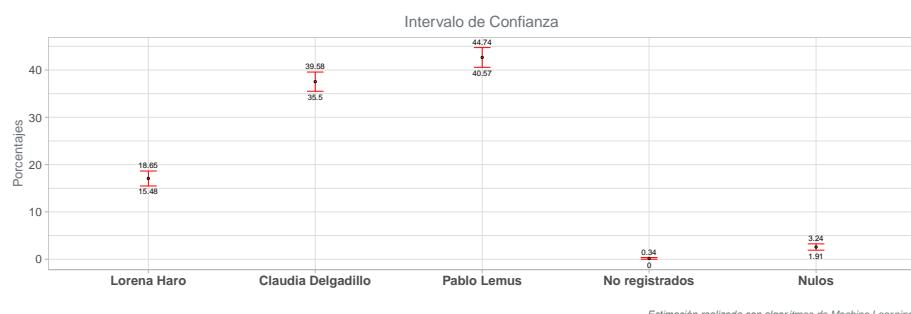
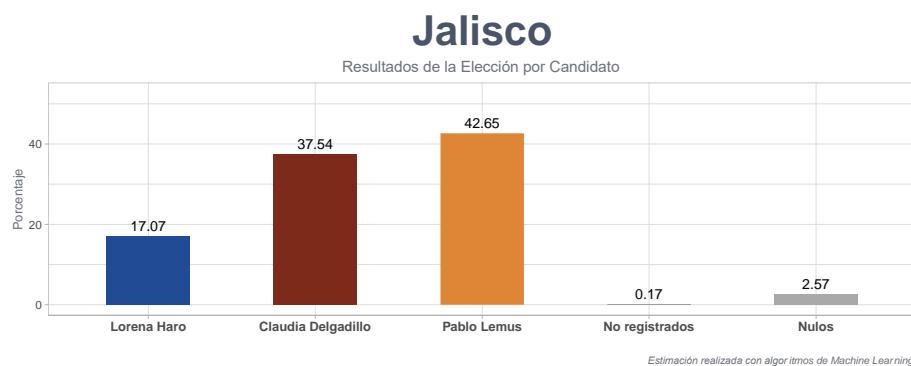
A las 23:25 hrs del 2 de junio de 2024, el Comité Técnico Asesor del INE reportó el resultado de 403 casillas de 500, i.e., el 80.6% de la muestra total, donde Libia Denisse García Muñoz Ledo obtuvo un rango de porcentaje de (49.9, 52.6), Alma Edwviges Alcaraz Hernández (39.8, 42.4) y Yulma Rocha Aguilar (5.3, 6.0). Por otro lado, con 7,715 de 8,166 actas (i.e., 94.4771%)<sup>15</sup>, el PREP reportó que Libia García ganó con el 51.2405%, Alma Alcaraz obtuvo el 40.8649% y Yulma Rocha 5.7076%. El CRA consideró  $n = 1,986$  actas, y en la Figura 6 se presentan las estimaciones. Alma Alcaraz obtuvo el 40.74%, Libia Denisse 51.15%, Yulma Rucha 5.86%, 0.04% No registrados y 2.22% Nulos. Todas las estimaciones del PREP se encuentran dentro de los intervalos de confianza obtenidos por el CRA.

<sup>14</sup> <https://iepcj2024.milenio.com/> a las 13:43hrs del 3 de junio 2024.

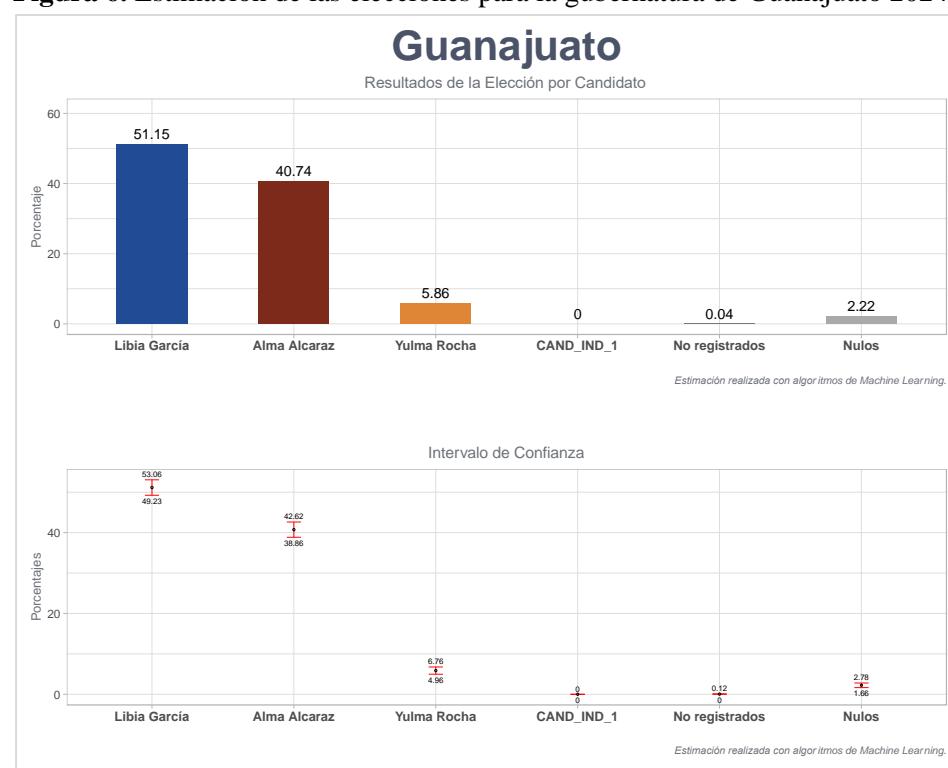
<sup>15</sup> De la página <https://prepto2024.ieeg.mx/#/gubernatura/entidad/votos-entidad/mapa> a las 13:56hrs



**Figura 5.** Estimación del CRA de las elecciones para la gubernatura de Jalisco 2024.



**Figura 6.** Estimación de las elecciones para la gubernatura de Guanajuato 2024.



## Yucatán

El Comité Técnico Asesor reportó a las 00:45 hrs del 3 de junio de 2024 información de 208 casillas de las 350 (i.e., 59.4% de la muestra total) en donde Renán Alberto Barrera Concha obtuvo un rango de

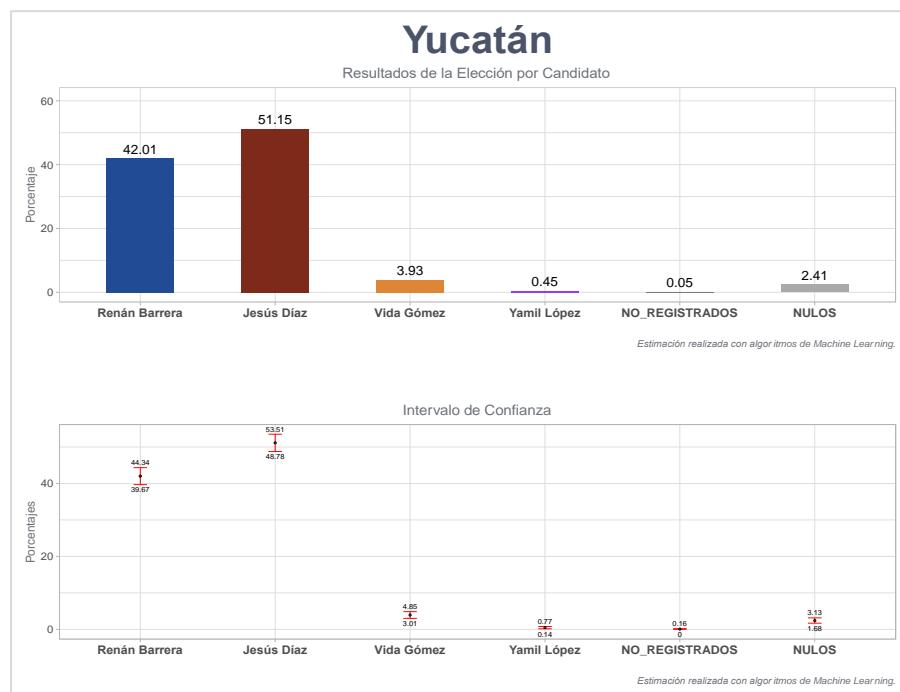


porcentaje de (41.4, 46.7), Yamil Jasmín López Marique (0.3, 1.0), Joaquín Jesús Díaz Mena (46.5, 51.6) y Vida Aravari Gómez Herrera (3.0, 4.4). Considerando 2,429 de 2,978 actas (i.e. 81.5648%)<sup>16</sup> el PREP reportó que Renán Barrera obtuvo 42.5263%, Jesús Díaz 50.9515%, Vida Gómez 3.8477% y Yamil López 0.4886%. El CRA consideró  $n = 1,087$  actas, en la Figura 7 se presentan las estimaciones obtenidas. Joaquín Medina obtuvo el 51.15%, Renán Barrera 42.01%, Vida Gómez 3.93%, Yamil López 0.05% No registrados y 2.41% Nulos.

## Veracruz

El Comité Técnico Asesor reportó a las 23:10 hrs del 2 de junio de 2024, la información de 261 casillas de un total de 350 (i.e., 74.6% del total), donde José Francisco Yunes Zorrila obtuvo un rango de porcentaje de (29.4, 33.2), Norma Rocío Nahle García (57.4, 61.5) e Hipólito Deschamps Espino Barros (6.1, 7.2). Habiéndose capturado 10,090 actas de 11,030, i.e., 91.4777%<sup>17</sup>; el PREP reportó que José Yunes obtuvo 32.1333%, Rocío Nahle 58.1575% e Hipólito Deschamps 7.0463%. El CRA consideró  $n = 1,992$  actas, en la Figura 8 se muestran las estimaciones. Rocío Nahle 58.54%, Pepe Yunes 31.59%, Polo Deschamps 7.16%, 0.10% No registrados y 2.62% Nulos.

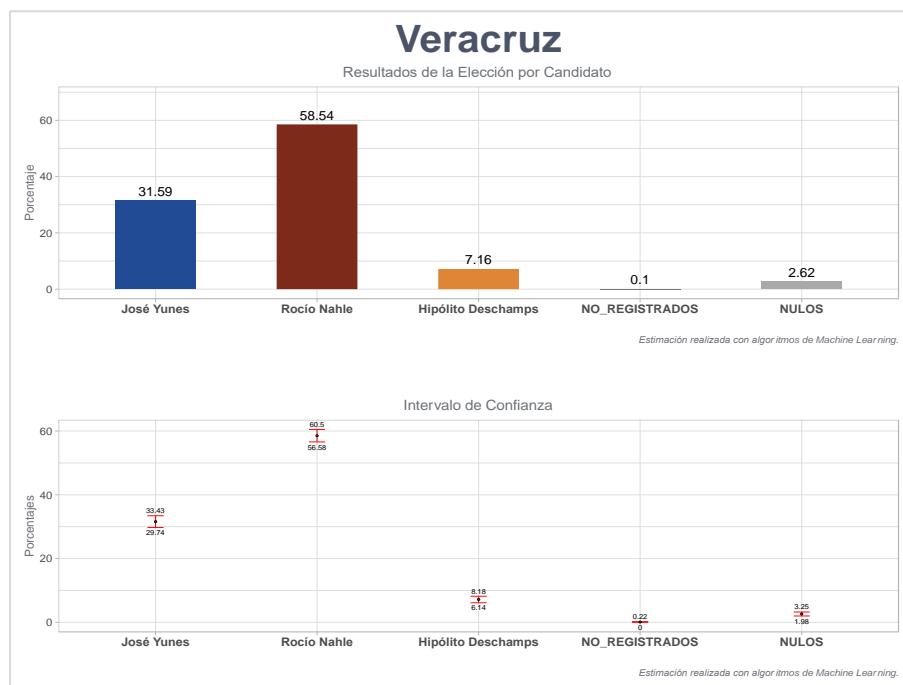
**Figura 7.** Estimación de las elecciones para la gubernatura de Yucatán 2024.



<sup>16</sup> <https://prep-yucatan-2024.sipse.com/gobernatura/votos-candidatura/grafica> 14:01hrs del 3 de junio de 2024

<sup>17</sup> <https://prep2024-ver.org.mx/gobernatura> a las 14:30hrs del 3 de junio de 2024.

**Figura 8.** Estimación de las elecciones para la gubernatura de Veracruz 2024.



## Tabasco

A las 00:00 hrs del 3 de junio de 2024, el Comité Técnico Asesor reportó la estimación considerando 111 casillas de las 200 en total, i.e., 55.5%, donde Lorena Beauregard De Los Santos obtuvo un rango de porcentaje de (3.4, 4.9), Juan Manuel Fócil Pérez (4.9, 7.2), Javier May Rodríguez (78.5, 83.0) y María Inés De La Fuente Dagdug (4.6, 7.2). Considerando 3,053 de 3,117 actas (i.e., 97.9467%)<sup>18</sup>, el PREP reportó que Lorena Beauregard obtuvo un 4.2779%, Javier May 80.4636%, Inés de la Fuente 5.3854% y Manuel Fócil 6.8334%. El CRA consideró  $n = 1,961$  actas. Las estimaciones para esta Entidad se presentan en la Figura 9. Javier May obtuvo el 80.50%, Lorena Bourregard 4.07%, Mines de la Fuente 5.84%, José Manuel Fócil 6.31%, además de 0.07% No registrados y 3.22% Nulos.

## Morelos

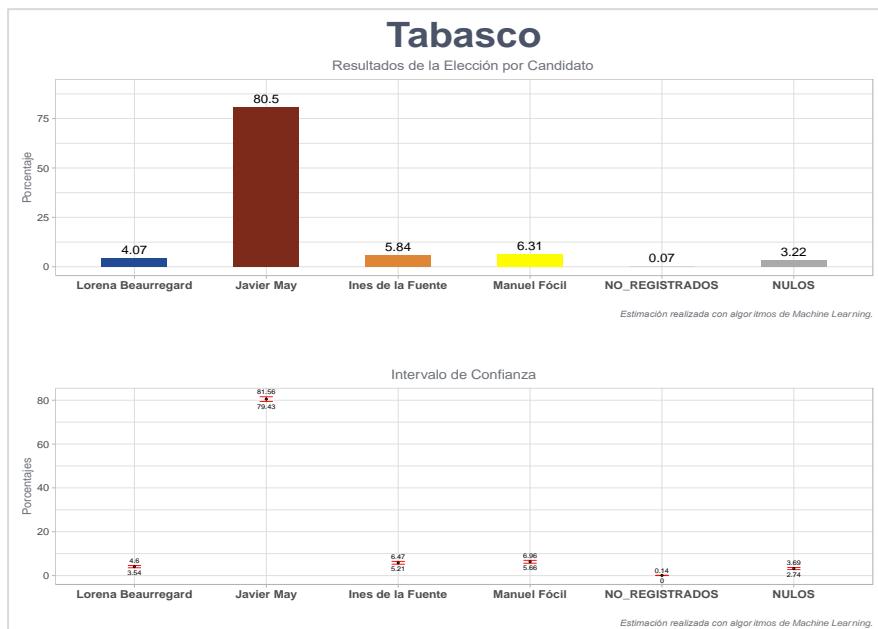
El Comité Técnico Asesor reportó a las 00:55 hrs del 3 de junio de 2024, considerando 132 de 200 casillas (i.e., 66.0%); que Lucía Virginia Meza Guzmán obtuvo un rango de porcentaje de (29.9, 34.4), Margarita González Saravia Calderón (45.7, 50.6) y Jéssica María Guadalupe Ortega De La Cruz (15.8, 19.0). Con 1,604 actas de 2,583 (i.e., 62.0983%)<sup>19</sup>, el PREP reportó que Lucía Meza obtuvo el 30.7702%, Margarita González el 47.8612% y Guadalupe Ortega el 15.2538%. El CRA consideró  $n = 1,046$  actas. En la Figura 10 se muestran las estimaciones, donde Margarita González obtuvo el 47.76%,

<sup>18</sup> <https://prepet.iepct.mx/gobernatura/votos-ambito/grafica> a las 14:35hrs del 3 de junio de 2024.

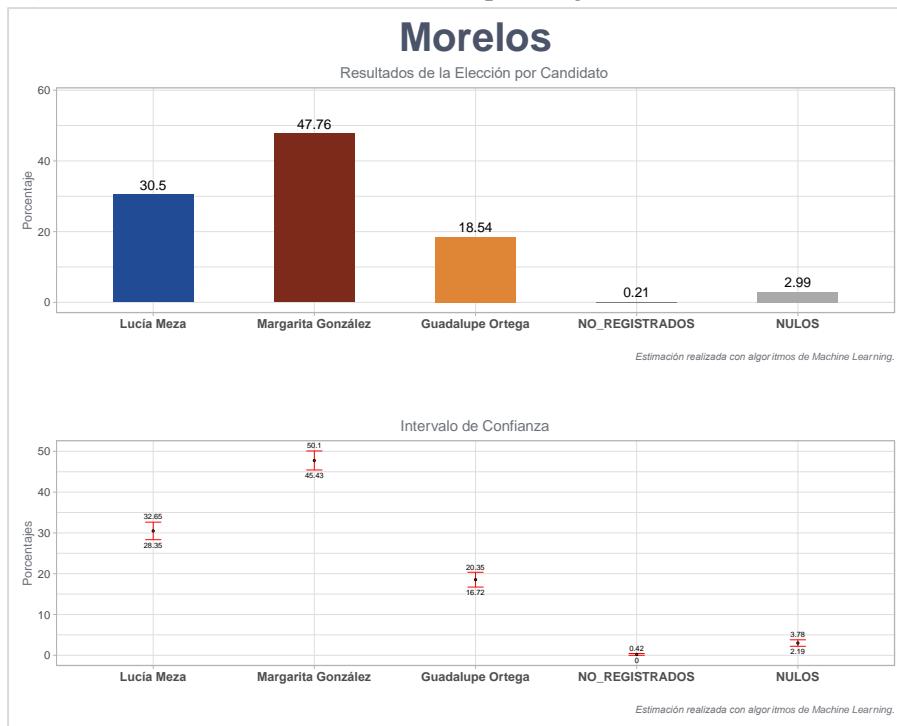
<sup>19</sup> <https://morelos2024.heraldodemexico.com.mx/gobernatura> a las 14:50hrs del 3 de junio de 2024.

Lucy Meza 30.5%, Jessica Ortega 18.54%, 0.21% No registrados y 2.99% Nulos.

**Figura 9.** Estimación de las elecciones para la gubernatura de Tabasco 2024.



**Figura 10.** Estimación de las elecciones para la gubernatura de Morelos 2024.



## Puebla

A las 23:10hrs del 2 de junio de 2024, el Comité Técnico Asesor reportó 347 casillas de 450 (i.e., 77.1%); donde Eduardo Rivera Pérez obtuvo un rango de porcentaje de (31.2, 34.0), Alejandro Armenta Mier (58.8, 61.7) y Fernando Morales Martínez (3.7, 4.4). Por otro lado, tomando 6,493 de 8,334 (i.e.,

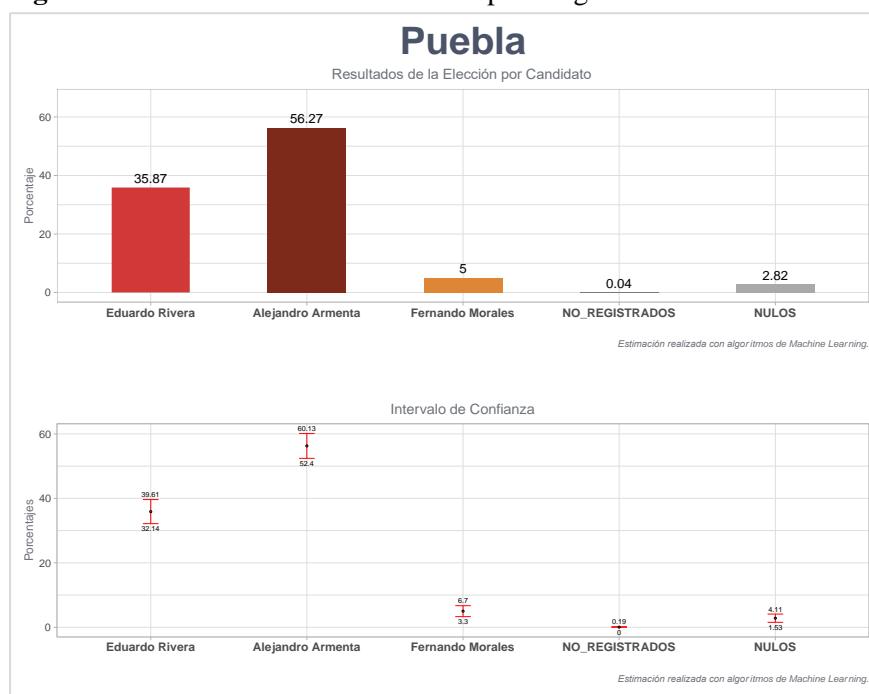


77.9097%)<sup>20</sup> actas, el PREP reportó que Eduardo Rivera obtuvo el 33.2753%, Alejandro Armenta el 59.1507% y Fernando Morales el 4.4475%. El CRA consideró  $n = 589$  actas. En la Figura 11 se muestran los resultados obtenidos por el CRA, donde Alejandro Armenta obtuvo el 56.27%, Eduardo Rivera 35.87%, Fernando Morales 5.0%, 0.04% No registrados y 2.82% Nulos.

### Ciudad de México

El Comité Técnico Asesor a las 23:05 hrs del 2 de junio de 2024, consideró 420 casillas de 600 (i.e., 70.0%), donde Santiago Taboada Cortina obtuvo un rango de porcentaje de (37.2, 40.5), Clara Marina Brugada Molina (49.0, 52.8) y Salomón Chertorivski Woldenberg (6.9, 9.0). Tomando en cuenta 13,109 de 13,460 actas, i.e. el 97.3923%<sup>21</sup>, el PREP reportó que Santiago Taboada obtuvo el 39.0029%, Clara Brugada el 51.7286% y Salomón Chertorivski el 7.3733%. El CRA consideró  $n = 1,945$  actas. En la Figura 12 se presenta la estimación. Clara Brugada obtuvo el 50.90%, Santiago Taboada 40.01%, Salomón Chertorivski 7.14%, 0.10% No registrados y 1.85% Nulos.

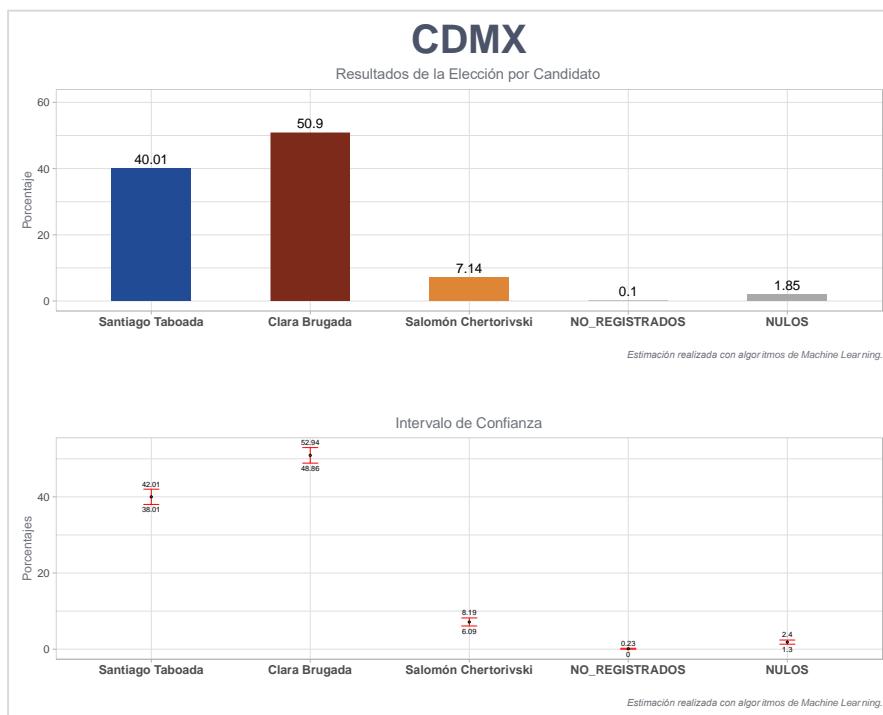
**Figura 11.** Estimación de las elecciones para la gubernatura de Puebla 2024.



<sup>20</sup> <https://prep2024puebla.mx/gobernatura/votos-candidatura/grafica> a las 14:39hrs del 3 de junio de 2024

<sup>21</sup> [https://iecm.contentv.com.mx/prep2024/jg\\_cdmx?\\_gl=1\\*1oc202u\\*\\_ga\\*MTE1MDA5ODM4My4xNzE0NDM0NzM1\\*\\_ga\\_Y0M20DM3VT\\*MTcxNzQ0NjE3OS4xLjAuMTcxNzQ0NjE3OS4wLjAuMA](https://iecm.contentv.com.mx/prep2024/jg_cdmx?_gl=1*1oc202u*_ga*MTE1MDA5ODM4My4xNzE0NDM0NzM1*_ga_Y0M20DM3VT*MTcxNzQ0NjE3OS4xLjAuMTcxNzQ0NjE3OS4wLjAuMA) del día 3 de junio de 2024 las 15:38 hrs.

**Figura 12.** Estimación de las elecciones para la gubernatura de CDMX 2024.

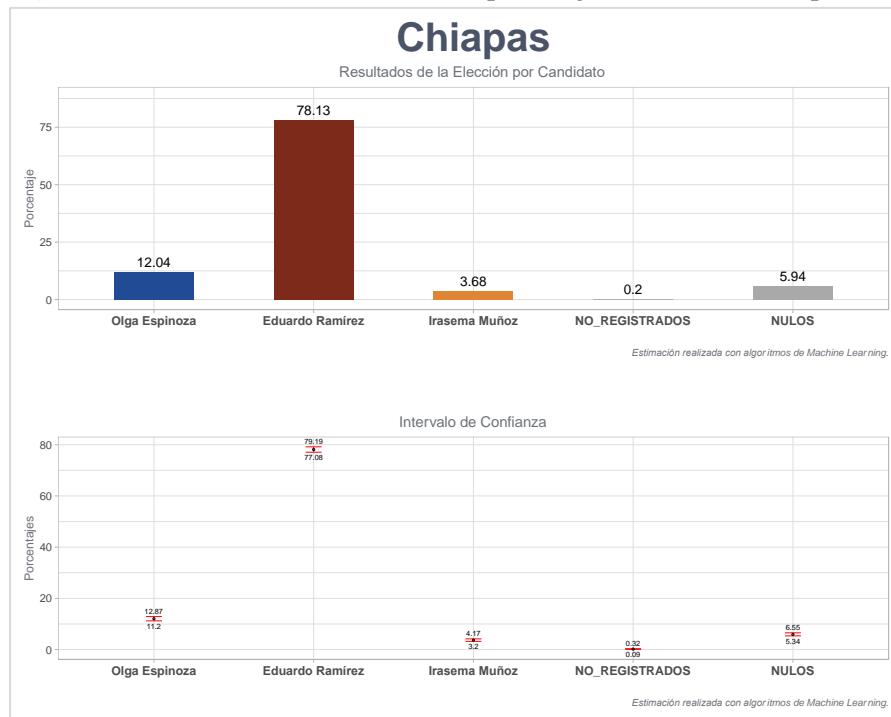


## Chiapas

El Comité Técnico Asesor a las 01:05 hrs del 3 de junio de 2024, consideró 271 casillas de 600 (i.e., 45.2%), donde Olga Luz Espinosa Morales obtuvo un rango de porcentaje de (10.0, 13.4), Óscar Eduardo Ramírez Aguilar (79.0, 82.6) y Karla Irasema Muñoz Balanzar (3.1,3.9). Por otro lado, tomando en cuenta el PREP y considerando 4,566 actas de 6,906 (i.e., 66.1164%)<sup>22</sup>, obtuvo que Olga Espinosa 12.3225%, Eduardo Ramírez 78.8034% e Irasema Muñoz 3.7793. El CRA consideró  $n = 3,166$  actas. En la Figura 13 se presentan los resultados. Eduardo Ramírez obtuvo el 78.13%, Olga Luz Espinoza 12.04%, Karla Muñoz 3.68%, 0.20% No registrados y 5.94% Nulos.

<sup>22</sup> <https://prep2024chiapas.mx/gobernatura/votos-candidatura/grafica> del día 3 de junio de 2024 a las 14:24hrs.

**Figura 13.** Estimación de las elecciones para la gubernatura de Chiapas 2024.



En la Tabla 6 se presentan los algoritmos de *machine learning* que mejor modelaron los datos para las gubernaturas, así como los métodos de remuestreo utilizados. Se observa nuevamente que los Bosques Aleatorios (ranger) han sido el algoritmo que mejor modeló los datos. El método de *subsampling* destacó, logrando un rendimiento destacado el 88% de las veces.

**Tabla 6.** Algoritmos de *machine learning* y métodos de resmuestreo que estimaron las elecciones para gubernamentales en 2024.

Algoritmo	Conteo	Porcentaje	Remuestreo	Conteo	Porcentaje
Redes neuronales	12	25	CV	0	0
Vecinos más cercanos	9	19	Repeated_C	1	2
Regresión lineal	8	17	Holdout	2	4
Árboles de decisión	2	4	Bootstrap	3	6
Bosques aleatorios	17	35	Subsampling	42	88

## CONCLUSIONES

Las elecciones de cualquier lugar son de suma importancia, ya que de ellas depende el futuro de ese lugar. A lo largo del tiempo, se han utilizado diversas metodologías para estimar los resultados finales de manera rápida y precisa. En México, en particular, se han implementado los Conteos Rápidos en las

últimas décadas con el objetivo de proporcionar a los ciudadanos una estimación de los resultados finales alrededor de las 10pm. Estas metodologías se han ido refinando a lo largo del tiempo, adaptándose a los avances en los métodos de muestreo. Como resultado, se han logrado obtener estimaciones cada vez más precisas con errores más pequeños.

En el ámbito de la estadística, las matemáticas y las ciencias de datos, los algoritmos de *machine learning* han emergido como herramientas poderosas que han contribuido significativamente a mejorar la precisión de estos estimadores. Estos algoritmos nos permitieron obtener estimaciones con errores muy pequeños, lo que ayudó a fortalecer la confianza en los resultados proporcionados por el conteo rápido propuesto (CRA).

El Comité Técnico Asesor del INE seleccionó una muestra aleatoria de las casillas. Cuando no llega información a tiempo de alguna de ellas, estiman sus resultados utilizando información de casillas contiguas mediante métodos estadísticos. Sin embargo, como pudimos constatar en estas elecciones, al no cumplirse con el 75% de la muestra, se tuvo que dar el resultado hasta las 23:30 hrs, cuando generalmente se daba a las 22:00 hrs.

Por otro lado, de acuerdo a los resultados finales de la elección obtenidos por el conteo de las actas de los 300 Consejos Distritales del INE, Claudia Sheinbaum obtuvo el 59.7594%, Xóchitl Gálvez el 27.4517% y Jorge Álvarez Mázquez el 10.3213%. Por lo que los resultados obtenidos por el CRA (ver Tabla 4) son altamente satisfactorios.

Dado que el CRA no utiliza una muestra fija, una ventaja destacada del CRA incluye su implementación rápida y por tanto una estimación de los resultados más oportuna, obteniéndose estimadores confiables, pues los algoritmos utilizados construyen aprendizaje del comportamiento electoral fundamentado en las características generales de las casillas, como su ubicación, las casillas vecinas, entre otros.

Además, el tiempo necesario para obtener estimaciones precisas en función del flujo de llegada de las actas electorales, proporcionó información valiosa sobre el rendimiento y la eficacia del CRA. Así, el CRA propuesto demuestra ser una herramienta efectiva y confiable para realizar estimaciones precisas del total y porcentaje de votos en elecciones de cualquier lugar, aprovechando la capacidad del *machine learning* para adaptarse dinámicamente a los datos en tiempo real.

Como trabajo futuro, sería valioso considerar las secciones electorales en el análisis. Incorporar esta



dimensión permitirá una comprensión más granular y precisa del comportamiento electoral.

## Agradecimientos

Expresamos nuestro agradecimiento al Consejo de Ciencia y Tecnología del estado de Durango (COCYTED) por su orientación y apoyo para la realización de esta investigación enmarcada en la iniciativa estratégica DuranIA. De igual manera a la empresa Código 27 por el soporte financiero para la realización de los trabajos de experimentación.

## REFERENCIAS BIBLIOGRAFICAS

- Anzarut, M., González, L. F., & Ortiz, M. T. (2019). A heavy-tailed multilevel mixture model for the quick count in the Mexican elections of 2018. In *Selected Contributions on Statistics and Data Science in Latin America: 33 FNE and 13 CLATSE, 2018, Guadalajara, Mexico, October 1–5* (pp. 1-13). Springer International Publishing.  
[https://link.springer.com/chapter/10.1007/978-3-030-31551-1\\_1](https://link.springer.com/chapter/10.1007/978-3-030-31551-1_1)
- Aparicio, J. (2009). Análisis estadístico de la elección presidencial de 2006: ¿Fraude o errores aleatorios?. *Política y gobierno*, 16(SPE2), 225-243.  
[https://www.scielo.org.mx/scielo.php?pid=S1665-20372009000400010&script=sci\\_arttext](https://www.scielo.org.mx/scielo.php?pid=S1665-20372009000400010&script=sci_arttext)
- Argandoña-Mamani, A., Ormeño-Alarcón, T., Iparraguirre-Villanueva, O., Paulino-Moreno, C., Cabanillas-Carbonell, M. (2024). Predicting Election Results with Machine Learning—A Review. In: Yang, XS., Sherratt, R.S., Dey, N., Joshi, A. (eds) Proceedings of Eighth International Congress on Information and Communication Technology. ICICT 2023. Lecture Notes in Networks and Systems, vol 695. Springer, Singapore. [https://doi.org/10.1007/978-981-99-3043-2\\_82](https://doi.org/10.1007/978-981-99-3043-2_82)
- Barreto, M. A., Guerra, F., Marks, M., Nuño, S. A., & Woods, N. D. (2006). Controversies in exit polling: Implementing a racially stratified homogenous precinct approach. *PS: Political Science & Politics*, 39(3), 477-483. doi:10.1017/S1049096506060689
- Bernardo, J. M., & Girón, F. J. (1992). Robust sequential prediction from non-random samples: The election night forecasting case. *Bayesian Statistics*, 4, 61-77.
- Brown, P. J., Firth, D., & Payne, C. D. (1999). Forecasting on British election night 1997. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 162(2), 211-226.



<https://doi.org/10.1111/1467-985X.00131>

Cochran, W. G. (2001). *Sampling techniques*. John Wiley & Sons.

Curtice, J., & Firth, D. (2008). Exit polling in a cold climate: the BBC–ITV experience in Britain in 2005. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 171(3), 509–539.

<https://doi.org/10.1111/j.1467-985X.2007.00536.x>

del Tronco Paganelli, J., Flores Ivich, G., & Madrigal Ramírez, A. (2016). La utilidad de las encuestas en la predicción del voto. La segunda vuelta de Argentina 2015. *Revista mexicana de opinión pública*, (21), 73-92.

[https://www.scielo.org.mx/scielo.php?pid=S2448-49112016000200073&script=sci\\_arttext](https://www.scielo.org.mx/scielo.php?pid=S2448-49112016000200073&script=sci_arttext)

Da Silva Santos , F., & López Vargas , R. (2020). Efecto del Estrés en la Función Inmune en Pacientes con Enfermedades Autoinmunes: una Revisión de Estudios Latinoamericanos. *Revista Científica De Salud Y Desarrollo Humano*, 1(1), 46–59.

<https://doi.org/10.61368/r.s.d.h.v1i1.9>

Erdely Ruiz, A. (2019). Análisis estadístico ex post del conteo rápido institucional de la elección de gobernador del Estado de México en 2017. *Apuntes Electorales*, 18(60), 227–262. Recuperado a partir de <https://aelectorales.ieem.org.mx/index.php/ae/article/view/139>

Eslava, G. (2009). Las elecciones de 2006 un análisis del conteo rápido. *Ciencias*, (084). Recuperado a partir de <https://revistas.unam.mx/index.php/cns/article/view/12072>

Estok, M., Nevitte, N., Cowan, G., 2002. The Quick Count and Election Observation. A NDI Handbook for Civil Organizations and Political Parties. National Democratic Institute. Available at: [https://www.ndi.org/files/1417\\_elect\\_quickcounthdbk\\_0.pdf](https://www.ndi.org/files/1417_elect_quickcounthdbk_0.pdf)

Fernández-Delgado, M., Sirsat, M. S., Cernadas, E., Alawadi, S., Barro, S., & Febrero-Bande, M. (2019). An extensive experimental survey of regression methods. *Neural Networks*, 111, 11-34. <https://doi.org/10.1016/j.neunet.2018.12.010>

Gómez Ramón , R. del C., Diez Irizar , G. A., Herrera Sánchez , B., & Pérez Rejón , J. Ángel. (2024). Impacto del curso propedéutico de matemáticas básicas: caso estudiantes de nuevo ingreso a nivel superior. *Estudios Y Perspectivas Revista Científica Y Académica* , 4(1), 179–194.

<https://doi.org/10.61384/r.c.a.v4i1.91>



Gómez Ramón , R. del C., Diez Irizar , G. A., Herrera Sánchez , B., & Pérez Rejón , J. Ángel. (2024).

Impacto del curso propedéutico de matemáticas básicas: caso estudiantes de nuevo ingreso a nivel superior. *Estudios Y Perspectivas Revista Científica Y Académica* , 4(1), 179–194.

<https://doi.org/10.61384/r.c.a.v4i1.92>

Hartlyn, J., McCoy, J., & Mustillo, T. M. (2009). La importancia de la gobernanza electoral y la calidad de las elecciones en la América Latina contemporánea. *América Latina Hoy*, 51, 15-40.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: with applications in R*. Springer. <https://link.springer.com/book/10.1007/978-1-0716-1418-1>

Kavanagh, D. (2011). Las encuestas de opinión pública. *La Sociología en sus Escenarios*, (24).

<https://revistas.udea.edu.co/index.php/ceo/article/view/10975>

Mendoza, M., & Nieto-Barajas, L. E. (2016). Quick counts in the Mexican presidential elections: A Bayesian approach. *Electoral Studies*, 43, 124-132.

<https://doi.org/10.1016/j.electstud.2016.06.007>

Martínez, O., Aranda , R., Barreto , E., Fanego , J., Fernández , A., López , J., Medina , J., Meza , M., Muñoz , D., & Urbieta , J. (2024). Los tipos de discriminación laboral en las ciudades de Capiatá y San Lorenzo. Arandu UTIC, 11(1), 77–95. Recuperado a partir de

<https://www.uticvirtual.edu.py/revista.ojs/index.php/revistas/article/view/179>

Nanguce López , M., & Sierra Ramírez, J. A. (2024). Funcionalidad familiar y depresión en adultos mayores con diabetes. *Revista Científica De Salud Y Desarrollo Humano*, 5(1), 378–391.

<https://doi.org/10.61368/r.s.d.h.v5i1.101>

Pavía-Miralles, J. M. (2005). Forecasts From Nonrandom Samples: The Election Night Case. *Journal of the American Statistical Association*, 100(472), 1113–1122.

<https://doi.org/10.1198/016214504000001835>

Probst, P., Boulesteix, A. L., & Bischl, B. (2019). Tunability: Importance of hyperparameters of machine learning algorithms. *Journal of Machine Learning Research*, 20(53), 1-32.

<https://www.jmlr.org/papers/v20/18-444.html>

Ruiz, A. E. (2018). La falacia del empate técnico electoral/The fallacy of technical tie in electoral polls. *Revista Mexicana de Estudios Electorales*, 2(20), 11-47.



<https://rmee.org.mx/index.php/RMEstudiosElectorales/article/view/260>

Sedransk, J. H., & Clyde, R. W. (1966). A Generalized Technique for Predicting Election Results From Early Returns. *The Public Opinion Quarterly*, 30(3), 416-422.

<https://www.jstor.org/stable/2747141>

v, H., & Quispe Coca, R. A. (2024). Tecno Bio Gas. Horizonte Académico, 4(4), 17–23. Recuperado a partir de <https://horizonteacademico.org/index.php/horizonte/article/view/14>

Woldenberg, J. (2012). *Historia mínima de la transición democrática en México*. El Colegio de Mexico AC.[https://books.google.com.mx/books?hl=es&lr=&id=MfkZAwAAQBAJ&oi=fnd&pg=PT4&dq=conteos+r%C3%A1pidos+en+mexico+historia+empiezan+en+1994&ots=OMJxT2mtCE&sig=Gyps0b6u1Ow3i\\_08NGfU2SIc-0U&redir\\_esc=y#v=onepage&q&f=false](https://books.google.com.mx/books?hl=es&lr=&id=MfkZAwAAQBAJ&oi=fnd&pg=PT4&dq=conteos+r%C3%A1pidos+en+mexico+historia+empiezan+en+1994&ots=OMJxT2mtCE&sig=Gyps0b6u1Ow3i_08NGfU2SIc-0U&redir_esc=y#v=onepage&q&f=false)

