

Pronóstico electoral: ¿Cuál sería el resultado si el plebiscito fuese mañana?

20-08-2022

Presentación

El presente documento es el quinto reporte en serie que indica los resultados de un análisis predictivo sobre el avance del clima electoral de cara al plebiscito de salida para la propuesta de Nueva Constitución en Chile para el año 2022. Este análisis trabaja con datos masivos aleatorizados y normalizados que se obtienen desde las tendencias de Google, registrando el interés en el tiempo que generan ciertos conceptos buscados, en esta ocasión referidos al *apruebo* y al *rechazo*. No es una encuesta, sino un modelo matemático en base a la interacción en el mundo virtual, específicamente a las series de interés en el tiempo de Google Trends. Se basa en el supuesto de que las personas que participarán de un proceso electoral buscan información sobre las alternativas disponibles para informar su voto, o bien, participan de campañas de forma activa, interactuando con los conceptos asociados a cada campaña en redes sociales e internet. Buena parte de estas interacciones son recogidas por los motores de búsqueda de Google. El uso de esta técnica de análisis es relativamente reciente en la literatura, pero con buenos resultados en investigaciones similares¹. En este caso, se aplica un método estadístico de modelación futura por serie de tiempo usando la técnica ARIMA, o modelo autorregresivo integrado de media móvil². Esta técnica toma una serie de datos

RESUMEN

- Se presenta modelo matemático que ha pronosticado con buena precisión resultados electorales recientes en Chile.
- Para el plebiscito de entrada, el modelo arrojó como resultado que la opción Apruebo ganaría con un 77.80% de las preferencias, mientras resultado final fue de 78.28%.
- Para la segunda vuelta entre Boric y Kast el modelo pronosticó un triunfo de Boric por un 57% contra 43% de Kast, mientras resultado final fue de 55.64% para Boric.
- Si el plebiscito fuera mañana, el modelo indica que la opción Apruebo obtendría con un 55.1% en series por horas desde hace 21 días.

¹ Riccardo Cantini et al., 'Analyzing Political Polarization on Social Media by Deleting Bot Spamming', *Big Data and Cognitive Computing* 6, no. 1 (4 January 2022): 3, <https://doi.org/10.3390/bdcc6010003>; Maryam Bilal et al., eds., '2018 Pakistan General Election: Understanding the Predictive Power of Social Media', in *2018 12th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics (MACSS)*, 12 (Karachi, 2018); Harald Schmidbauer, Angi Rösch, and Fabian Stieler, 'The 2016 US Presidential Election and Media on Instagram: Who Was in the Lead?', *Computers in Human Behavior* 81 (April 2018): 148–6 <https://doi.org/10.1016/j.chb.2017.11.021>.

² Damodar N. Gujarati and D. C. Porter, *Basic of Econometric, Fifth Edition., Econometrics*, 2009.

ordenadas temporalmente e identifica a partir de su propia historia estadística, cual es el valor futuro más probable para esa serie de datos, indicando también sus márgenes de error. Este método se prefiere por sobre otros por tener más etapas metodológicas que permiten curar mejor los datos a usar, sin embargo, el resultado aplicando otras técnicas univariantes es similar a los que hemos entregado hasta ahora.

El uso de esta técnica de análisis fue calibrado usando datos coyunturales durante otras elecciones recientes en Chile. En el plebiscito 202 el modelo recogió datos para las últimas tres semanas de la campaña electoral hasta el día anterior del plebiscito por una Nueva Constitución a celebrarse el 25 de octubre 2020. En esa primera aproximación el resultado del modelo fue de un 77.92% para el apruebo, mientras que el resultado electoral real fue de 78.28% para el apruebo. En esta primera modelación se usaron datos diarios, no datos por horas. De manera similar, para la elección Boric y Kast se hizo el mismo ejercicio. Los resultados de este ejercicio predictivo fueron publicados en este [vínculo](#), indicando que Gabriel Boric obtendría un 57.27% y José Antonio Kast un 42.73%, registrando en esta ocasión una diferencia de 1.63% con el resultado definitivo de esa segunda vuelta, logrando así similar precisión que la encuesta Critería (diferencia de 1.6%), CADEM (diferencia de 2.64%), Pulso Ciudadano (3.36%) y Atlas Intel (4.64%). Si bien el modelo fue menos preciso que el del plebiscito, frente a otros instrumentos de medición logra resultados competitivos. En este tercer ensayo, se puede comenzar a idear la hipótesis de que en Chile se pueden generar pronósticos en base a modelación matemática a partir de datos de redes sociales. **El supuesto aun se encuentra en discusión dado que el plebiscito de salida tiene una modalidad de voto obligatorio universal, lo que genera un panorama distinto a las elecciones anteriores.** Para enfrentar teóricamente esta incertidumbre, se plantea que la interacción entre las personas y las opciones electorales en búsquedas, noticias y actividad en general en internet, forma parte de un proceso tendiente a informar la decisión de voto o bien sobre la adhesión sobre una de las opciones. Una vez que el voto esta informado y la adhesión consolidada -es decir, más cerca de la elección, el modelo indicaría el dominio de una opción electoral sobre otra con más precisión. Justamente, este reporte se realiza cerca de la elección. A partir de este supuesto, el impacto del voto obligatorio debiera estar representado en este análisis, aunque eso recién lo confirmaremos o rechazaremos el 5 de septiembre de 2022.

Pronóstico 20/08/2022

La entrega de hoy es excepcional, dado que la realizamos un día sábado. Este es el último día en que se pueden hacer públicos resultados de encuestas, por lo que nuestra entrega de hoy busca, justamente, comparar nuestros resultados con los que entreguen las encuestas antes del período de veda. Si bien este reporte no surge de una encuesta sino de un modelo matemático, desde el SERVEL nos han indicado que el espíritu de la ley electoral indica que este tipo de análisis bien se puede interpretar como un sondeo de opinión y quedará al juicio de tribunales su eventual sanción. En esta entrega realizamos 3 pronósticos con distintos sets de datos: series cada 1 hora desde el 4 de julio hasta el 20 de agosto, series de los últimos 21 días en frecuencias cada 1 hora y finalmente un estudio de datos normalizados desde el 4 de julio hasta el 16 de agosto en series diarias.

Las series diarias se nutren de interacciones en redes en tiempo real, es decir, marcan los registros instantáneos de las interacciones que a partir del muestreo aleatorio que realiza Google indican la relación de interés de un termino en una hora durante el día. Como indica

Simon Rogers de Google, las series en tiempo real muestran las reacciones de las personas a los términos que buscamos en el tiempo. Las series diarias registran esas reacciones, mientras que las series mensuales muestran datos normalizados, más fiables para estudios a largo plazo pero en menor frecuencia. Por un lado, los datos diarios permiten medir climas específicos (en intervalos de semanas), las series mensuales permiten medir impactos sostenidos. Sobre la modelación del voto obligatorio, posiblemente el efecto sobre el muestreo se verá muy encima de las elecciones y el resultado que se presenta aquí es más indicativo de grupos de personas que se informan e interactúan activamente con las opciones disponibles, mientras que quienes votarán aparecerán recién hacia el final de la campaña, pocos días antes del 4 de septiembre 2022. En lo concreto para este reporte, tomamos la serie diaria y dos exploraciones mensuales, respondiendo así a la pregunta:

¿Cuál sería el resultado electoral si el plebiscito de salida fuera mañana?

Para medir el impacto coyuntural del Apruebo sobre el Rechazo, se recogen los datos de los últimos 1 días de interacciones en internet para revisar el interés que despiertan las opciones Apruebo o Rechazo, con filtros de exclusividad para una o para otra (es decir, tomar solo aquellas que hablan de Apruebo o solo aquellas que hablan de Rechazo). El resultado para la serie construida con datos cada 1 hora desde el 4 de julio hasta el 20 de julio es el siguiente:

Origen de datos	Resultado	Observaciones	Sigma2	Mínimo por error estándar	Máximo por error estándar
Modelación series por hora desde el 4 de julio al 20 de agosto	51.73%	906	0.005939	42.65%	60.81%
Modelación series por hora en los últimos 21 días	55.1%	504	0.0049	64,02%	46.18%

Tabla 1. Resultados de modelación comparado entre datos por hora desde el 4 de julio y datos por horas desde el 30 de julio (últimos 21 días).

Si la elección fuese en 24 horas, la opción Apruebo tendría un 51.73% de las preferencias y la opción rechazo un 48.27% para series desde el 4 de julio. En lo estrictamente coyuntural, es decir, el éxito de la campaña de uno o del otro, al medir hoy Sábado los datos, nos dice el modelo que apruebo gana la campaña con un 55.1% de las opciones. En este punto es necesario hacer una observación, que se indica en la siguiente tabla, que fue realizada con datos hasta el día viernes 19 de agosto (ayer):

Origen de datos	Resultado	Observaciones	Sigma2	Mínimo por error estándar	Máximo por error estándar
Modelación series por hora desde el 4 de julio al 19 de agosto	49.17%	882	0.005	39.07%	59.27%
Modelación series por hora en los últimos 21 días HASTA 19 DE AGOSTO	48.13%	504	0.0056	38.84%	57.42%

Tabla 2. Resultados de modelación comparado hasta el día 19 de agosto (ayer).

En solo 1 día, el interés por la opción apruebo logra pasar a la delantera. Los fines de semana se han convertido en los únicos días de la semana en que Apruebo domina las interacciones. Todo el resto de la semana la opción apruebo permite que rechazo crezca, dado que de momento la estrategia principal se ha enfocado en eventos masivos y trabajos puerta a puerta, más que una ofensiva mediática y por redes sociales, donde el rechazo no crece pero sí inunda el espacio virtual. Desde que iniciamos las mediciones vemos este patrón, como indicamos en el informe anterior, donde el apruebo es competitivo con picos de rendimiento los días sábados, mientras el rechazo se mantiene constante en su estrategia el resto de los días. La campaña se puede estar ganando en estos detalles y eso es lo que veremos en el siguiente modelo, basado en series diarias que identifican la ganancia mas estructural de una opción sobre la otra en el tiempo.

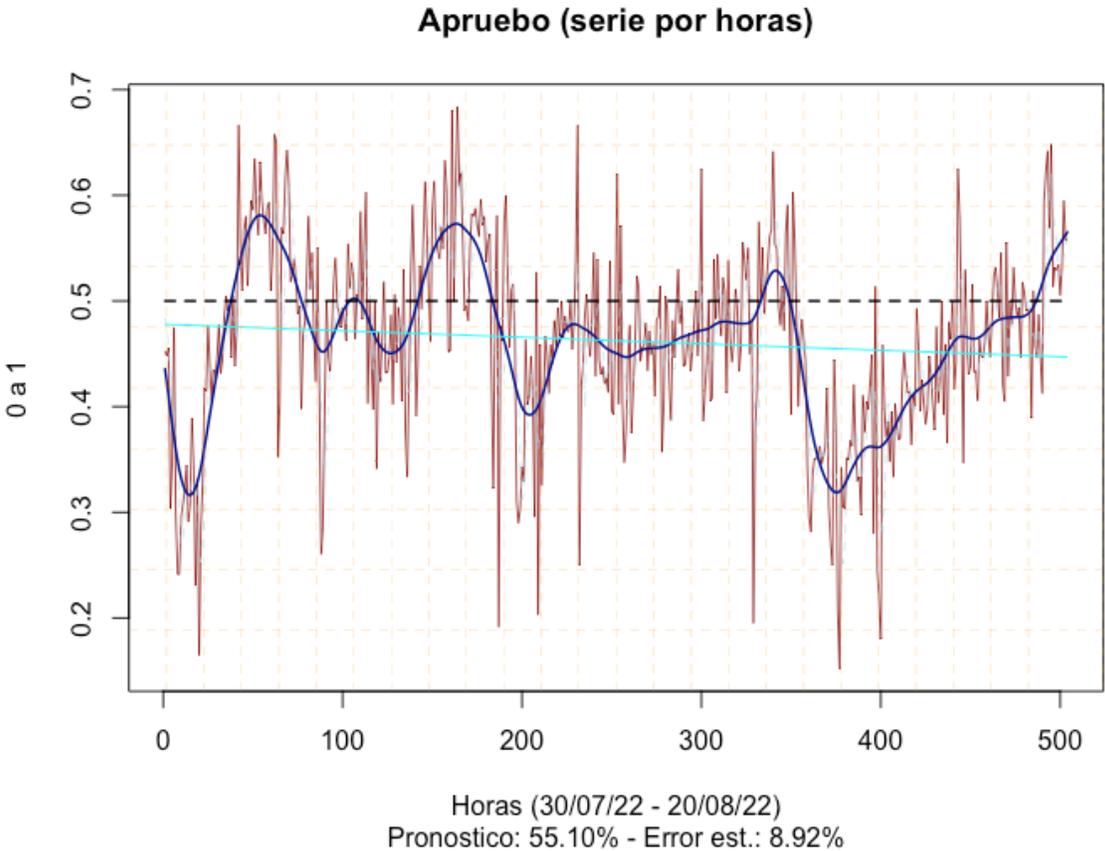


Fig 1. La figura indica la variación del peso del apruebo sobre rechazo entre el 30 de julio 2022 y el 20 de agosto 2022. En rojo las interacciones, en celeste la línea de tendencia por regresión lineal.

En este gráfico, es importante fijarse en la línea de tendencia de regresión en celeste. Si bien el pronóstico para el día domingo indica triunfo del apruebo por un 55.1%, la tendencia va a la baja, es decir, las ventajas que entrega el apruebo en días de semana podrían generar un daño estructural a sus pronósticos, donde los momentos exitosos de apruebo son momentáneos mientras rechazo se mantiene constante.

El pronóstico por series diarias indica un resultado estrecho, con un margen de error del 10.65%. Es decir, se presenta como una elección abierta. Es importante indicar que estos datos solo son hasta el día martes 16 de agosto y dan por perdedora a la opción Apruebo con un 43.19%. Los datos normalizados se entregan desde Google con un atraso de cerca de 4 días, por lo que el auge que muestra los fines de semana la opción apruebo tardan en registrarse en estos conjuntos de datos.

Origen de datos	Pronóstico	Minimo	Maximo	Rango (error estándar)
Modelación con datos entre el 4 de julio 2022 y el 16 de agosto 2022.	43.19%	32.53%	53.84%	10.65%

Tabla 3. Modelación de datos normalizados hasta el martes 16 de agosto.

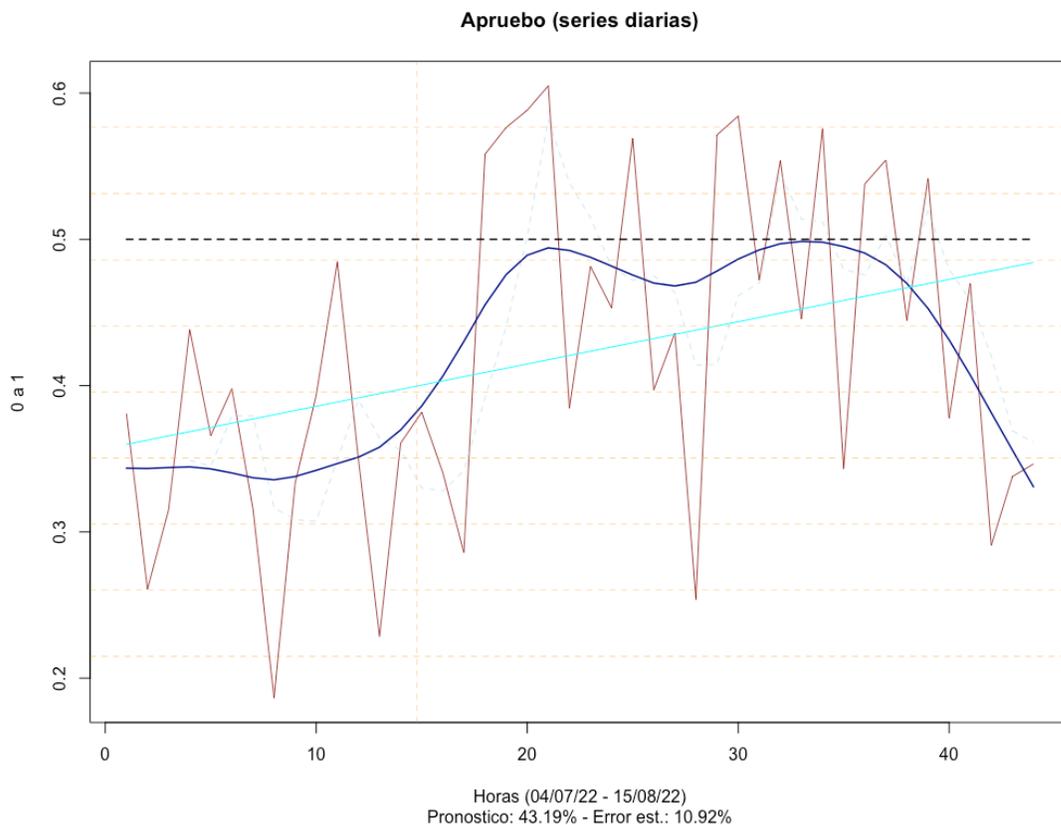


Fig 2. La figura indica la variación del peso del apruebo sobre rechazo entre el 4 de julio 2022 y el 16 de agosto 2022. En rojo las interacciones, en violeta la línea de tendencia suavizada por Hodrick-Prescott.

Si bien el gráfico da por perdedora a la opción Apruebo, la curva de tendencia (celeste) indica un auge sostenido desde el 4 de julio en las series normalizadas, con lo que se puede decir que los resultados de las series por horas presentadas en la tabla 1, podrían prevalecer en lo que queda de campaña y terminar dando por ganadora a la opción aprueba, contrario a lo que indican las encuestas en sus últimas entregas.

Según nuestra experiencia anterior, es con las series mensuales con las que se logra un mejor pronóstico. Para el plebiscito de entrada de 2020, cuando faltaba cerca de 1 mes para las

votaciones, realizamos un pronóstico altamente preciso del resultado final, por eso compartimos este resultado. Sin embargo, han cambiado las condiciones de base: 1. Ya no solo la extrema derecha esta por el rechazo, 2. Esta es una elección con voto obligatorio, 3. En 2020 no había texto constitucional, sino la esperanza de hacer uno y posiblemente una idealización de ese proceso de redacción antes que comenzara, 4. No había el nivel de fake news que hoy circulan, 5. Las encuestas y este método daban por ganador a la opción apruebo, en este caso, el método matemático entrega un resultado contradictorio al de las encuestas.

Si hoy tuviéramos que elegir un pronóstico, basado en el avance de la campaña y tendencias observadas, es más probable que gane Apruebo a que gane Rechazo, como se indica en la figura 2. Nos basamos principalmente en que Google plantea que para investigar la estructura de una tendencia es mejor usar los datos normalizados, dado que permiten revisar si las interacciones se sostienen en el tiempo, mientras que las series diarias registran reacciones de búsqueda más intempestiva y muy referidas a la coyuntura específica de cada día. Si bien hasta el martes 16 de agosto se indica que gana la opción Rechazo, la línea de tendencia indica que la opción apruebo repuntará el fin de semana, como es habitual. Aun así, para uno u otro escenario, los márgenes de error siguen indicando que la elección sigue abierta pero el interés relativo a otras tendencias comienza a ser dominante en el territorio nacional.

Tabla de pronósticos, mínimos y máximos para series por horas desde el 4 de julio

Apruebo	Mínimo	Máximo	Error estándar
0.5415851	0.4646888	0.6184813	0.07689624
0.5373963	0.4568481	0.6179444	0.08054814
0.5348033	0.4524871	0.6171195	0.08231623
0.5324773	0.4486900	0.6162645	0.08378722
0.5302521	0.4451314	0.6153728	0.08512070
0.5281050	0.4417601	0.6144498	0.08634485
0.5260309	0.4385591	0.6135027	0.08747180
0.5240272	0.4355164	0.6125379	0.08851072
0.5220913	0.4326217	0.6115609	0.08946957
0.5202210	0.4298656	0.6105764	0.09035539
0.5184140	0.4272396	0.6095885	0.09117445
0.5166682	0.4247359	0.6086006	0.09193238
0.5149816	0.4223474	0.6076159	0.09263424
0.5133521	0.4200675	0.6066367	0.09328460
0.5117778	0.4178902	0.6056654	0.09388758
0.5102568	0.4158099	0.6047037	0.09444693
0.5087873	0.4138213	0.6037534	0.09496606
0.5073677	0.4119196	0.6028157	0.09544807
0.5059960	0.4101002	0.6018918	0.09589580
0.5046709	0.4083591	0.6009827	0.09631183
0.5033906	0.4066921	0.6000892	0.09669854
0.5021537	0.4050956	0.5992118	0.09705811
0.5009587	0.4035662	0.5983512	0.09739254
0.4998042	0.4021005	0.5975078	0.09770366

Tabla de pronósticos, mínimos y máximos para series por horas desde el 30 de julio

Apruebo	Mínimo	Máximo	Error estándar
0.552033	0.4821751	0.6218909	0.06985788
0.550932	0.4770043	0.6248598	0.07392773
0.550932	0.4753468	0.6265173	0.07558523
0.550932	0.4737249	0.6281392	0.07720716
0.550932	0.4721363	0.6297277	0.07879572
0.550932	0.4705792	0.6312849	0.08035287
0.550932	0.4690516	0.6328124	0.08188041
0.550932	0.4675521	0.6343120	0.08337998
0.550932	0.4660790	0.6357851	0.08485305
0.550932	0.4646311	0.6372330	0.08630098
0.550932	0.4632070	0.6386570	0.08772501
0.550932	0.4618057	0.6400583	0.08912629
0.550932	0.4604262	0.6414379	0.09050588
0.550932	0.4590673	0.6427968	0.09186475
0.550932	0.4577282	0.6441359	0.09320382
0.550932	0.4564081	0.6454559	0.09452391
0.550932	0.4551062	0.6467579	0.09582583
0.550932	0.4538217	0.6480423	0.09711029
0.550932	0.4525541	0.6493100	0.09837798
0.550932	0.4513025	0.6505616	0.09962954
0.550932	0.4500665	0.6517976	0.10086557
0.550932	0.4488454	0.6530187	0.10208664
0.550932	0.4476388	0.6542253	0.10329328
0.550932	0.4464461	0.6554180	0.10448598