

Redes sociales, inteligencia computacional y predicción electoral: el caso de las primarias presidenciales de Chile 2017

Social networks, computational intelligence and electoral prediction: the case of the presidential primaries of Chile 2017

Redes sociais, inteligência computacional e previsão eleitoral: o caso das primárias presidenciais do Chile 2017

PEDRO SANTANDER, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile (pedro.santander@pucv.cl)

CLAUDIO ELÓRTEGUI, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile (claudio.elortegui@pucv.cl)

CRISTIÁN GONZÁLEZ, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile (cristian.gonzalez@pucv.cl)

HÉCTOR ALLENDE, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile (hector.allende@pucv.cl)

WENCESLAO PALMA, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile (wenceslao.palma@pucv.cl)

RESUMEN

Este artículo muestra resultados de una investigación interdisciplinaria aplicada a la capacidad predictiva de las redes sociales, específicamente Twitter, en las primarias legales en Chile de 2017. Mediante la incorporación de inteligencia computacional, se monitoreó la interacción de todos los usuarios chilenos que mencionaron al menos una vez a algunos de los cinco candidatos en competencia para diseñar un modelo de pronóstico que consideró el contexto propio de la comunicación política. Como resultado, se lograron modelos cuyos pronósticos estuvieron bajo 2% en el error absoluto medio (MAE), es decir, con mayor precisión que las encuestas electorales.

Palabras clave: predicción electoral; redes sociales; elecciones primarias; comunicación política; Twitter; inteligencia computacional.

ABSTRACT

This article shows the results of an interdisciplinary research applied to the predictive capacity of social networks, specifically Twitter, in the legal primaries in Chile in 2017. Through the incorporation of computational intelligence, we monitored the interaction of all Chilean users who mentioned at least once some of the five candidates competing to design a forecast model that considered the context of political communication, which delivered results under 2% in the mean absolute error (MAE), with more precision than the electoral polls.

Keywords: electoral prediction; social networks; primary elections; political communication; Twitter; computational intelligence.

RESUMO

O seguinte artigo fornece os resultados de uma pesquisa interdisciplinária aplicada à capacidade preditiva das redes sociais, especificamente o Twitter, nas primárias legais constituídas no Chile durante o ano de 2017. Por meio da incorporação de inteligência computacional, a interação de todos e todos os usuários chilenos que mencionaram pelo menos uma vez a alguns dos cinco candidatos na competição eleitoral, com o objetivo de projetar no modelo metodológico uma previsão que considerou o contexto da comunicação política, apresentando resultados menores de 2% no significância erro absoluto (MAE), com mais precisão do que as pesquisas eleitorais como a principal descoberta.

Palavras-chave: previsão eleitoral; redes sociais; eleições primárias; comunicação política; Twitter; inteligência computacional.

Forma de citar:

Santander, P., Elórtégui, C., González, C., Allende, H. & Palma, W. (2017). Redes sociales, inteligencia computacional y predicción electoral: el caso de las primarias presidenciales de Chile 2017. *Cuadernos.info*, (41), XX-XX. <https://doi.org/10.7764/cdi.41.1218>

INTRODUCCIÓN

El 2016 podría ser recordado como el año en el que fallaron dos clásicas instituciones que las sociedades modernas han creado para tratar de entender y leer los procesos políticos: los medios de comunicación y las encuestas de opinión pública. Fallaron en su capacidad de sondear importantes dinámicas sociopolíticas y, sobre todo, en su capacidad predictiva respecto de acontecimientos de alto impacto.

Esto sucedió ese año con situaciones de relevancia mundial. Los ejemplos son de público conocimiento: las encuestas de opinión no anticiparon la victoria del Brexit, el triunfo del No en el plebiscito de Colombia o el triunfo de Donald Trump en Estados Unidos; ocurrió también en Holanda, en las primarias francesas y en las elecciones municipales en Chile. Respecto de los casos de Gran Bretaña, Colombia y Holanda, los pronósticos de las encuestadoras resultaron totalmente errados, y en lo relativo a Estados Unidos, si bien Hillary Clinton ganó en el voto popular, las encuestas y los analistas que se basaban en ellas no pudieron desentrañar el comportamiento electoral de estados que a la postre fueron ganados por Trump, y le dieron los representantes suficientes para ser electo presidente en el colegio electoral. En cuanto a Chile, ninguna encuesta logró anticipar que la abstención electoral en las municipales 2016 bordearía 70%, haciendo de esta la elección menos participativa de la historia.

No solo las encuestas, también los medios de comunicación, demostraron no ser una herramienta adecuada para la lectura social, y menos para anticipar lo que podría suceder en los casos señalados. Como también es de público conocimiento, los medios no fueron imparciales en estas contiendas: apoyaron mayoritariamente a Clinton, al Sí en Colombia o a la permanencia de Gran Bretaña en la Unión Europea. La campaña negativa o de contraste en medios tradicionales no le hizo el daño esperado a Trump (lo que abre la posibilidad de reevaluar científicamente los efectos de los medios sobre las audiencias). Mientras a Clinton la respaldaron 243 diarios y 148 semanarios, Trump solo recibió el apoyo de 20 diarios y un semanario. La proporción fue de 18 a 1 (Boczkowski, 2016).

Sin embargo, el comportamiento de los usuarios de redes sociales (RR.SS., en adelante) mostraba otra realidad: cuatro días antes de las elecciones generales en Estados Unidos, la página de Facebook (FB) de Trump acumulaba 11,9 millones me gusta y su cuenta de Twitter contaba con 12,9 millones de seguidores, mientras que el número de Clinton fue de 7,8 millones de me

gusta y 10,1 millones de seguidores. En ese momento Clinton tenía 35% menos me gusta en Facebook y 27% menos seguidores en Twitter que su contrincante.

Lo anterior nos hace pensar en la posibilidad –incluso hipótesis– de que, a pesar de los grandes despliegues técnicos y económicos, los esfuerzos humanos y las capacidades profesionales de los medios tradicionales para hacer cobertura de eventos electorales, estos probablemente cumplirán en adelante un papel secundario en comparación con las prácticas de comunicación más centrales que conectarán a los líderes políticos con la ciudadanía en y mediante las redes sociales.

Parece que hay una intención de voto que permanece oculta. Las situaciones señaladas demuestran que los electores no contestan con veracidad o que son recelosos de compartir en público sus preferencias y, al parecer, los estudios demoscópicos tradicionales tienen dificultades crecientes para realizar predicciones de voto; a su vez, los medios tradicionales son los mismos que difunden dichas encuestas.

Ocurre que a veces el electorado oculta su intención, sea porque se niega a contestar la encuesta, porque dice que votará por una opción distinta a la que finalmente escoge o porque –de acuerdo al criterio de deseabilidad social– dice lo que supone es más aceptado y correcto socialmente. Esto pareciera ser más frecuente en situaciones en que los candidatos no son apoyados por los medios de referencia, como Donald Trump o el Brexit, opciones rechazadas por la élite. Asimismo, ante definiciones radicalmente antagónicas, pareciera ser que aumenta el voto oculto y, por lo mismo, esas voces, aun siendo mayoritarias, son silenciosas, y no se detectan anticipadamente ni en los estudios de opinión ni leyendo la agenda mediática.

Al respecto, la teoría de la espiral del silencio (Noëlle-Neuman, 1995) plantea la adaptación de las personas a lo que consideran los climas de opinión predominantes, como una forma de evitar el aislamiento frente a posturas contrarias y que asumen que son las mayoritarias/hegemónicas. Las personas estarían sondeando permanentemente la relación entre sus opiniones y las del espacio público, tendiendo a enmudecer si presentan posiciones diferentes a la (supuesta) mayoría, incluso a invisibilizarse cuando se enfrentan a los núcleos duros que, aunque pueden estar integrados por pocos, plantean con insistencia sus opiniones, por ejemplo, gracias a la cobertura medial de la que gozan. Los medios suelen ser una fuente fundamental de información para que las personas sepan cuáles son las opiniones hegemónicas. El fenómeno se califica

como espiral del silencio, pues cuanto más los medios difunden la versión dominante, más guardarán silencio las voces individuales contrarias, en un proceso en espiral (Noëlle-Neuman, 1995). Probablemente, el fracaso en la capacidad predictiva de las últimas encuestas en procesos políticos de trascendencia se vincule con este fenómeno.

DISCUSIÓN CONCEPTUAL

En este contexto, el llamado entorno digital (Jenkins, 2006) y las RR.SS. introducen una complejidad adicional. Las RR.SS. actúan con otra velocidad e influyen en el sistema tradicional de medios (Campos, 2008), lo que puede favorecer que los climas de opinión sean más cambiantes y difíciles de medir, pues se aprecia un nivel mayor de fragmentación y movilidad (Feijóo, Maghiros, Abadie, & Gómez Barroso, 2009). En este entorno, las RR.SS. generan dinámicas que afectan el modo en el que los ciudadanos se relacionan con la información, pero también cómo se genera la información por parte de los actores políticos y cómo esta es canalizada por los medios (Van Dijck, 2016; Wescott, 2008). Ello significa que deberíamos considerar a las RR.SS. como una gran fuente de datos que, teóricamente, pueden proporcionarnos información acerca de muchas preferencias, opiniones, gustos, estados de ánimo, entre otros, de los usuarios (Aguado & Navarro, 2013). De hecho, hoy las RR.SS. pueden ser consideradas como uno de los principales canales para recoger y difundir información política y comercial (Fábrega & Paredes, 2012; García, 2015), así como para la elaboración de mensajes micro-segmentados (Sánchez Medero, 2016).

Este contexto en el que lo analógico se interrelaciona con lo digital genera inquietudes y preguntas nuevas, abre campos inexplorados, todo lo cual motiva nuestra investigación. Sabemos que existe un escenario dinámico en el cual los sondeos basados en mediciones estadísticas y los medios de comunicación se ven afectados (Harvey, 2013). Clásicamente, se ha sostenido que los sondeos ayudan a comprender, orientar y coordinar las acciones de los individuos en entornos complejos (Lippmann, 1922; McCombs & Shaw, 1972; McQuail, 1991; Thompson, 1998). En ese marco, se ha destacado su eficacia para realizar lectura macrosocial y para operar como un termómetro social. Las encuestas, por su parte, ofrecerían la posibilidad de conocer estados de opinión, preferencias electorales, gustos, etc., es decir, aquello que muchos autores denominan opinión pública (Habermas, 1986; Price, 1994). Los medios, a

su vez, darían cuenta de las preocupaciones que tienen las audiencias y la élite, como de los principales temas que se debaten, y serían, por lo mismo, una herramienta socio-discursiva adecuada para orientar a especialistas y a amplios segmentos del público en torno a tópicos de importancia nacional e internacional (McCombs & Shaw, 1972). Esto explica, por ejemplo, el gran interés académico y la variedad de procedimientos analíticos que existen en torno al discurso de los medios (análisis del discurso, análisis ideológico, análisis semiótico, análisis de contenido de los medios, etc.).

Estas premisas, hipotetizadas y verificadas a lo largo del siglo XX, están siendo puestas a prueba en la actualidad y, a nuestro entender, las certezas que en las ciencias sociales había en relación a ellas se encuentran en una situación de inestabilidad. Estos postulados se enfrentan a un nuevo contexto —el ecosistema mediático-digital (Guzmán & Sánchez Medero, 2016; Scolari, 2015)— y el desarrollo de la inteligencia artificial (IA). Los cambios sobre el entorno comunicacional en el contexto digital han impactado fuertemente en el desarrollo de nuevas formas de expresión de la política. Para Castells (2015), por ejemplo, el cambio fundamental en las comunicaciones de los últimos años es el uso de Internet y de las redes inalámbricas como plataformas de comunicación digital (Castells, 2009). Para los ciudadanos, se hace posible un escenario de mayor interacción con otros desde plataformas dispuestas al contacto micro-comunicativo y, a su vez, masivo. El sentido y la vivencia de la política se despliegan también en lo *online*, en un entorno móvil, multidispositivo, con accesos inéditos al contenido y a los servicios (Scolari, 2013).

El nuevo entorno medial y la IA desafían a las encuestas de opinión tradicionales, generalmente hechas cara a cara o telefónicamente con tamaños muestrales, sesgos y procedimientos propios de un contexto analógico que conlleva instrumentos de medida que trabajan con variables continuas y magnitudes limitadas, además, en situaciones a menudo no naturales. El desarrollo de la IA (Russell, Norvig, & Davis, 2016) y el altísimo y continuo uso que los ciudadanos hacen de las tecnologías de la información y de las comunicaciones (TIC) a escala global permiten postular, a modo de supuesto conceptual, el inicio de una nueva era en la comunicación digital que comienza a caracterizar el siglo XXI, en cuyo centro ya no está lo industrial, que es algo propio del desarrollo económico del siglo XX, sino la generación de conocimiento basado en la IA.

Los formatos de contenido político no son una

excepción, pues están en constante proceso de transformación, resultado inestable de la confluencia entre formatos adaptados de otros medios y la innovación específica de la movilidad, fusión entre contenido y comunicación (Castells, 2015; Cotarelo, 2013). La pérdida del privilegio de la información por parte de los medios tradicionales tiene fuertes consecuencias para la comunicación política, sobre todo en momentos de elecciones (Rogers, 2004), por lo que la pérdida de una posición central de los medios tradicionales da lugar a un escenario nuevo en el que las predicciones tradicionales de la política pierden fuerza (Gutiérrez-Rubí, 2011). En este marco, se observa un traslado creciente de la actividad política hacia las campañas digitales a nivel global (Issenberg, 2012). Si la socialización política y su expresividad cambian, las formas de medir esas percepciones también deberían hacerlo. En ese ámbito se inscribe justamente nuestra investigación.

LEER LAS REDES, LEER LA SOCIEDAD: USO Y ESTUDIO DE LAS RR.SS. EN CAMPAÑAS

Como a menudo ocurre, el conocimiento científico y la elaboración de teoría no avanzan a la misma velocidad que los hechos extra-científicos. Es así como las RR.SS. y, en general, las TIC están siendo empleadas hace ya más de una década en el campo político, especialmente en campañas electorales. En ese sentido, la trazabilidad digital de los usuarios que, a la vez son potenciales electores, se ha convertido en una herramienta poderosa para la indagación política. El rastro digital que dejamos se ve afecto a una verdadera minería de datos que permite obtener información acerca de gustos, comportamientos, relaciones (Gutiérrez-Rubi, 2011; Issenberg, 2012). Hoy vemos ambientes que proporcionan datos reales a escalas sin precedentes. Es lo que se conoce como *Big Data*, que implica que lo que hacemos deja una huella que puede ser recolectada y procesada gracias a la IA (Provost & Fawcett, 2013).

Existe hoy la capacidad de acceder a inmensos volúmenes de datos que además muchas veces son muy personales, incluso íntimos. A menudo son proporcionados de manera voluntaria por los usuarios, lo que convierte a las RR.SS. en herramientas con potencial de sondeo permanente y capacidad para ofrecer segmentadamente a cada usuario diversos mensajes. Facebook (FB) es un buen ejemplo de esto.

Por lo mismo, las RR.SS. se están convirtiendo crecientemente en una herramienta para responder a los desafíos políticos de campaña, por un lado, debido a las imprecisiones que las encuestas tradicionales muestran

y, por otro, por su capacidad de llegada hiper segmentada a las personas. Esas potencialidades, basadas en un monitoreo computacional del comportamiento de los usuarios, hace que se tenga una creciente confianza en la capacidad política y demoscópica de las RR.SS. (Magnani, 2017).

El primero en hacer un uso estratégico de alto impacto de herramientas digitales y del *Big Data* en campaña electoral fue Barack Obama, en 2008. Desde ese momento, importantes universidades y centros de investigación comenzaron a explorar el modo en que las RR.SS. se relacionan con los votantes, sobre todo FB, Twitter y YouTube (Deltell, Claes, & Osteso, 2013). Cuatro años después, en 2012, la campaña de Obama trabajó con una base de datos de 16 millones de perfiles de indecisos que permitió clasificarlos individualmente, para luego operar comunicacionalmente sobre ellos. En las elecciones de 2016, la sofisticación tecnológica aumentó y el equipo electoral de Trump pudo, durante el tercer debate televisivo que mantuvo con Clinton, tomar uno de los argumentos planteados en televisión y, por medio del uso de algoritmos, sus asesores crearon 175 mil versiones de dicho mensaje, los que fueron enviados con variaciones y matices de acuerdo al perfil de cada usuario (Hilbert, 2017).

Hoy estas lógicas de filtrado, basadas en algoritmos, se están usando de manera regular, con intencionalidad política, generando lo que algunos autores denominan una nueva “economía política de control de la información” (Ochigame & Holston, 2016, p. 94).

En consecuencia, hoy no se concibe una campaña electoral sin el uso de las RR.SS. Y si bien ya no hay dudas en el campo político-comunicacional acerca de la centralidad de las redes—por ejemplo, los bots para la comunicación política, (Gutiérrez-Rubi, 2011), la crucial relación entre nodos y *spread influence* (Riquelme & González-Cantergiani, 2016) o *feeds* y las categorizaciones algorítmicas de millones de usuarios en la red (Ceron, Curini, & Iacus, 2017)—el conocimiento respecto de la capacidad predictiva de las RR.SS. en situaciones de campaña se encuentra en una fase más experimental y exploratoria. De acuerdo a Schoen, Gayo-Avello, Takis Metaxas, Mustafaraj, Strohmaier y Gloor (2013), esta reside en la inmensa cantidad de datos provistos por las RR.SS., el factor expansivo de su comportamiento y la capacidad tecnológica de procesamiento. Asimismo, en el hecho de que las RR.SS. permiten una aproximación empírica no intrusiva a las conductas, preferencias y discursos de los usuarios, pues en ellas las opiniones se manifiestan de

forma libre, con el deseo de darlas a conocer. A ello se suma que la recolección de datos permite mantener una adecuada codificación de las respuestas gracias al desarrollo tecnológico (Tumasjan, Sprenger, Sandner, & Welp, 2010). Twitter, por ejemplo, actúa como una sonda, ya que permite seguir y medir opiniones de forma continua en extensos periodos. A diferencia de las tradicionales encuestas, las RR.SS. se caracterizan por constituir redes de opinión voluntarias. Dicho de otro modo, entre los encuestados puede no haber vínculo personal ni conocimiento mutuo; es el investigador quien construye el grupo a partir de criterios sociodemográficos. El estudio de las RR.SS., en tanto, permite estudiar al ciudadano como un actor que adhiere a ciertas figuras y proyectos y que está dispuesto a actuar como un agente reproductor en el proceso de difusión. Entonces, lo que se analiza no solo es la opinión de un individuo particular, sino el flujo discursivo del cual participa. Esto permite tener una visión más dinámica y sofisticada de la ciudadanía y, en particular, de su relación con la clase política (Wescott, 2008).

En esa línea, podemos observar que, sobre todo en países de Europa central y en Estados Unidos, se ha comenzado a testear científicamente la capacidad predictiva de las RR.SS., especialmente en contextos electorales y también en estudios de mercado (Taylor & Lethman, 2016). Ya no solamente se trata del uso de la IA para la comunicación política, se avanza un paso y se explora la red para predecir resultados electorales (Schoen et al., 2013). Esto último está basado en la hipótesis de correspondencia, que supone tendencias coincidentes entre declaraciones positivas hechas por los usuarios de redes respecto de partidos o candidatos y votos efectivamente obtenidos por éstos (Aparaschivei, 2010; Deltell, 2012; Tumasjan et al., 2010).

Bajo el contexto de dicha hipótesis se ha comenzado a explorar la relación entre RR.SS. y elecciones. Por ejemplo, en las elecciones alemanas de 2009 (Tumasjan et al., 2010), en las elecciones catalanas (Congosto, Fernández, & Moro, 2011), en las suecas (Larson & Moe, 2011), en las elecciones andaluzas de 2012 (Deltell, Claes, & Osteso, 2013) o la rumanas (Momoc, 2012). Estas investigaciones se centran mayoritariamente en el uso de Twitter que hacen los usuarios, correlacionándose dicho comportamiento con datos efectivos de las elecciones.

Este estudio del campo es reciente; a partir de 2007 comienzan los primeros intentos (Ceron et al., 2017), es decir, se trata de una dimensión investigativa aún en su

infancia. Es, además, un área muy original, pues obliga a un cruce entre ciencias computacionales, comunicación política y lingüística. Aún hay mucho que explorar respecto del potencial y del poder predictivo de las RR.SS., así como de sus limitaciones y de su aplicabilidad general en áreas como la comunicación política y las campañas electorales. En ese sentido, hay aún un amplio campo de estudio que en Chile permanece poco explorado.

DISCUSIÓN METODOLÓGICA

Con la base conceptual señalada, nuestra investigación quiso explorar la capacidad predictiva de las RR.SS. en relación con las primarias, que son las segundas oficiales —es decir, supervisadas por el Servicio Electoral— que se realizaban en Chile. Estas tuvieron lugar el domingo 2 de julio de 2017 y concurrieron cinco candidatos, tres por la alianza de derecha Chile Vamos (Sebastián Piñera, Manuel José Ossandón y Felipe Kast) y dos por la alianza de izquierda Frente Amplio (Beatriz Sánchez y Alberto Mayol).

Dado el carácter exploratorio ya mencionado, no optamos por una hipótesis, sino por una pregunta de investigación y un objetivo general, elaborados colectivamente por un grupo interdisciplinario de académicos pertenecientes a la lingüística, la ingeniería informática y el periodismo. La pregunta guía fue: ¿Qué algoritmos de inteligencia computacional que monitoreen la interacción de los usuarios de Twitter permiten predecir con mayor precisión los resultados electorales de las primarias presidenciales chilenas? Con base en esta pregunta, se consensuó el siguiente objetivo general: evaluar la capacidad predictiva de diversas métricas aplicadas a la red social Twitter, para predecir con la mayor precisión posible los resultados electorales de las primarias, mediante la construcción de un modelo de pronóstico que considere elementos de la interacción entre usuarios y del contexto social, empleando técnicas de inteligencia computacional.

Esta labor fue abordada interdisciplinariamente por el equipo de investigadores, al que se sumó un equipo de tres profesionales de apoyo (dos ingenieros informáticos y una periodista especializada en redes sociales), y cinco ayudantes que clasificaron de manera manual.

A partir del 14 de mayo de 2017, se trabajó con una reconocida empresa de monitoreo de *social media* para extraer los tuits de la red social y se diseñó una base de datos con los mensajes y metadatos de todos los usuarios

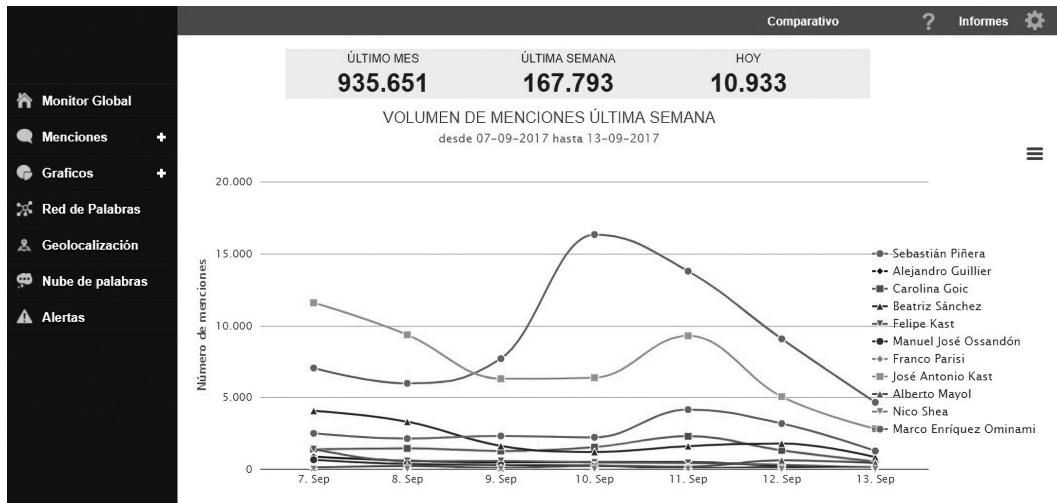


Imagen 1. Plataforma con bases de datos de Twitter y del proyecto de investigación

Fuente: Elaboración propia.

chilenos que mencionaran al menos una vez a alguno de los cinco candidatos. Metadatos son, por ejemplo:

- Nombre de usuario que emitió el tuit.
- Fecha de publicación del mensaje.
- Cantidad de seguidores del usuario.
- Cantidad de personas a las que el usuario sigue.
- Aplicación utilizada para generar el tuit (Android, Facebook, HootSuite, etc.).
- Cantidad total de mensajes emitidos por el usuario a la fecha en la que publicó el tuit.

Se guardan estos mensajes en bases de la propiedad de la empresa y el equipo de investigadores los rescata mediante un algoritmo de recolección de datos, para luego hacer un respaldo de los mensajes y de los metadatos en nuestros servidores (imagen 1).

Esa información fue procesada y analizada por nuestro equipo desde el 14 de mayo al 1 de julio, un día antes de las elecciones, y con ella se elaboró una gran cantidad de gráficos y tablas en un sitio propio.

Una vez recolectados, comenzó el proceso de clasificación. Aquí entramos a la etapa denominada análisis de sentimiento, tarea que —mediante el procesamiento

computacional del lenguaje natural— permite clasificar comentarios de usuarios (Bakliwal, Foster, van der Puil, O'Brien, Tounsi, & Hughes, 2013). Inicialmente, se encargó a los ayudantes que durante mayo y junio clasificaran manualmente los miles de tuits que se generaban en los momentos *peak*, o sea, de mayor actividad en las RR.SS. La clasificación manual siguió el criterio positivo/negativo/neutro, es decir, catalogando las opiniones en polaridades, guiados por la hipótesis de correspondencia ya mencionada (Aparaschivei, 2010; Deltell, Claes, & Osteso, 2013; Tumasjan et al., 2010). Para ello, se elaboró un libro de códigos que orientó a los ayudantes; por ejemplo, definimos que un post expresa una intención de voto a favor de un candidato solo si se satisface al menos una de las tres siguientes condiciones:

1. El post incluye una declaración explícita relacionada con la disposición a votar para un candidato/partido. Ejemplo: En las primarias voy a votar por @mjossandon #PrimariasPresidenciales #Chile2017.
2. El post incluye una declaración a favor de un candidato/partido, junto con un mensaje o un hashtag conectado a la campaña electoral de ese candidato/partido. Ejemplo: Beatriz Sánchez es una cara nueva que limpiará la política #ConfianzaQueCambiaChile.

3. El post incluye una declaración negativa que se opone a un candidato/partido con un mensaje o un hashtag conectado a la campaña electoral de un candidato/partido rival. Ejemplo: Me carga lo sinvergüenza que es @sebastianpinera. Vamos todos por @fkast #FueraPiñera.

Con estos criterios, los ayudantes etiquetaron manualmente 158 mil tuits (9,7% del total). Este trabajo de catalogación fue la base para las técnicas de aprendizaje automático (*machine learning*) necesarias para generar modelos predictivos (Sebastiani, 2002). Es decir, los criterios de etiquetación humana deben ser incorporados por las máquinas (Pang, Lee, & Vaithyanathan, 2002) gracias al uso de algoritmos de aprendizaje supervisados y de algoritmos de imitación, en el entendido de que se trata de programas que, con el insumo proporcionado (los datos de la clasificación humana), logran entender, mediante criterios estadísticos-matemáticos, los patrones existentes entre datos (mensajes, texto) y etiquetas (positivo/negativo/neutro) y luego discriminarlas automáticamente. La finalidad es que se imite el comportamiento humano al clasificar. Es por esto que la clasificación manual llevada a cabo fue de suma importancia, tanto por la gran cantidad de datos que se clasificó (158 mil tuits que fueron la base para el entrenamiento de los algoritmos de aprendizaje automático), como por la posibilidad de reducir el sesgo clasificatorio.

Para entrenar los algoritmos con los tuits clasificados manualmente, se empleó una programación algorítmica que transforma dichos datos, mediante técnicas de procesamiento de lenguaje natural, en una representación vectorial. Esta representación, denominada bolsa de palabras (*bag of words*) (Zhang, Jin, & Zhou, 2010) traduce los tuits a un lenguaje que las máquinas entienden, y esa traducción se usa como entrada de datos, para que los sepan interpretar (en polaridades) y usarlos (predictivamente). Con los algoritmos ya parametrizados y entrenados, generamos nuestros propios clasificadores para el ejercicio predictivo (que denominamos C1 y C2, ver gráfico 5). Cabe señalar que entendemos un clasificador como un algoritmo ya entrenado con los datos relevantes al ejercicio.

Con toda la base metodológica y conceptual mencionada, se procedió a realizar ejercicios predictivos a partir del 18 de junio hasta el 1 de julio. Es importante señalar que, para esas mediciones, se probaron distintas ventanas, es decir, rangos temporales entre un día X y el día antes de las elecciones. Por ejemplo, una ventana

de tamaño 5 para pronosticar el resultado de las primarias del 2 de julio implica que los datos a recolectar y a clasificar son todos los tuits generados desde el 28 de junio hasta el 1 de julio. Por lo tanto, el ejercicio predictivo debe considerar y hacer trabajar en conjunto a tres variables: los clasificadores (C1 y C2), el rango de ventana en que se recolectan y procesan los datos, y el mejor día a partir del cual se emiten las ventanas (para nosotros fue el 27 de junio). Estas tres variables se cruzan y se comparan luego con el resultado electoral real.

El margen de error que se produce con la predicción se denomina error absoluto medio (MAE, por sus siglas en inglés), métrica utilizada para ver el rendimiento en las predicciones y que da cuenta de la diferencia promedio entre las estimaciones obtenidas por los clasificadores utilizados y los resultados reales (Willmott & Matsuura, 2005). Mientras más próximo sea el MAE a 0, más cercano al resultado real de las elecciones. Esta métrica se calcula con la siguiente fórmula:

$$MAE = \frac{\sum_{i=1}^n \hat{y}_i - y_i}{n}$$

En ella, y corresponde al valor real de un evento i , \hat{y} corresponde al valor de la estimación realizada para un evento i , y finalmente n corresponde a la cantidad de eventos existentes a estimar. Un MAE bajo 5 califica como aceptable, al igual que ocurre en las métricas de las encuestas tradicionales (Birmingham & Smeaton, 2011).

Para enfrentar la siempre posible actividad de bots en campañas electorales, partíamos de la desventaja de contar con tuits publicados previamente en el contexto de primarias o de las elecciones presidenciales chilenas. Para suplir aquella falta de conocimiento, se utilizaron datos (Varol, Ferrara, Davis, Menczer, & Flammini, 2017) de investigaciones sobre la caracterización y detección de *bots*, extrayéndose características para detectar su presencia, agrupadas de la siguiente manera:

- a. Basadas en el usuario: el número de amigos y de seguidores, descripción del perfil y la configuración, número de tuits y retuits producidos, entre otros.
- b. Basadas en el tiempo: aquí se considera el tiempo entre tuits y retuits consecutivos.

- c. Basadas en la red: la comunicación entre usuarios de Twitter a través de retuits, menciones y *hashtag* construye una red, lo que permite caracterizar el tipo de comunicación entre los usuarios.

Luego, se tomaron todos los tuits generados durante el período de estudio y se filtraron las características antes mencionadas. Finalmente, se realizó un proceso llamado transferencia de conocimiento (del inglés *transfer learning*), que implica que el conocimiento generado para detectar *bots* en estudios previos se utiliza para detectar *bots* en las elecciones chilenas. No se detectó la presencia de *bots* durante el período comprendido entre el 14 de mayo y el 1 de julio de 2017.

RESULTADOS

En total, 162 mil usuarios comentaron en el rango de fecha señalado al menos una vez algo acerca de los candidatos, generándose un total de 1.619.631 mensajes. Esta cifra incluye tuits (30,32% del total de mensajes) y retuits (69,68%). Los retuits (RT) doblaron a los tuits, en todos los candidatos (gráfico 2).

El candidato con mayor cantidad de menciones (tuits y RT) fue Sebastián Piñera: 609.887 post (38,1%) se refirieron a él, seguido por Beatriz Sánchez, quien se llevó casi un cuarto de las menciones (gráfico 3).

Considerando que, tal como se menciona en la formulación del objetivo general, nos interesaba relacionar la actividad de las RR.SS. con el contexto-país, optamos por hacer –en el mismo período– un seguimiento de hechos mediáticos (debate y entrevistas a candidatos) y eventos no mediáticos impulsados por las propias campañas (proclamaciones, concentraciones, inscripciones, etc.). Partimos del supuesto conceptual de que el comportamiento de las curvas en Twitter a menudo está relacionado con situaciones que ocurren fuera de la red, ya sea en medios tradicionales o en eventos políticos. Por lo mismo, elaboramos cuadros (gráfico 4) por coalición, con sus respectivos candidatos, que cruzan la cantidad de menciones (eje Y) con los sucesos que en esas fechas ocurrían (eje X).

Claramente, los *peaks* en Twitter están muy relacionados con ocurrencias en los medios tradicionales, sobre todo en televisión abierta. Por ejemplo, la mayor cantidad de mensajes que se generaron en un solo día mencionando a Beatriz Sánchez (29.371) ocurrieron el 29 de mayo, cuando asistió al programa político Tolerancia Cero, en horario *prime time* del domingo. Para el caso de los tres candidatos de la derecha, eso

sucede el 26 de junio, cuando los canales de la televisión abierta, agrupados en la Asociación Nacional de Televisión de Chile (Anatel), organizaron un debate en cadena nacional. Esa noche, se generaron 41.083 mensajes en torno a Piñera, 32.211 en torno a Manuel José Ossandón y 25.615 en torno a Felipe Kast; son las cifras diarias más altas en todo el rango temporal medido. Es decir, para nuestro caso, comprobamos que los ciclos de actividad en las RR.SS. dependían claramente de lo que ocurriera en los medios tradicionales. Los ciclos de relativa calma o normalidad eran interrumpidos cuando en los medios se generaba alguna actividad importante relacionada con los candidatos, sobre todo entrevistas y debates. La dependencia, en ese sentido, de los *peaks* en Twitter respecto de los eventos mediáticos es clara. Solo una vez, el 20 de junio, un *peak* no dependió de los medios tradicionales. Fue cuando el candidato de la derecha, Sebastián Piñera, realizó un chiste misógino en un acto de campaña, que fue grabado por el celular de alguien y viralizado.

En cuanto a la capacidad predictiva demostrada por los clasificadores (C1 y C2) en relación con los resultados de las primarias, podemos concluir que, como muestra el gráfico 5, para el caso de Chile Vamos, el clasificador 1 (C1) mostró un MAE de 1,6% el 27 de junio, con una ventana de 11 días. Ese día, con ese rango, el clasificador predijo el siguiente resultado en las elecciones primarias de la derecha: Sebastián Piñera: 59,6%; Manuel José Ossandón: 23,7%; Felipe Kast: 16,7% (resultados reales: SP: 58,4%; MJO: 26,2%; FK: 15,4%). El C2, por su parte, con igual ventana, tuvo el menor MAE (2,9) el 25 de junio, pronosticando las siguientes cifras: Sebastián Piñera: 55,4%; Manuel José Ossandón: 20,2%; Felipe Kast: 24,3%.

Para el caso del Frente Amplio, el C1 fue el que mostró el mejor rendimiento, también el 27 de junio; con una ventana de 15 días, el error predictivo (MAE) fue de apenas de 0,5 (gráfico 6). El clasificador predijo los siguientes resultados: Beatriz Sánchez: 67,1%; Alberto Mayol, 32,9% (resultado real: BS: 67,6%; AM: 32,4%). En tanto, ese mismo día 27 de junio, con la misma ventana, el C2 generó con un MAE de 0,7 el siguiente pronóstico: Beatriz Sánchez: 66,9%; Alberto Mayol: 33,1%.

Esto nos lleva a concluir que el C1 tuvo un excelente rendimiento en su capacidad predictiva el día 27 de junio, tanto para el Frente Amplio como para la derecha, con un rango de ventana como el señalado. En ese sentido, hacer la predicción cinco días antes de las elecciones parece ser una fecha adecuada, sobre todo pensando en las características políticas de una

Detalle del total de mensajes de candidatos desde 14/05/2017 hasta 01/07/2017

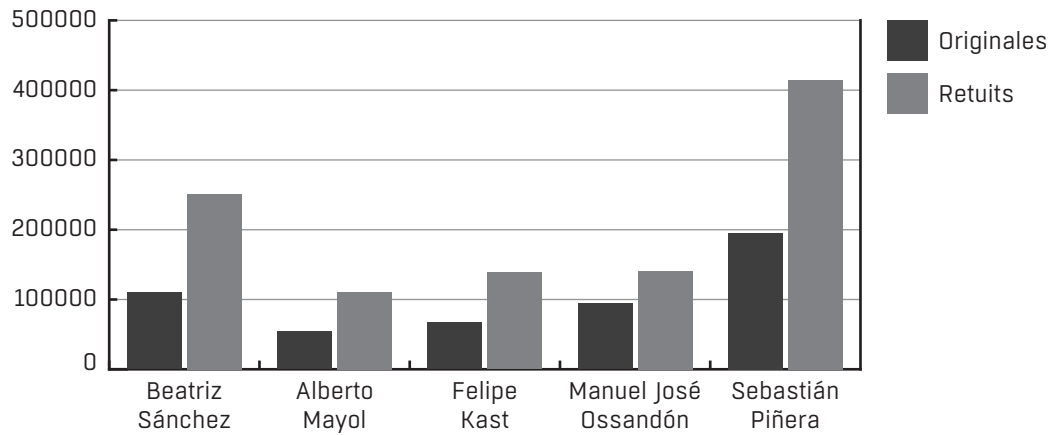


Gráfico 1. Cantidad de mensajes por candidatos

Fuente: Elaboración propia.

Mensajes generados de candidatos desde 14/05/2017 hasta 01/07/2017

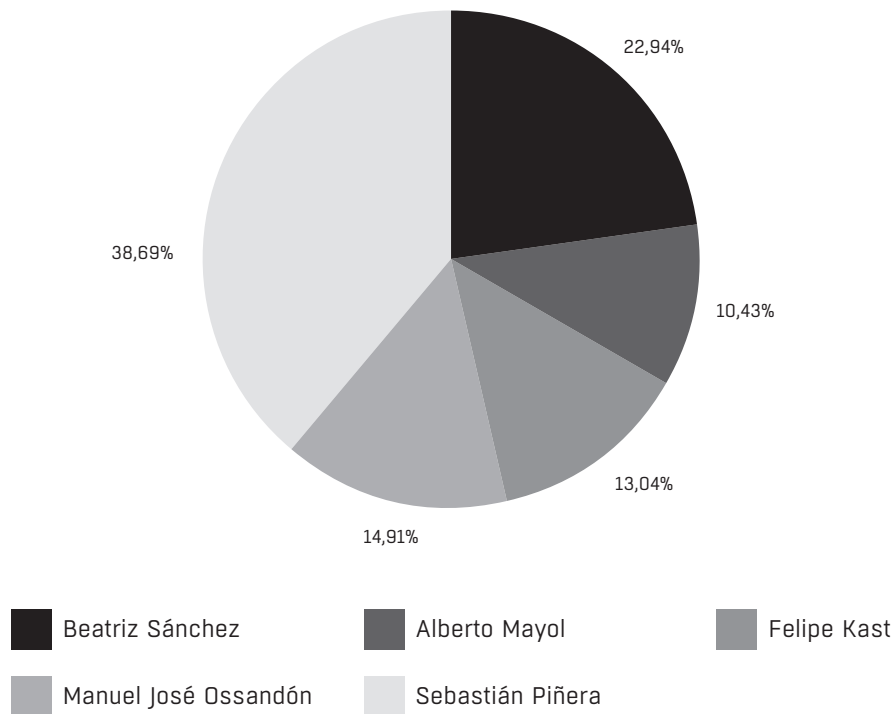


Gráfico 2. Distribución porcentual mensajes de candidaturas

Fuente: Elaboración propia.

Mensajes generados de candidatos desde 14/05/2017 hasta 01/07/2017

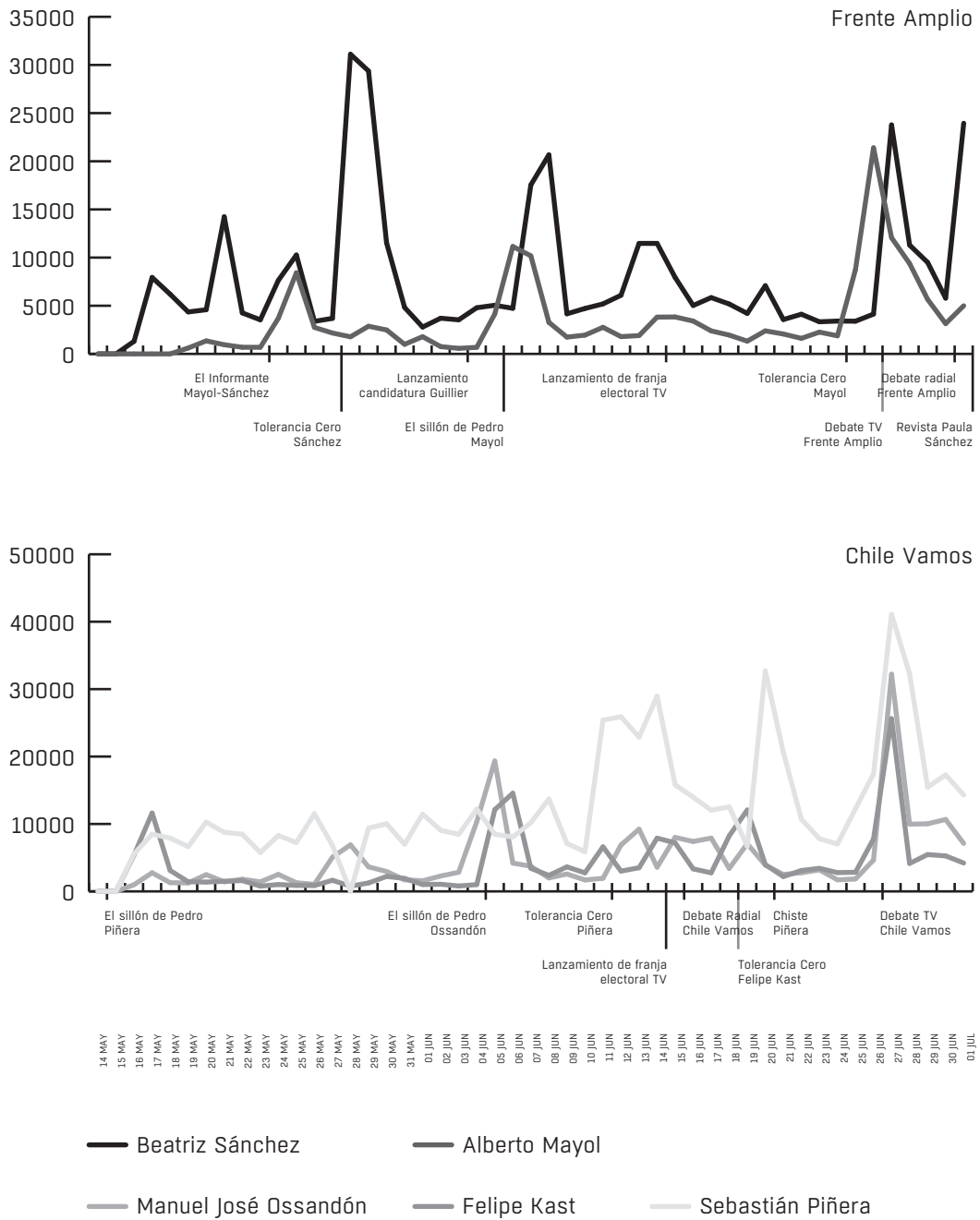


Gráfico 3. Menciones a candidatos y eventos

Fuente: Elaboración propia.

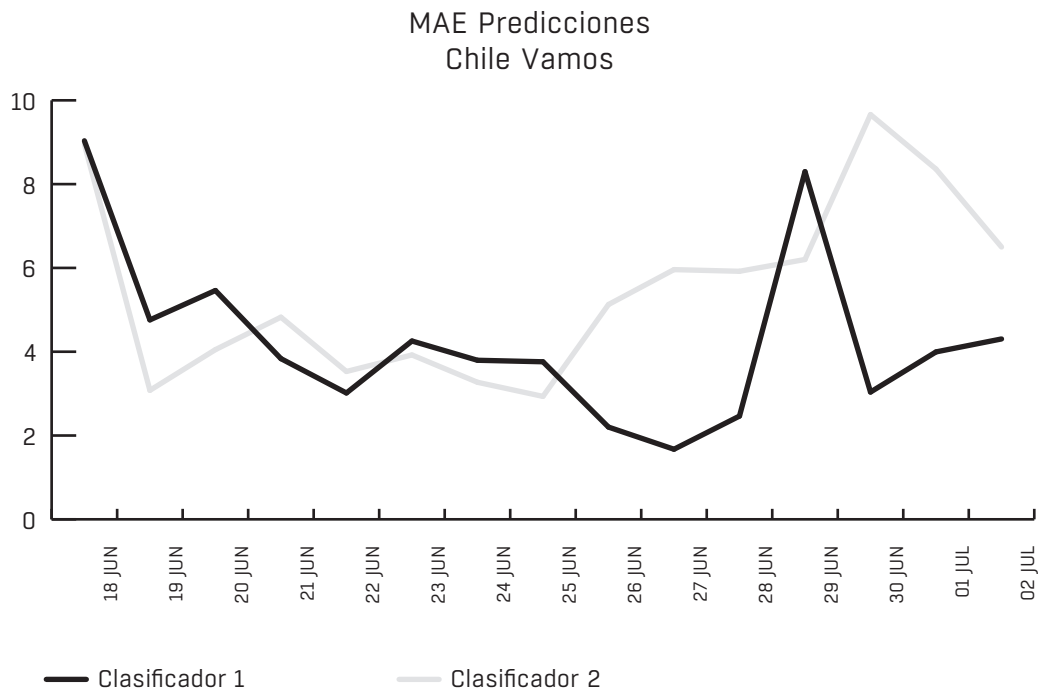


Gráfico 4. Error absoluto medio de clasificadores para Chile Vamos

Fuente: Elaboración propia.

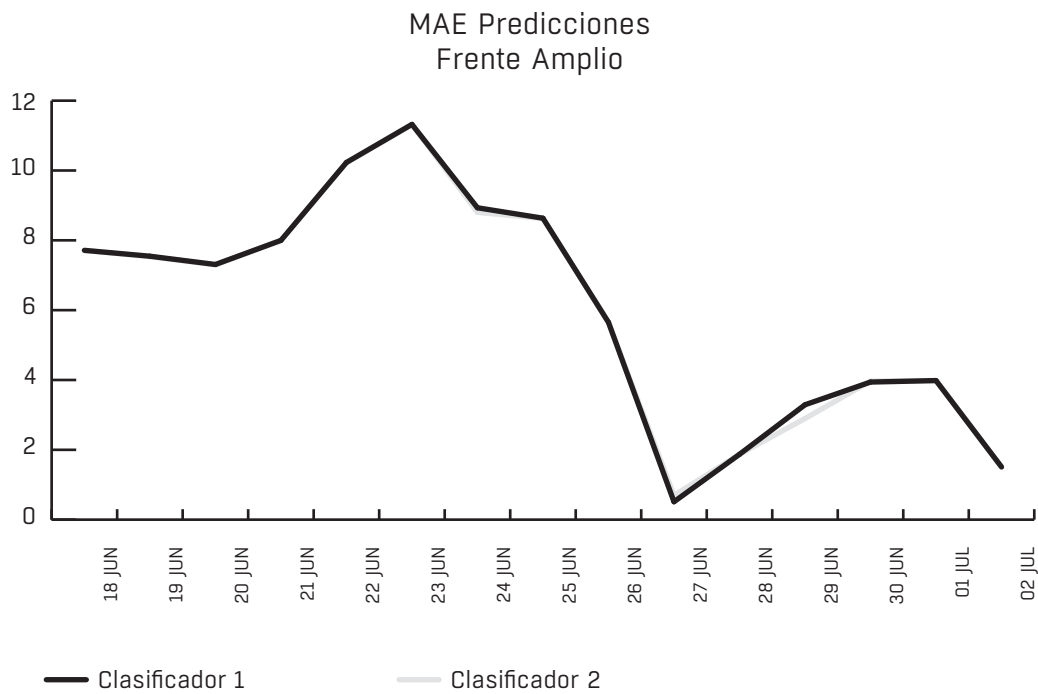


Gráfico 5. Error absoluto medio de clasificadores para Frente Amplio

Fuente: Elaboración propia.

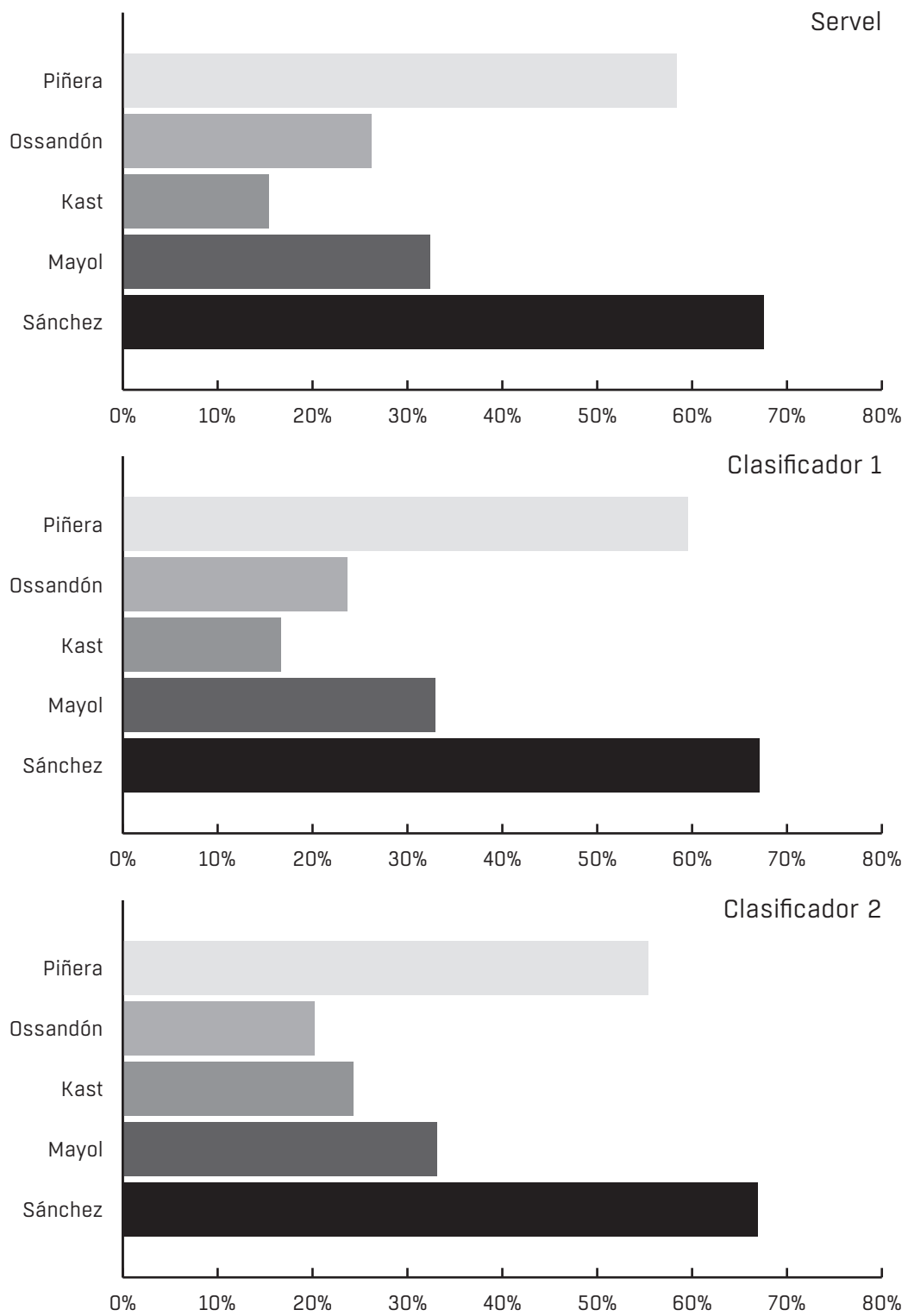


Gráfico 6. Resultados comparativos de Servel y clasificadores

Fuente: Elaboración propia.

primaria. El C2 predijo con bastante acierto y muy bajo MAE los resultados del Frente Amplio, pero no tuvo la misma capacidad con los de la derecha, lo que hace pensar en las dificultades naturales que presentan estos pronósticos cuando aumentan la cantidad de candidatos. Los indicadores también exhiben que la actividad estratégica impulsada por las candidaturas en las redes sociales pocos días u horas antes de las elecciones, distorsiona el flujo de Twitter, lo que hace aumentar el MAE de la predicción. En el siguiente gráfico se muestra comparativamente los resultados del Servicio Electoral y los nuestros.

CONCLUSIONES

Con los resultados entregados por los dos clasificadores (C1 y C2) que formaron parte de la investigación realizada, se aplica en Chile una línea de estudio que ha sido desarrollada en otras realidades electorales (Aparaschivei, 2011; Deltell, 2012; Larsson & Moe, 2011; Tumasjan et al., 2010). A nivel global, en determinados procesos políticos se han dado a conocer tendencias electorales emanadas de Twitter que fueron más certeras que las encuestas tradicionales (Deltell, Claes, & Osteso, 2013). Los MAE descritos en esta investigación evidenciaron una proyección y desarrollo de la predicción electoral a través de Twitter en Chile, que entrega espacios para su continuidad, profundización y seguimiento.

El alcance del estudio permitió destacar la importancia de integrar al análisis político predictivo una dimensión expansiva propia de la comunicación política, como es el entorno digital dialógico de las redes sociales y sus interacciones con otros medios. Además, la investigación analizó este comportamiento sobre una práctica de primarias legales que en Chile es reciente (es el segundo proceso que se celebra) y que incorporó para estas elecciones, por primera vez, el voto de los chilenos en el extranjero. Es decir, el campo de aprendizaje que se da en esta área es naciente, pero con buenos resultados.

Desde el punto de vista político y la actividad de las campañas en Twitter, también las primarias estudiadas eran inéditas, lo que conllevaba una dificultad y un desafío mayor para la predicción. Esto, debido a la incertidumbre en la participación del electorado por la ausencia en este proceso del bloque oficialista, la Nueva Mayoría. Sin embargo, la decisión de los partidos políticos de gobierno de no ingresar a las primarias no afectó la intensidad

online del proceso y tampoco los resultados de los dos clasificadores.

Si bien el Frente Amplio (327.613) obtuvo una cantidad de votos muy menor a Chile Vamos (1.418.167), esta nueva alternativa se observó con un protagonismo en redes sociales que posibilitó un rápido posicionamiento en el sistema de medios. Surge con ello la articulación de una tercera fuerza política luego de un sistema binominal que marcara la transición chilena, lo que establecerá mayor competitividad a las futuras elecciones chilenas y un desafío mayor para la predicción electoral en Twitter.

Nuestro estudio también concluye que la televisión abierta sigue siendo un medio que determina la agenda de los hechos que posteriormente son replicados con intensidad en Twitter, para luego, en una suerte de comportamiento espasmódico, bajar también intensamente. La red social se alimenta de la televisión y amplifica sus comentarios a partir de la visibilidad que obtienen los candidatos en eventos que son establecidos por el medio tradicional. Con esta dinámica de intensidad, el proceso electoral ingresa directamente en las discusiones de miles de usuarios que extienden la temporalidad de los programas de televisión, en un nuevo debate que se resignifica durante horas por Twitter y se hace más participativo. Se debe continuar estudiando si estas características evidenciadas en las primarias, posibilitaron una tasa de participación más alta que la proyectada por analistas.

Los desafíos metodológicos estarán dados también por el supuesto de representatividad que se haga respecto a los usuarios de Twitter y la exploración de otros tipos de variables de los mensajes, como las derivadas de la red de usuarios, impacto de los RT versus los mensajes originales, establecer una relación entre los mensajes positivos e intención de voto. La investigación concluye que existe una capacidad metodológica replicable a la realidad chilena, con sus particularidades electorales y el desafío para seguir mejorando la predicción en las siguientes elecciones.

DISCUSIÓN

La cantidad de mensajes, tanto tuits como retuits, fueron importantes para el liderazgo que cada candidato vencedor de su bloque obtuvo durante la campaña. Tanto Sebastián Piñera como Beatriz Sánchez mantienen una mayor producción y circulación de interacciones en Twitter, aunque Sánchez obtiene la tercera mayoría de los cinco candidatos sumando

ambos pactos (221.425 votos). Ese liderazgo en Twitter no genera necesariamente una tendencia al triunfo pensando en una presidencial abierta, ni tampoco valoraciones exclusivamente positivas como se apreció en el análisis de sentimiento.

Hipotetizamos que fue precisamente una combinación de monitoreo humano y artificial el que permitió el bajo MAE, obteniendo alentadoras respuestas de cara a las futuras elecciones. Los algoritmos fueron entrenados en el aprendizaje y en las expresiones de las interacciones de usuarios (para este caso 162.000) que, según pensamos, viven en una red de medios y comentarios que hacen de la comunicación política un espacio complejo de respuestas cognitivas, emocionales y culturales (Gerstlé, 2004). Nuestra investigación

entiende los procesos electorales como un continuo, que no puede desconectarse de la vida cotidiana, que forma parte de la expresividad simbólica de los sujetos y de las miles de conexiones que generan con sus entornos.

La capacidad de predicción en Twitter puede ser un espacio de interés para el estudio de la comunicación política, la medición de las estrategias electorales, en sus efectos y niveles de influencia, y para la calidad de la democracia. El seguimiento y la proyección de las tendencias artificiales seguirán mejorando, por lo que pueden ser oportunidades para restablecer vínculos con la ciudadanía digital. La inteligencia computacional permite un procesamiento de metadatos que plantea además desafíos que son urgentes en la discusión de la comunicación y sus disciplinas afines.

REFERENCIAS

- Aguado, J. & Navarro, H. (2013). Comunicación móvil, ecosistema digital e industrias culturales. In J. Aguado, C. Feijóo, & I. Martínez (coords.). *La comunicación móvil. Hacia un nuevo ecosistema digital* [Mobile communication: Towards a new digital ecosystem]. Barcelona: Gedisa.
- Aparaschivei, P. (2011). The Use of Media in Electoral Campaigns: Analysis on the Use of Blogs, Facebook, Twitter and YouTube in the 2009 Romanian Presidential Campaign. *Journal of Media Research*, 4(2), 39–60. Retrieved from <http://www.mrjournal.ro/docs/R2/10MR5.pdf>
- Bakliwal, A., Foster, J., van der Puil, J., O'Brien, R., Tounsi, L., & Hughes, M. (2013). Sentiment Analysis of Political Tuits: Towards an Accurate Classifier. In *Proceedings of the Workshop on Language Analysis in Social Media* (pp. 49–58). Stroudsburg, PA: Association for Computational Linguistics.
- Birmingham, A. & Smeaton, A. F. (2011). On using Twitter to monitor political sentiment and predict election results. In *Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology (SAAIP)* (pp. 2-10). Chiang Mai, Thailand: Workshop at the International Joint Conference for Natural Language Processing. Retrieved from <http://doras.dcu.ie/16670/>
- Boczkowski, P. (2016). Los diarios con Hillary, los fans con Trump. Campañas, medios y mensajes [The newspapers with Hillary, the fans with Trump. Campaigns, media and messages]. *Revista Anfibia*. Retrieved from <http://www.revistaanfibia.com/ensayo/los-diarios-hillary-los-fans-trump/>
- Campos, F. (2008). Las redes sociales trastocan los modelos de los medios de comunicación tradicionales [Social networks disrupt traditional media models]. *Revista Latina de Comunicación Social*, 11(63), 287-293. <https://doi.org/10.4185/RLCS-63-2008-767-287-293>
- Castells, M. (2009). *Comunicación y Poder* [Communication power]. Madrid: Alianza Editorial.
- Castells, M. (2015). *Redes de indignación y esperanza* [Indignation and hope networks]. Madrid: Alianza Editorial.
- Ceron, A., Curini, L., & Iacus, S. (2017). *Politics and Big Data*. London: Routledge
- Congosto M., Fernández, M., & Moro, E. (2011). Twitter y política: información, opinión y ¿predicción? [Twitter and politics: information, opinion and prediction?]. *Cuadernos de comunicación Evoca*, (4), 11-16. Retrieved from <http://neolabs.es/evoca/down/cuadernos4.pdf>

- Cotarelo, R. (Coord.) (2013). *Ciberpolítica. Las nuevas formas de acción y la comunicación política* [Cyberpolitics. New forms of action and political communication]. Valencia: Tirant Lo Blanc.
- Deltell, L., Claes, F., & Osteso, J. (2013). Predicción de tendencia política por Twitter: Elecciones Andaluzas 2012 [Political Trends Prediction on Twitter: Andalusian Election, 2012]. *Ámbitos. Revista Internacional de Comunicación*, (22), 91-100. Retrieved from <http://institucional.us.es/ambitos/?p=148>
- Fábrega, J. & Paredes, P. (2012). La política chilena en 140 caracteres [Chilean Politics in 140 Characters]. In A. Arriagada & P. Navia (Eds.), *Intermedios. Medios de comunicación y democracia en Chile* [Intermedia. Media and democracy in Chile] (pp. 199-224). Santiago: Universidad Diego Portales.
- Feijóo, C., Maghiros, I., Abadie, F., & Gómez Barroso, J.L. (2009). Exploring a heterogeneous and fragmented digital ecosystem: Mobile content. *Telematics and Informatics*, 25(3), 282-292. <https://doi.org/10.1016/j.tele.2008.11.009>
- García, F. (2015). Twitter en las campañas políticas en Latinoamérica [Twitter in Political Campaigns in Latin America]. In D. Ivoskus (Ed.), *VI Cumbre Mundial de Comunicación Política* [VI World Political Communication Summit]. Santo Domingo: Editorial Funglode.
- Gerstlé, J. (2004). *La communication politique* [Political communication]. Paris: Éditions Dalloz.
- Gutiérrez-Rubí, A. (2011). *La política vigilada. La comunicación en la era de Wikileaks* [Monitored politics. Communication in the Wikileaks era]. Barcelona: Editorial UOC.
- Guzmán, G. & Sánchez Medero, R. (2016). El ecosistema digital de la comunicación política. In R. Sánchez Medero (Ed.), *Comunicación Política. Nuevas dinámicas y ciudadanía permanente* [Political Communication. New dynamics and permanent citizenship]. Madrid: Tecnos.
- Habermas, J. (1986). *L'espace public. Archéologie de la publicité comme dimension constitutive de la société bourgeoise* [The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society]. Paris: Payot.
- Harvey, L. (2013). Communication Issues and policy implications. In F. Biocca & M. Levy (Eds.). *Communication in the age of virtual reality*. New Jersey: Lawrence Erlbaum Associates.
- Hilbert, M. (2017). The More You Know, the More You Can Grow: An Information Theoretic Approach to Growth in the Information Age. *Entropy*, 19(2), 82. <https://doi.org/10.3390/e19020082>
- Issenberg, S. (2012). *The Victory Lab*. New York: Crown Publishers.
- Jenkins, H. (2006). *Convergence culture. Where old and new media collide*. New York: University Press.
- Larsson, A. O. & Moe, H. (2011). Studying political microblogging: Twitter users in the 2010 Swedish election campaign. *New Media and Society*, 14(5), 729-747. <https://doi.org/10.1177/1461444811422894>
- Lippmann, W. (1922). *Public opinion*. New York: Harcourt, Brace and Co.
- Magnani, E. (2017). Big data y política. El poder de los algoritmos [Big data and politics. The Power of Algorithms]. *Nueva Sociedad*, (269), 45-55. Retrieved from <http://nuso.org/articulo/big-data-y-politica/>
- McCombs, M. & Shaw, D. (1972). The agenda setting function of mass media. *The Public Opinion Quarterly*, 36(2), 176-187. <https://doi.org/10.1086/267990>
- McQuail, D. (1991). *Introducción a la teoría de la comunicación de masas* [Introduction to the mass communication theory]. Barcelona: Paidós.
- Momoc, A. (2013). Social Media - PR Tools for Romanian Politicians? *Procedia-Social and Behavioral Sciences*, 81, 116-121. <https://doi.org/10.1016/j.sbspro.2013.06.398>
- Noëlle-Neumann, E. (1995). *La espiral del silencio. Opinión pública: nuestra piel social* [The spiral of silence. Public opinion: our social skin]. Barcelona: Paidós.
- Ochigame, R. & Holston, J. (2016) Filtering Dissent. Social Media and Land Struggles in Brazil. *New Left Review*, (99), 85-108. Retrieved from <https://newleftreview.org/II/99/rodrigo-ochigame-james-holston-filtering-dissent>

- Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10* (pp. 79-86). Stroudsburg, PA: Association for Computational Linguistics.
- Price, V. (1994). *La Opinión Pública* [Public opinion]. Madrid: Gedisa.
- Provost, F. & Fawcett, T. (2013). Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Big Data*, 1(1), 51-59. <https://doi.org/10.1089/big.2013.1508>
- Riquelme, F. & González-Cantergiani P. (2016). Measuring user influence on Twitter. *Information Processing & Management*, 52(5), 949–975. <https://doi.org/10.1016/j.ipm.2016.04.003>
- Rogers, R. (2004). *Information Politics on the Web*. Cambridge, MA: MIT Press.
- Russell, S. J., Norvig, P., & Davis, E. (2016). *Artificial intelligence: a modern approach*. Harlow: Prentice Hall.
- Sánchez Medero, R. (2016). Nuevos métodos para viejos propósitos: marketing político para el siglo XXI [New Methods for Old Purposes: Political Marketing for the 21st Century]. *Cuadernos Hispanoamericanos*, (792), 25-41. Retrieved from https://issuu.com/publicacionesaacid/docs/web_cha_792_junio_2016
- Schoen, H., Gayo-Avello, D., Takis Metaxas, P., Mustafaraj, E., Strohmaier, M., & Gloor, P. (2013). The power of prediction with social media. *Internet Research*, 23(5), 528-543. <https://doi.org/10.1108/IntR-06-2013-0115>
- Scolari, C. (2012). Media ecology: exploring the metaphor to expand the theory. *Communication Theory*, 22(2), 204-225. <https://doi.org/10.1111/j.1468-2885.2012.01404.x>
- Scolari, C. (2013). *Narrativas transmedia. Cuando todos los medios cuentan* [Transmedia Narrative: when all media count]. Barcelona: Deusto.
- Scolari, C. (2015). *Ecología de los medios. Entorno, evoluciones e interpretaciones* [Media ecology: Environment, evolutions and interpretations]. Barcelona: Gedisa.
- Sebastiani, F. (2002). Machine Learning in Automated Text Categorization. *ACM Computing Surveys (CSUR)*, 34(1), 1–47. <https://doi.org/10.1145/505282.505283>
- Taylor, S. & Lethman, B. (2016). Forecasting at scale. *The American Statistician*, (just-accepted). <http://doi.org/10.1080/00031305.2017.1380080>
- Thompson, J. (1998). *Los media y la modernidad. Una teoría de los medios de comunicación* [The Media and Modernity: A Social Theory of the Media]. Barcelona: Paidós Comunicación.
- Tumasjan, A., Sprenger, T., Sandner, P., & Welpe, I. (2010) Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. *Icwsn*, 10(1), 178-185. Retrieved from <http://www.aai.org/ocs/index.php/ICWSM/ICWSM10/paper/view/1441>.
- Van Dijck, J. (2016). *La cultura de la conectividad: una historia crítica de las redes sociales* [The Culture of Connectivity: A Critical History of Social Media]. Buenos Aires: Siglo Veintiuno Editores.
- Varol, O., Ferrara, E., Davis, C., Menczer, F., & Flammini, A. (2017). Online Human-Bot Interactions: Detection, Estimation, and Characterization. *arXiv preprint arXiv:1703.03107*. Retrieved from <https://arxiv.org/pdf/1703.03107.pdf?ref=il>
- Wescott, N. (2008). Digital Diplomacy: The Impact of the Internet on International Relations. *Oxford Internet Institute Research Paper Series*, (16), 1-20. <https://doi.org/10.2139/ssrn.1326476>
- Willmott, C. J. & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30(1), 79–82. <https://doi.org/10.3354/cr030079>
- Zhang, Y., Jin, R., & Zhou, Z.-H. (2010). Understanding bag-of-words model: a statistical framework. *International Journal of Machine Learning and Cybernetics*, 1(1–4), 43–52. <https://doi.org/10.1007/s13042-010-0001-0>

SOBRE LOS AUTORES

Pedro Santander, licenciado en Comunicación y doctor en Lingüística. Profesor titular de la Escuela de Periodismo de la Pontificia Universidad Católica de Valparaíso, investigador en el área de análisis de discurso y análisis de medios.

Claudio Elórtogui, doctor en Periodismo y Ciencias de la Comunicación de la Universidad Autónoma de Barcelona. Profesor de la Escuela de Periodismo de la Pontificia Universidad Católica de Valparaíso, investigador en el área de comunicación política.

Cristian González, doctor en Lingüística de la Pontificia Universidad Católica de Valparaíso, Chile y doctor en Ciencias del Lenguaje en la Universidad Paris 13, Francia. Actualmente es profesor adjunto del Instituto de Literatura y Ciencias del Lenguaje de la Pontificia Universidad Católica de Valparaíso, donde realiza docencia en programas de pregrado y posgrado. Investiga en el ámbito de la lingüística del discurso, desde la perspectiva semiolingüística, centrado en el discurso periodístico, político y académico.

Héctor Allende, doctor en Ingeniería Informática de la Universidad Técnica Federico Santa María. Es profesor e investigador de la Escuela de Ingeniería Informática de la Pontificia Universidad Católica de Valparaíso en el área de aprendizaje automático y reconocimiento de patrones.

Wenceslao Palma, doctor en Informática de la Universidad de Nantes. Es profesor e investigador de la Escuela de Ingeniería Informática de la Pontificia Universidad Católica de Valparaíso en el área de *Big Data*.