CHILECLIPPING.COM
INFORMACION DIARIA

Informe "Big Data": Correlaciones vs. evidencia

Domingo, 19 de Enero de 2020 - Id nota:901151

Medio : El Mercurio
Sección : Economía y Negocios
Valor publicitario estimado : $4192240.-
Página : B6
Tamaño : 29 x 16

Ver en formato web

Informe "Big Data": Correlaciones vs. evidencia

ANÁLISIS
Eduardo Engel



EL INFORME "BIG DATA" SIGUE EN LA AGENDA NOTICIOSA. ESTA SEMANA INTERFERENCIA REVELÓ QUE EL LUNES 2 DE DICIEMBRE DICHO INFORME FUE ANALIZADO EN UNA REUNIÓN QUE TUVO LUGAR EN DEPENDENCIAS DE LA AGENCIA NACIONAL DE INTELIGENCIA (ANI) Y DONDE PARTICIPARON EL GERENTE LEGAL DE QUIÑENCO, RODRIGO HINZPETER; UN EXPERTO DE LA EMPRESA QUE PREPARÓ EL INFORME (ALTO ANALYTICS) Y REPRESENTANTES DEL MINISTERIO DEL INTERIOR, DE LAS FUERZAS ARMADAS Y DE LA ANI.

Quiñenco y el gobierno confirmaron que la primera contactó al gobierno con Alto Analytics y no se refirieron a la reunión en que se habría analizado el estudio. El gobierno también insistió en que no encargó el estudio y que este no fue financiado con fondos públicos. Cabe recordar que la opinión pública supo de la existencia del informe el 19 de diciembre, cuando el ministro del Interior, Gonzalo Blumel, se reunió con el fiscal nacional, Jorge Abbott, para entregárselo. Esto en medio de una polémica entre el Ejecutivo y el Ministerio Público acerca de si existía información de inteligencia sobre la participación de grupos organizados en los ataques al Metro de octubre.

Más allá de varias preguntas que siguen abiertas, el informe "Big Data" permite ilustrar el potencial y las limitaciones que tienen las nuevas metodologías para analizar grandes cantidades de datos, conocidas por diversos anglicismos, entre ellos "big data", "data analytics", "machine learning" y "data science". Con tal objeto, es conveniente partir por resumir lo que sabemos del contenido del estudio.

El informe no es público, lo que se conoce de su contenido proviene de un reportaje de La Tercera del 21 de diciembre. Según este medio, el estudio clasifica los autores de comentarios en redes sobre el estallido social en cinco grupos. Para ello analiza 60 millones de comentarios de casi 5 millones de autores únicos, entre el 18 de octubre y el 21 de noviembre. Un primer grupo incluye canales de televisión rusos y venezolanos y figuras peronistas que critican el modelo económico chileno por generar desigualdad. Un segundo grupo también consiste de cuentas internacionales que, a diferencia del primer grupo, no apoya las manifestaciones. Este grupo incluye al analista argentino Jorge Lanata y se constata que es menos numeroso que el primero. Un tercer grupo, el más grande de todos, lo constituyen jóvenes que antes del 18 de octubre registraban poco interés en temas políticos y que son aficionados al K-pop, un género musical que se originó en Corea. Este grupo se enfoca en las violaciones de derechos humanos y trata con tono irónico afirmaciones según las cuales Nicolás Maduro esté tras el estallido social. Un cuarto grupo, más politizado que el tercero, menciona la desigualdad como la causa de las movilizaciones y tiene a los seleccionados nacionales Claudio Bravo y Gary Medel, junto a líderes del Partido Comunista y del Frente Amplio, entre sus miembros más importantes. Por último, está un grupo de fuentes nacionales que defiende al gobierno y donde destacan las cuentas del Ejército, Carabineros y Teresa Marinovic. El informe afirma que el impacto de este último grupo es mucho menor que aquel de quienes critican al gobierno.

¿Para qué sirve la clasificación anterior? Podría usarse para predecir cómo piensa una persona determinada sobre la crisis social. Los seguidores de Karol Cariola tienen percepciones muy distintas que los seguidores de Teresa Marinovic, por ejemplo. Y los fanáticos del K-pop son muy críticos de las violaciones de derechos humanos. La mayoría de estas correlaciones son bastante obvias y las que no lo son tampoco parecen particularmente de relevancia.

El problema central del estudio es que la clasificación no sirve para responder las preguntas relevantes, la más importante de éstas, siendo: ¿quiénes son los responsables de la destrucción de las estaciones de metro el 18 de octubre? Que un medio chavista o ruso comente de manera recurrente y obsesiva eventos en Chile no significa que tengan responsabilidad alguna en ellos. Y que líderes del Frente Amplio y el Partido Comunista expresen su indignación por abusos de derechos humanos tampoco.

En general, las aplicaciones más exitosas de "big data" usan grandes cantidades de datos para hacer predicciones en los ámbitos más diversos. Por ejemplo, evalúan el riesgo de incumplimiento de un deudor a partir del análisis de millones de préstamos. El tema clave es que una buena regla predictiva o de clasificación sirve poco y nada para entender las causas profundas de un fenómeno. Por ejemplo, existe un sinnúmero de maneras de clasificar los más de cinco millones de personas que comentaron el estallido social en las redes en grupos relativamente homogéneos. Salvo que encontremos comentarios de personas organizándose para delinquir, estas clasificaciones no sirven de nada para responder las preguntas de interés y menos para establecer una supuesta injerencia extranjera.

"En general, las aplicaciones más exitosas de "big data" usan grandes cantidades de datos para hacer predicciones en los ámbitos más diversos (...) El tema clave es que una buena regla predictiva o de clasificación sirve poco y nada para entender las causas profundas de un fenómeno (...) Salvo que encontremos comentarios de personas organizándose para delinquir, estas clasificaciones no sirven de nada para responder las preguntas de interés y menos para establecer una supuesta injerencia extranjera".