

Nuevas estrategias para la construcción de datos del conflicto Mapuche-Estado chileno

Por Alejandro Corvalán

Académico del Departamento de Ingeniería Industrial (DII) e investigador del Instituto Milenio para la Investigación de Imperfecciones de Mercado y Políticas Públicas (MIPP).

Los datos son la base de cualquier investigación empírica. La recolección y generación de datos en violencia política, desde guerras civiles hasta conflicto social, ha progresado fuertemente en las últimas décadas. En este contexto, una serie de investigadores hemos trabajado durante varios años en la construcción de MACEDA (Mapuche-Chilean State Conflict Events Database), la primera base armonizada sobre episodios recientes de conflicto entre el pueblo Mapuche y el Estado chileno. La base se enmarca en el Proyecto de Datos Mapuche (MDP, por su sigla en inglés), el cual es financiado por el Instituto Milenio de Imperfecciones de Mercado y Políticas Públicas (MIPP) y auspiciado por las Universidades de Chile y de Talca.

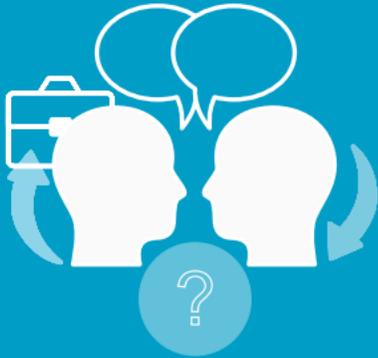
MACEDA es la primera base de datos sobre un conflicto indígena de autodeterminación en América y fue aceptada para su publicación en 2021 en el *Journal of Peace Research*, la principal revista sobre el tema. La base incluye más de 2.600 eventos de conflicto, codificados de manera manual a partir de medios de prensa, con información en más de 40 dimensiones como tipo de eventos, actores involucrados, nivel de violencia y costos económicos del evento. La versión original de MACEDA, que identifica los eventos a nivel comunal para el



período 1990 a 2016, se encuentra disponible como fuente abierta para su uso público. Su versión actual, aún no liberada, incluye los eventos hasta el año 2020 y la georreferenciación de cada uno de ellos.

La construcción de bases de conflicto como MACEDA presenta significativas sinergias con el estudio aplicado de datos y el uso de inteligencia artificial. Dos líneas de colaboración que se podrían explorar son las que siguen:

Desde hace casi una década se ensaya el uso de *machine learning* para la construcción de



información en eventos de conflicto. De hecho, en la actualidad existen múltiples algoritmos de búsqueda para la generación de datos en violencia social a partir de medios escritos y de prensa. No obstante, la literatura sugiere que la codificación inicial de los eventos debe hacerse de forma manual, dada la novedad que involucran las distintas experiencias investigadas. En el caso de MACEDA, por ejemplo, uno de los principales tipos de eventos son las tomas de tierra, que constituyen eventos muy específicos a conflictos de autodeterminación indígena. Al respecto, MACEDA presenta la ventaja que ya ha sido íntegramente clasificada de manera manual, de modo que están los insumos necesarios para avanzar hacia una búsqueda automatizada basada en el reconocimiento de patrones de texto.

Por otra parte, la creciente popularidad de los sitios de microblogs como Twitter ha dado un nuevo ímpetu a la construcción de bases de datos y análisis de eventos de violencia social. La literatura en esta materia está en sus pri-



meros pasos, dado que los mensajes cortos están dotados de poca estructura y no tienen patrones gramaticales que permitan asociarlos trivialmente a situaciones de conflicto. Sin embargo, las experiencias recientes son prometedoras. Para el caso mapuche, ya se ha usado Twitter para estudiar la reacción de las redes sociales frente a los casos de violencia; técnica conocida como *Sentiment Analysis*. La implementación de estrategias más complejas para la identificación de esos mismos hechos de violencia es aún una agenda abierta.

Hoy tenemos condiciones idóneas en nuestro país y nuestra facultad para levantar nuevas líneas de trabajo transdisciplinar en las interfaces de las matemáticas, la inteligencia artificial y la ciencia de datos, las que potenciarán tanto la investigación científica que desarrollamos en estas áreas como el impacto que tendrán en el desarrollo del país.