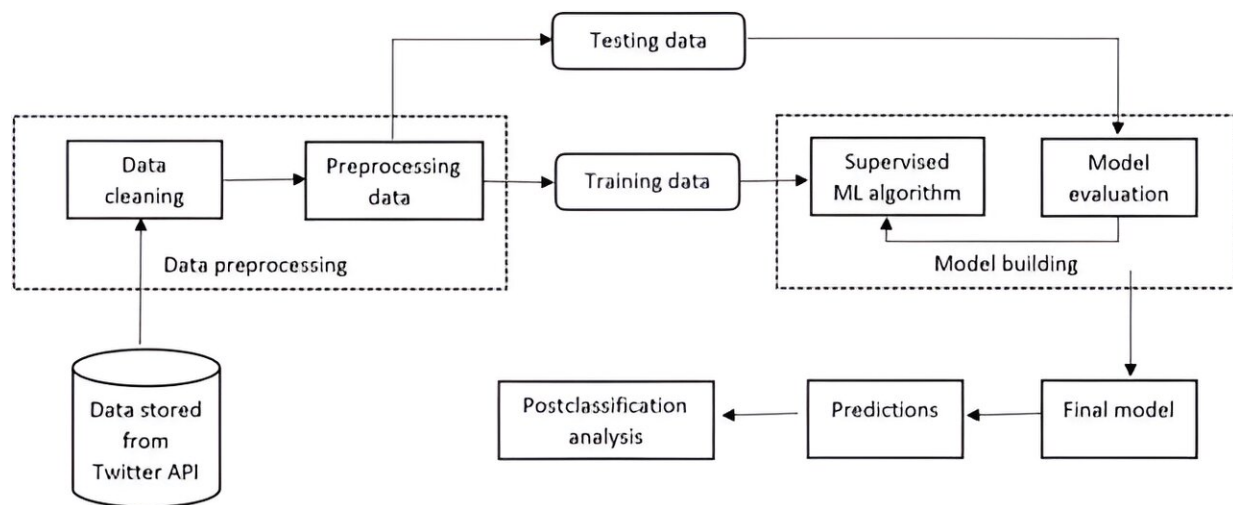# Automated chronic pain cohort mined from Twitter using machine learning

October 10 2023

End-to-end data processing framework involving data collection/storage via the Twitter API, data preparation and preprocessing, training, evaluation, and classification using the final model and postclassification analyses. Credit: *Health Data Science* (2023). DOI: 10.34133/hds.0078

Social media platforms like Twitter (now X) offer valuable insights into self-reported chronic pain, according to a recent study by multiple institutions. The researchers have automated the process of establishing a chronic pain cohort, setting the stage for future data mining and causal association studies.

The study was published in *Health Data Science*.

"We aimed to detect large-scale discussions related to chronic pain on Twitter, develop methods for automatic detection of self-disclosures, and collect and analyze longitudinal data," said Abeed Sarker, associate professor at Emory University.

Chronic pain and subsequent opioid use pose significant health and economic challenges. Gathering public knowledge and experiences could guide alternative therapies for specific pain types.

The team collected data via Twitter's academic API, followed by manual annotation and classification.

For instance, posts mentioned therapies like meditation and chiropractic, and sentiment analysis revealed varying public opinions on these therapies.

"Social media is a rich source for chronic pain-related information. Methods like NLP and machine learning can extract key insights from this large dataset," explained Prof. Sarker.

The study's automation features include detecting relevant posts and identifying self-reported cases, enabling the inclusion of individuals usually not reachable through traditional research settings.

Social media-based cohorts may also lend a unique perspective to chronic pain. For example, subscribers' social support (e.g., numbers of followers and interactions with a given post), its influence on their quality of life, and the connection between alternative therapies and social aspects of living.

This is the first publication that builds an elaborate NLP and machine learning pipeline to curate knowledge about chronic pain from patient-generated social media data. The automated collection of cohort

members via classification and their posts over time will yield various hypotheses about chronic pain control and an unprecedented resource for long-term studies on the topic.

"The next step will be to do a more comprehensive analysis of the data posted by our chronic pain cohort, particularly to generate hypotheses about effective opioid alternatives for specific chronic pain types," said Prof. Sarker.

Provided by Health Data Science