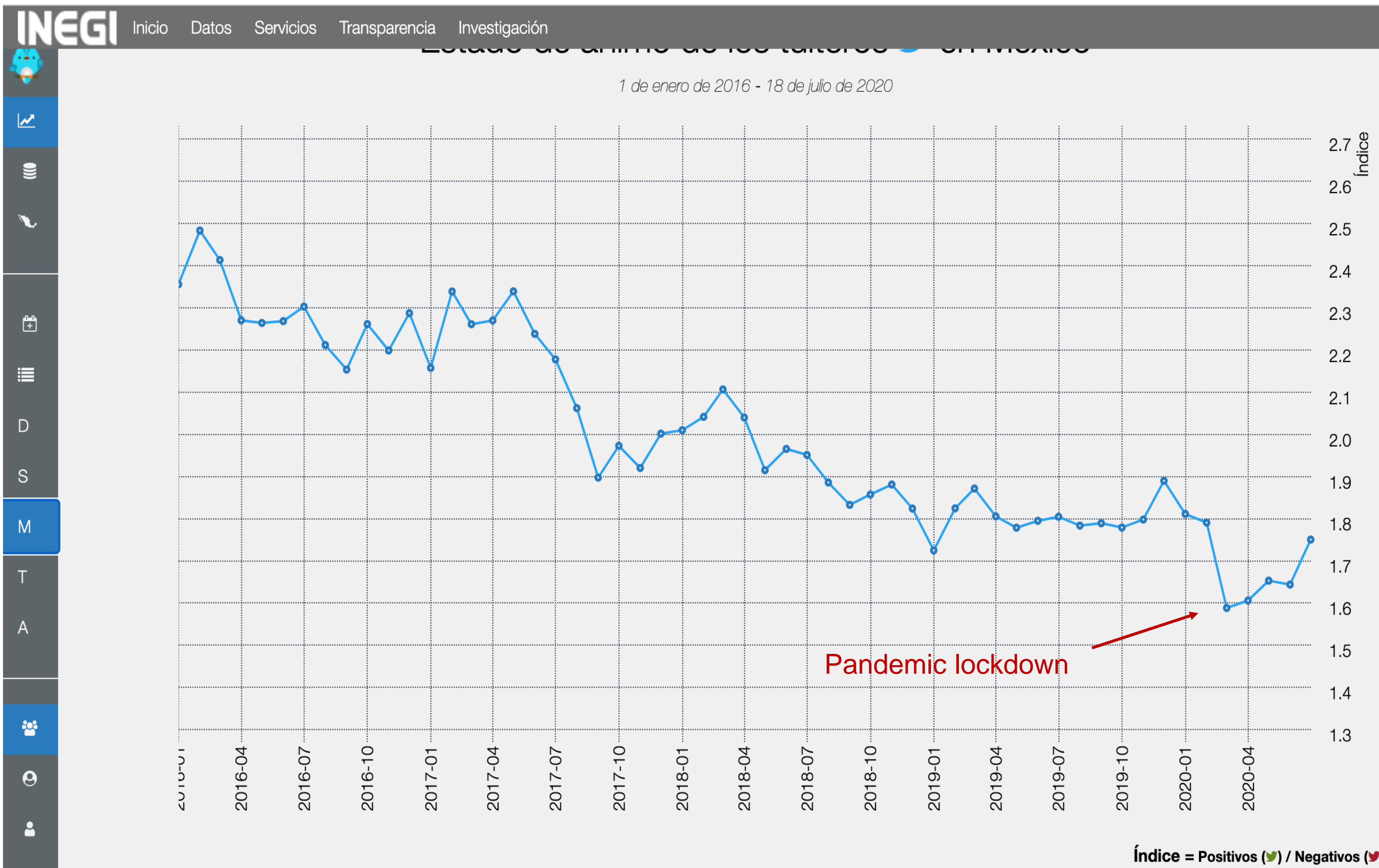




ML at INEGI

Alejandro Ruiz
Researcher, Social Scientist

Sentiment Analysis from twitter

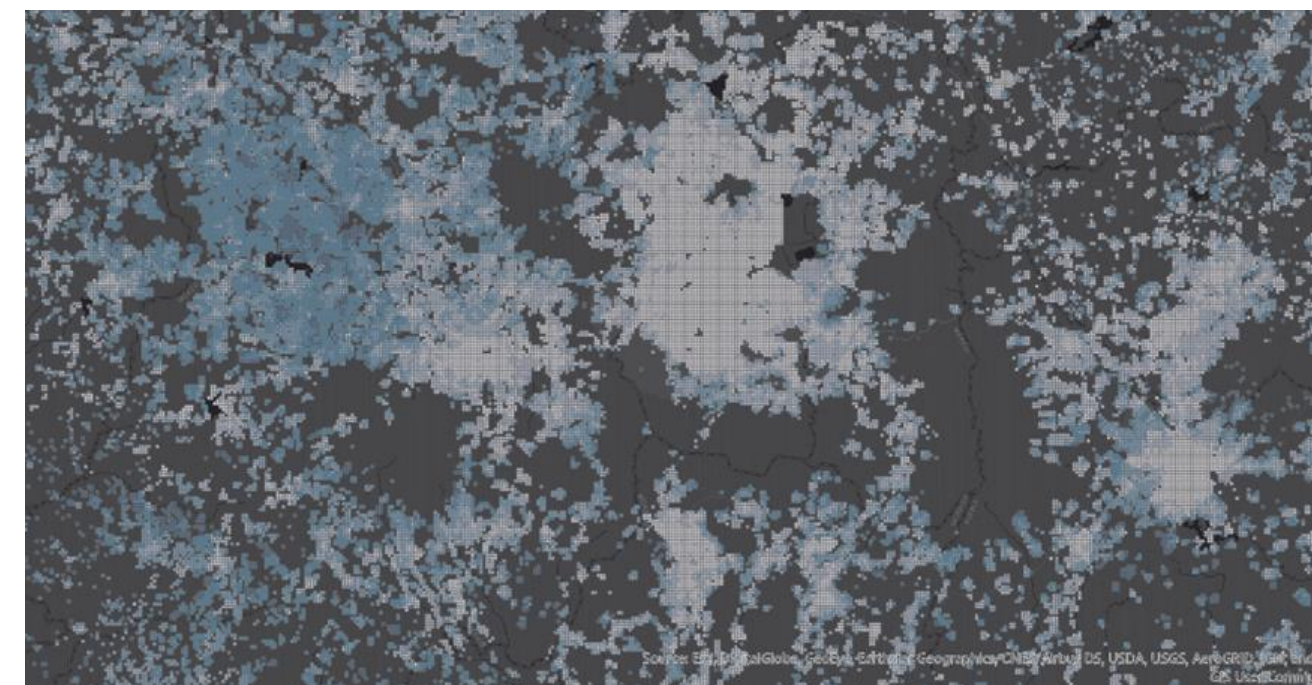
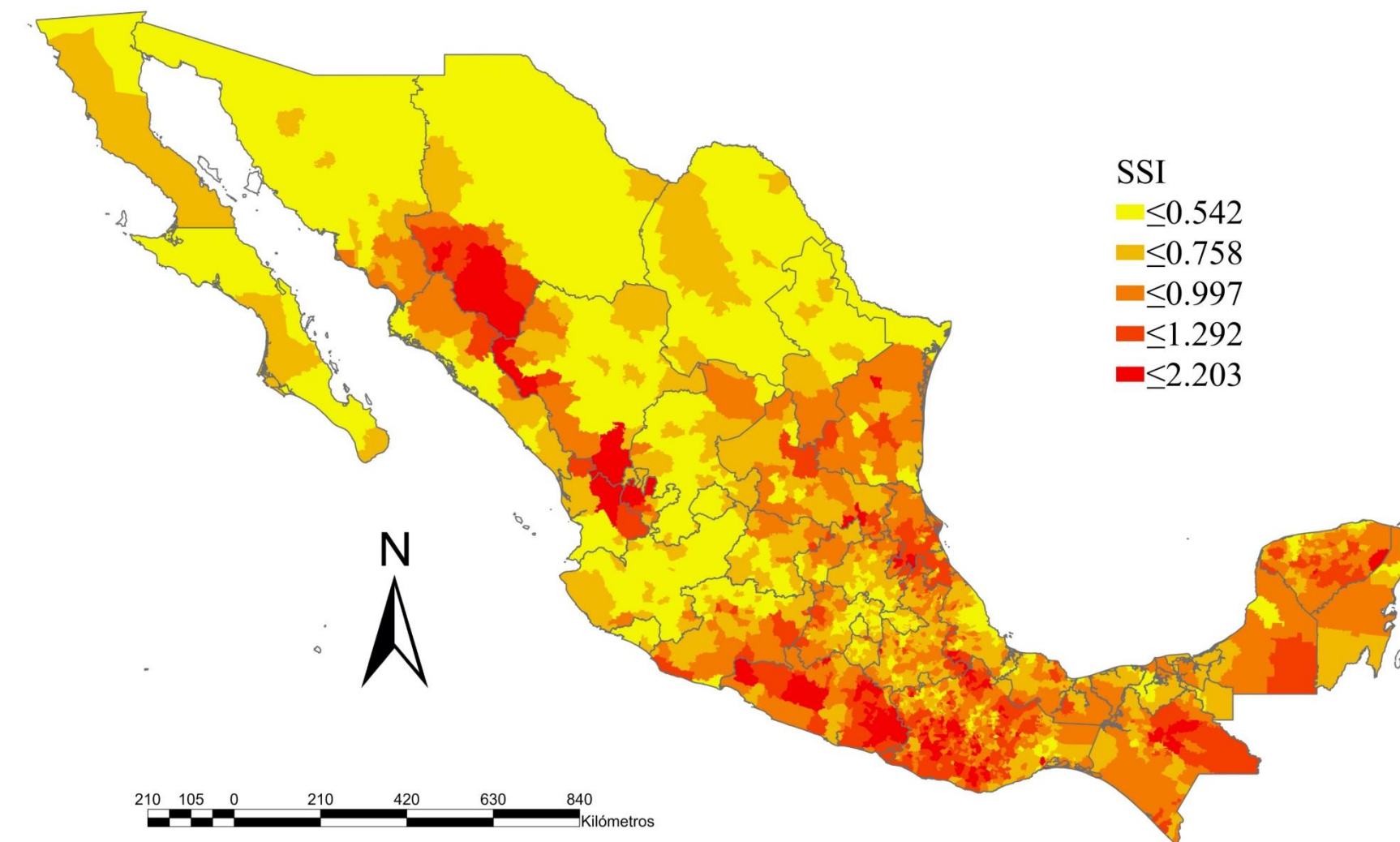


- » 10,000 people labeled over 54,000 tweets based on the emotion transmitted
 - › Positive
 - › Negative
- » 93,000 georeferentiated tweets on a daily basis
- » + 300 millions tweets
- » TFIDF + SMV
- » positivity quotient = $\frac{\#positive}{\#negative}$
- » It only reflects the emotion of people actually tweeting
 - › Restrictions from the Public API

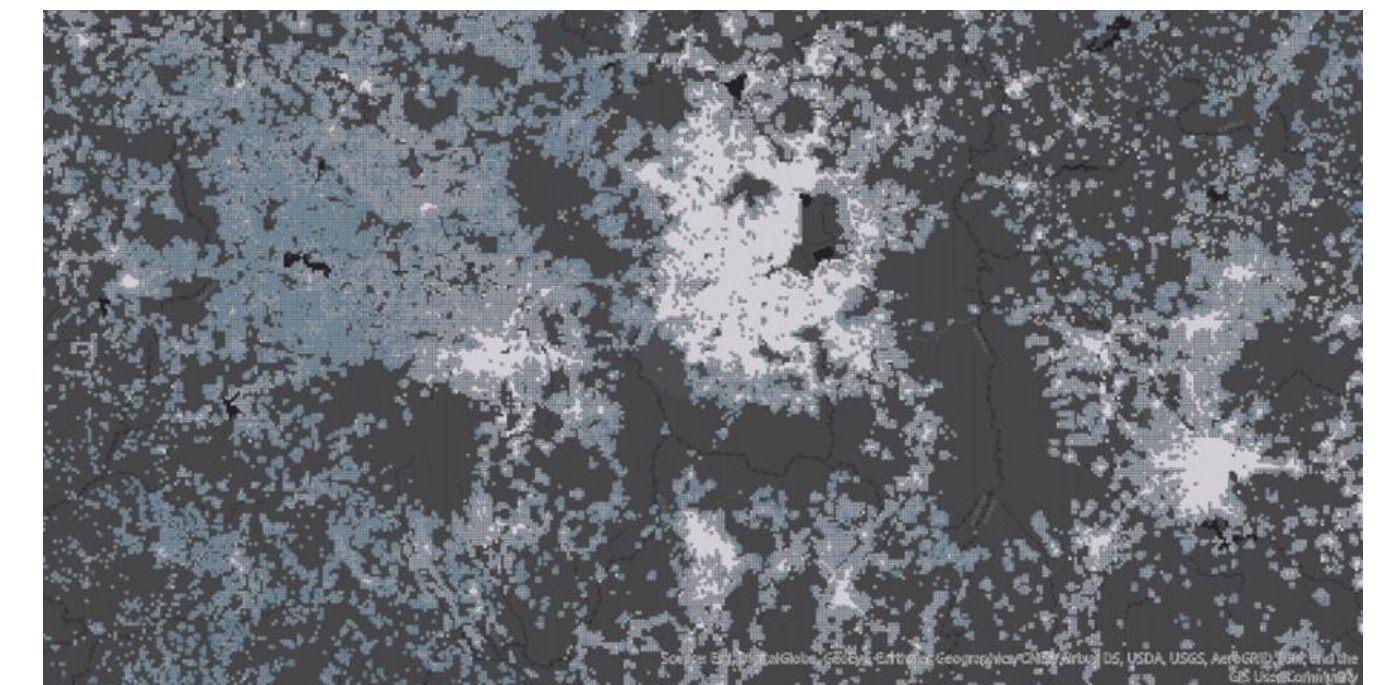
<https://www.inegi.org.mx/app/animotuitero/#/app/multiline>

ML + Satellite imagery to identify vulnerable settlements

- » Vulnerable settlements to be identified at the microscale and in a more timely manner
- » 600x600 meters grid, matching a vulnerability index to an image
 - › DL algorithm to train a model
- » Multidisciplinary team
- » Challenge:
 - › to improve accuracy metrics

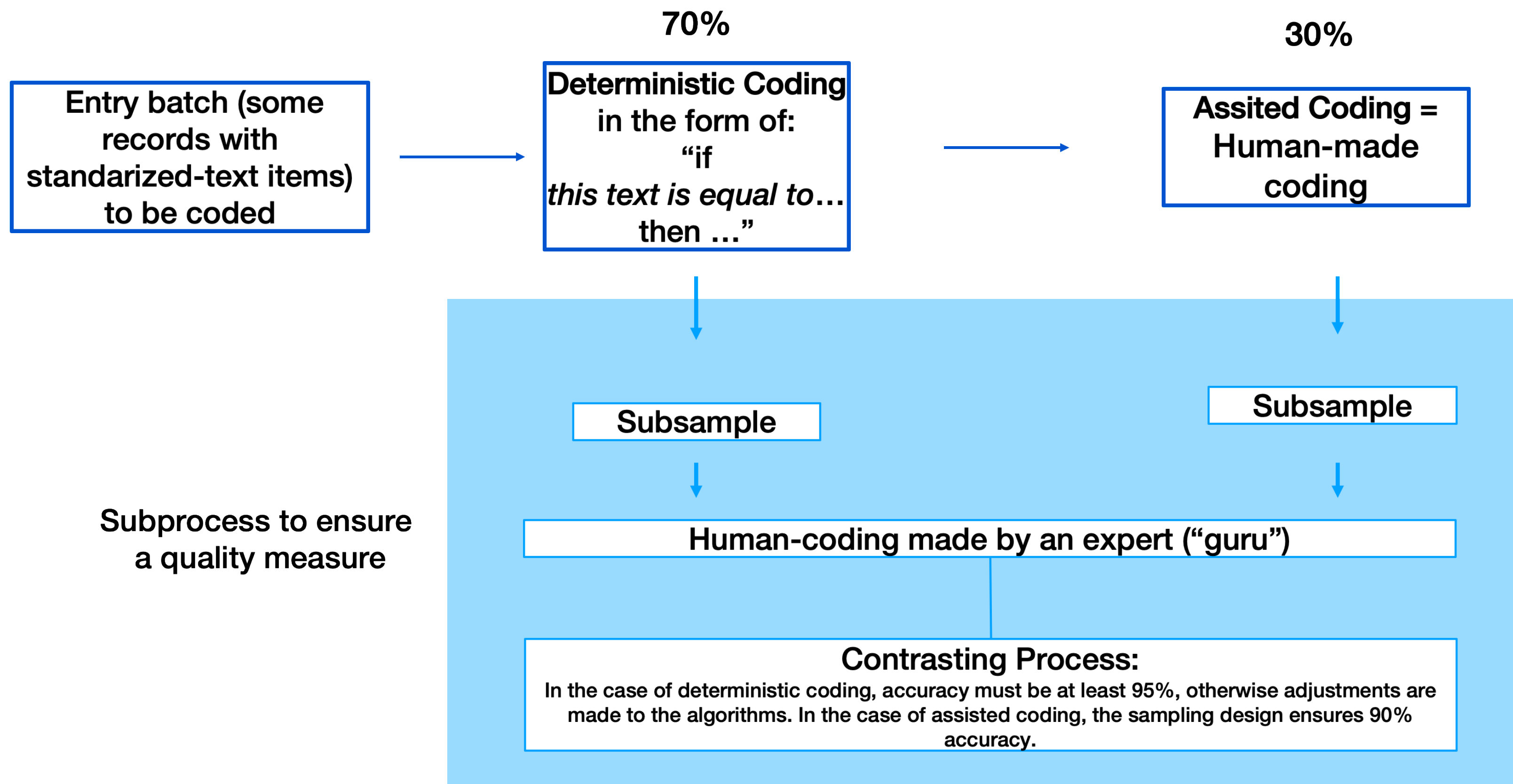


Real

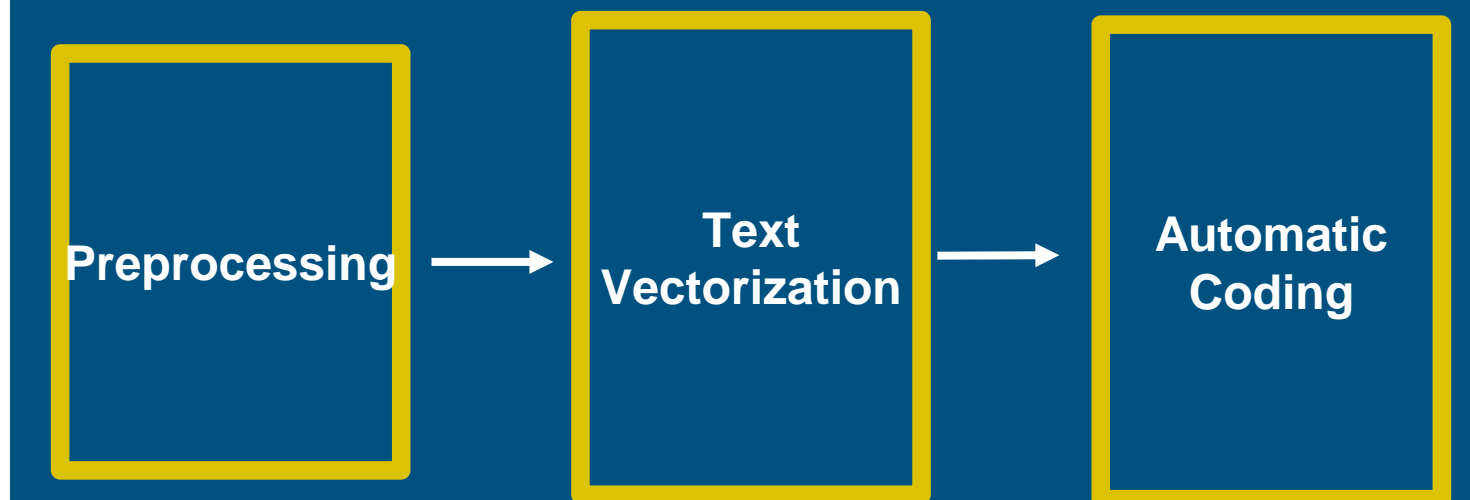


Predicted

Coding Occupation and Economic activity variables



- » Our goals:
 - › Reduce the percentage human-made coding
 - › Verify the quality of deterministic coding



- » We can reduce the human workload up to 40%
- » Challenges:
 - › To adjust the current process to add this new block based on NLP & ML



ML at INEGI

jose.ruizs@inegi.org.mx