**Article Selection Process**

Unclassified articles from initial search

Classified articles by coders from initial search

**Evaluation:**

Capture-Recapture model on adding search terms

**First Stage: Keyword Selection**

The aim of this stage is to identify highly predictive keywords of election violence articles based on changes in predictability of keywords from the classified to unclassified articles. This allows us to account for changes in the language and reporting.

Human selection of additional keywords

Feed keywords to crawler (fast mode)

**Evaluation:**

Random sample of non-selected articles for human evaluation

**Second Stage: Select Downloadable Set of Articles**

The aim of this stage is to identify a downloadable set of articles to further analyse. We use two classifiers trained on classified articles by coders to do so: (1) a classifier on the search terms and (2) a classifier on the descriptions. The downloadable set is the set of articles either classifiers considers relevant.

Feed document ids to crawler (full mode)

**Evaluation:**

Random sample of both non-selected and selected articles for human evaluation

**Third Stage: Identify Codable Set of Articles**

The aim of this stage is to identify the set of articles to be coded for further analysis. We use a classifier trained on classified articles by coders to identify relevant election violence articles based on the full OCR.

Download PDF image, upload to platform, and assign to coders