

codebook

Christy Yu

2023-06-14

Codebook 1: Dataset for Logistic Regression, KNN, and Random Forest

author: String of author who wrote the article

published: factor that represents the month article was published in (10 means october, 11 means november)

site_url: string that represents the url domain where the article is from

main_img_url: string that represents the url of the image

type: string that represents smaller label type of article (bias/conspiracy/fake/bs/satire/hate/junksci/state)

label: outcome variable string of if article is real or fake

has_image: whether the article contains an image or not (0 is no, 1 is yes)

Codebook 2: Dataset for Neural Network

text_without_stopwords: string of the article that does not contain stopwords – used as the NLP part of the neural net model.