Model Info Sheet

**Section 1: Information about paper or report**

1) Author(s): Names of the authors of the paper or report

*Hannah Rashkin et al.*

2) Title of the paper or report which introduces the model

*Truth of Varying Shades: Analyzing Language in Fake News and Political Fact-Checking*

3) DOI or permanent link to the paper or report (for example, link to arxiv.org webpage)

https://drive.google.com/drive/u/2/folders/1\_vpD0tXgCIxTMU8npSFgj2KxCfHXKAAB

4) License: Under which license(s) are the data and/or model shared?

5) Email address of the corresponding author

*hrashkin@cs.washington.edu*

**Section 2: Scientific claim(s) of interest**

6) Does your paper make a generalizable claim based on the ML model? If yes, what is the scientific claim? For example, “Our ML model can be used to diagnose Covid-19 using chest radiographs of adult patients”.

*The authors make the hedgiest and hand-waviest of general claims. To paraphrase: linguistic analysis of text features can help determine the truthfulness of a text.*

7) Is the scientific claim made about a distribution or population from which you can sample? If yes: (a) what is the population or distribution about which the scientific claim is being made? (b) What is the sample used for the study? For example, “(a) Population: adult patients with symptoms of Covid-19. (b) Sample: We use a random sample of adult patients who present at a U.S. based hospital between April 2020 and June 2020”.

*This is a bit murky – the sample set is a corpus of articles pulled from 8 news and satire sites. The articles are not identified in-text or as a part of the supplement. We could sample from the universe of all news articles on these sites published within the timeframe of the study, but wouldn’t be able to guarantee any consistency in terms of random selection procedure*

8) Does the scientific claim only apply to certain subsets of the distribution mentioned in Q6? For example, “Our model works on chest radiographs of U.S.-based adult patients and might not generalize to radiographs taken in other places or using different machines.”

*Yes, though not explicitly – the authors train their LSTM model on news articles from a subset of news sites, then test the trained model on news articles from the remaining news sites. Each of these news sites has a notably unique voice and style of writing, however: it seems likely that the training set would be markedly different from the test set.*

**Section 3: Train-test split is maintained across all steps in creating the model**

9) Train-test split type: How was the dataset split into train and test sets? (For example, cross-validation; separate train and test sets).

*Per the paper: there were ‘20k total articles from the Onion, American News, The Activist, and the Gigaword news excluding ‘APW’, ‘WPB’ sources’ for the train set, and the test set consisted of ‘3k articles from the remaining sources.’*

10) Are there duplicates in the dataset? If yes, explain how duplicates are handled to ensure the train-test split.

*This is unclear – the researchers seem to assume that different sources have unique authors, but this might not be the case, esp. considering that the Onion and Clickhole were affiliate sites in 2017.*

11) In case the dataset has dependencies (e.g., multiple rows of data from the same patient), describe how the dependencies were addressed (for example, using block-cross validation).

*Again, see 10): unclear if cross-validation for bylines was performed across sources.*

*Authorship is the main dependency here. The researchers specify that their analysis is blind to authorship, but neglect to mention that Clickhole and the Onion were affiliate sites in 2017, and that there were very possibly writers who contributed articles to both sites.*

12) List all the pre-processing steps used in creating your model. For example, imputing missing data, normalizing feature values, selecting a subset of rows from the dataset for building the model.

*Paper is pretty sparse on these details – pre-processing to create id-idf vectors, no imputation of missing data, no normalization across article length or style, no subsetting aside from selection of test-train. Further, no detail on what hyperparameter tuning consisted of, and what hyperparameters the authors eventually used …*

13) How was the train-test split observed during each pre-processing step? If applicable, use a separate line for each step mentioned in Q12.

*Unclear (see 12 above).*

14) List all the modeling steps used in creating your model. For example, feature selection, parameter tuning, model selection.

*Unclear! Precious little detail included about tuning steps; the features appear to be those turned up by an LICW analysis on corpora.*

15) How was the train-test split observed during each modeling step? If applicable, use a separate line for each step mentioned in Q14.

16) List all the evaluation steps used in evaluating model performance. For example, cross-validation, out-of-sample testing.

*No out-of-sample testing conducted. F1 scores were calculated using ground-truth dataset labels.*

17) How was the train-test split observed during each evaluation step? If applicable, use a separate line for each step mentioned in Q16.

**Section 4: Test set is drawn from the distribution of scientific interest.**

18) Why is your test set representative of the population or distribution about which you are making your scientific claims?

*Probably not, for the simple reason that articles posted on each of these sites have very distinct styles and areas of coverage.*

19) Explain the process for selecting the test set and why this does not introduce selection bias in the learning process.

*The test set absolutely* did *introduce selection bias, for the reason outlined above.*

20) In case your model is used to predict a future outcome of interest using past data, detail how data in the training set is always from a date earlier than the data in the test set.

**Section 5:** **Each feature used in the model is legitimate for the task**

21) List the features used in the model, alongside an argument for their legitimacy. A legitimate feature is one that would be available when the model is used in the real world and is not a proxy of the outcome being predicted. You can also include this list in an appendix and reference the relevant section of your Appendix here.

An example of a feature that should not be included (for illustration only; you do not need to include these in your model info sheet): “Anti-hypertensive drugs: We do not include the use of anti-hypertensive drugs as a feature in our ML model for hypertension diagnosis since that information is only available after diagnosis and would not be available when a new patient presents with symptoms of hypertension.”

***Note:*** *You do not need to list each feature used in your model here. However, you must provide an argument for the legitimacy of each feature included in your model to ensure that your model does not suffer from leakage due to illegitimate features. For example, “our model only uses data from the previous year as features. For instance, to predict civil war in 2017, we only use lagged features from the year 2016. Since these features are always available in advance of when we want to make predictions using our model, none of these features can lead to leakage.”*