Model Info Sheet

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

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

*Verónica Pérez-Rosas, Bennett Kleinberg, Alexandra Lefevre, Rada Mihalcea*

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

*Automatic detection of fake news*

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

https://arxiv.org/pdf/1708.07104.pdf

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

5) Email address of the corresponding author

*vrncapr@umich.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”.

*Not so much a generalizable claim based on an ML model – moreso a general statement about*

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”.

*We might be hard-pressed to precisely sample from FakeNewsAMT – the authors generated a `fake news’ corpus written by mechanical turkers (i.e., actually presented them with real news articles, then had them write their `fake’ versions).*

*Might be easier to sample from or build a corpus similar to Celebrity.*

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.”

*Per the authors, ‘different linguistic properties underlying different kinds of deception are more likely to explain cross-domain performance decreases than poorly performing feature sets.’*

**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).

*Five-fold cross validation for the model design portion of the study.*

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

*Seems possible that there were certain articles within the fake news corpus that were written by the same author, and that bore similar stylistic signatures (for both the AMT articles and those pulled from fake news sites).*

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).

*Unclear if certain authors contributed multiple articles, or if bias toward certain domain-specific features was related to extremely platform-specific writing styles (e.g., each celebrity gossip site possibly has its own style).*

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.

*During the development of the fake news corpus, authors conducted some fact-checking of claims found on celebrity news sites. Authors also selected celebrity news to reflect a diversity of topics. Unclear what maneuvers authors might have made within each text itself (tokenizing / lemmatizing text; removing stop words, etc).*

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.

*Five-fold x validation.*

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

*Feature selection; model selection; testing on expanded data set;*

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

*Five-fold x validation.*

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

*Five-fold x validation; testing on a larger dataset to gauge learning curve / any improvements to be observed on larger set;*

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

*Five-fold x validation.*

**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?

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

*The handwritten fake news set might introduce some biases as a result of 1) idiosyncrasies associated with individual writers, and 2) the scope of articles in question (e.g., possibly ‘negating’ true news articles does not necessarily yield convincing fake news articles).*

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.

*Unclear if this is something that you might have to worry about for the AMT dataset. Authors also don’t specify timeframe of their data collection, only raw counts of articles.*

**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.

*All features selected / developed using established methods. Listed here:*

*ngrams - from bag of words representation of each article*

*punctuation - constructed from linguistic inquiry and word count (LIWC) software*

*psycholinguistic features - constructed from LIWC lexicon*

*readability - number of syllables, complexity of words, number of paragraphs …*

*syntax - features-derived production rules using Stanford Parser.*