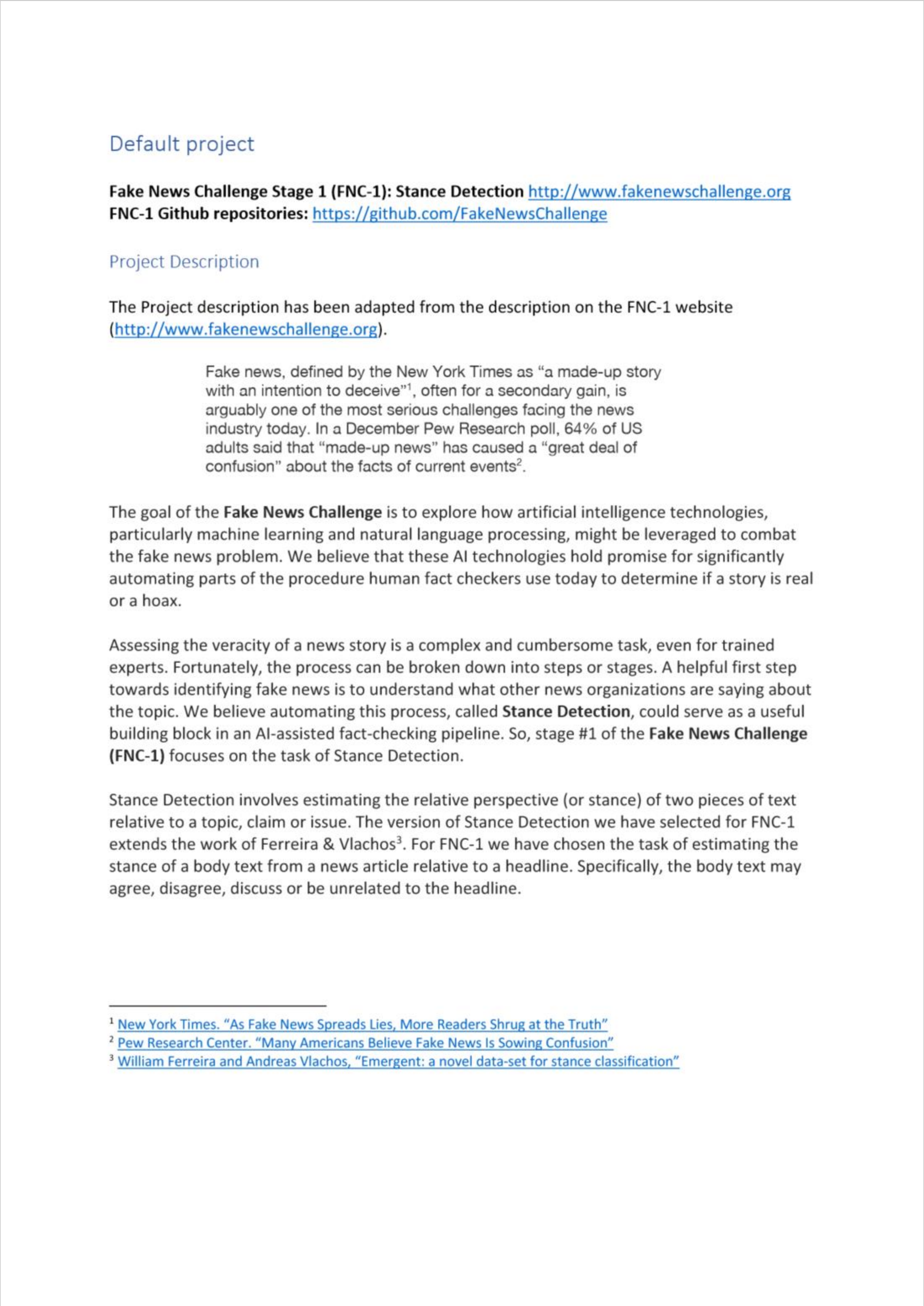
Project Outline

January 6, 2022 9:22 AM



Default project

Fake News Challenge Stage 1 (FNC-1): Stance Detection http://www.fakenewschallenge.org
FNC-1 Github repositories: https://github.com/FakeNewsChallenge

Project Description

The Project description has been adapted from the description on the FNC-1 website (http://www.fakenewschallenge.org).

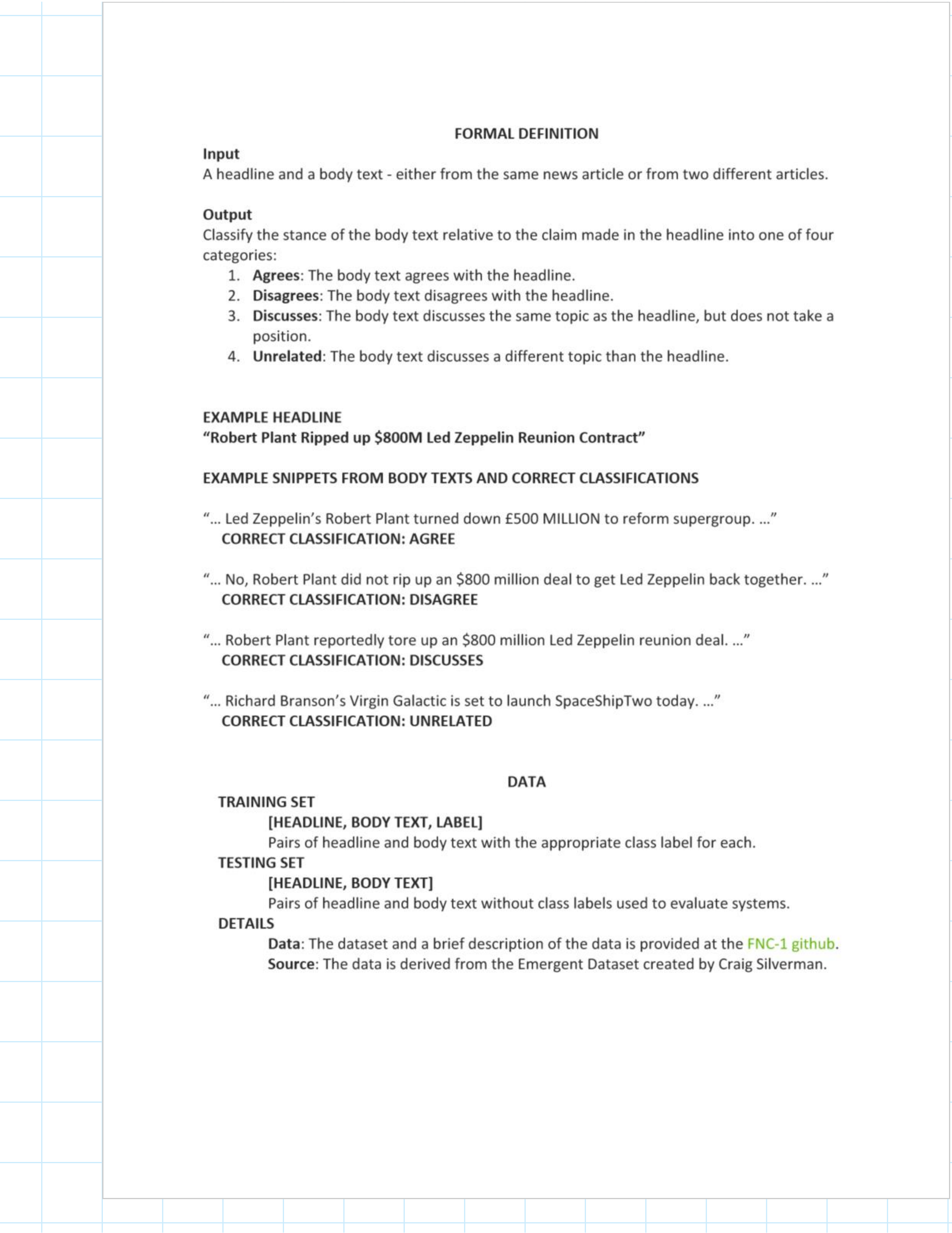
> Fake news, defined by the New York Times as "a made-up story with an intention to deceive"[1], often for a secondary gain, is arguably one of the most serious challenges facing the news industry today. In a December Pew Research poll, 64% of US adults said that "made-up news" has caused a "great deal of confusion" about the facts of current events[2].

The goal of the **Fake News Challenge** is to explore how artificial intelligence technologies, particularly machine learning and natural language processing, might be leveraged to combat the fake news problem. We believe that these AI technologies hold promise for significantly automating parts of the procedure human fact checkers use today to determine if a story is real or a hoax.

Assessing the veracity of a news story is a complex and cumbersome task, even for trained experts. Fortunately, the process can be broken down into steps or stages. A helpful first step towards identifying fake news is to understand what other news organizations are saying about the topic. We believe automating this process, called **Stance Detection**, could serve as a useful building block in an AI-assisted fact-checking pipeline. So, stage #1 of the **Fake News Challenge (FNC-1)** focuses on the task of Stance Detection.

Stance Detection involves estimating the relative perspective (or stance) of two pieces of text relative to a topic, claim or issue. The version of Stance Detection we have selected for FNC-1 extends the work of Ferreira & Vlachos[3]. For FNC-1 we have chosen the task of estimating the stance of a body text from a news article relative to a headline. Specifically, the body text may agree, disagree, discuss or be unrelated to the headline.

[1] New York Times. "As Fake News Spreads Lies, More Readers Shrug at the Truth"
[2] Pew Research Center. "Many Americans Believe Fake News Is Sowing Confusion"
[3] William Ferreira and Andreas Vlachos, "Emergent: a novel data-set for stance classification"



FORMAL DEFINITION

Input

A headline and a body text - either from the same news article or from two different articles.

Output

Classify the stance of the body text relative to the claim made in the headline into one of four categories:

1. **Agrees**: The body text agrees with the headline.
2. **Disagrees**: The body text disagrees with the headline.
3. **Discusses**: The body text discusses the same topic as the headline, but does not take a position.
4. **Unrelated**: The body text discusses a different topic than the headline.

EXAMPLE HEADLINE

"Robert Plant Ripped up $800M Led Zeppelin Reunion Contract"

EXAMPLE SNIPPETS FROM BODY TEXTS AND CORRECT CLASSIFICATIONS

"... Led Zeppelin's Robert Plant turned down £500 MILLION to reform supergroup. ..."
CORRECT CLASSIFICATION: AGREE

"... No, Robert Plant did not rip up an $800 million deal to get Led Zeppelin back together. ..."
CORRECT CLASSIFICATION: DISAGREE

"... Robert Plant reportedly tore up an $800 million Led Zeppelin reunion deal. ..."
CORRECT CLASSIFICATION: DISCUSSES

"... Richard Branson's Virgin Galactic is set to launch SpaceShipTwo today. ..."
CORRECT CLASSIFICATION: UNRELATED

DATA

TRAINING SET

[HEADLINE, BODY TEXT, LABEL]

Pairs of headline and body text with the appropriate class label for each.

TESTING SET

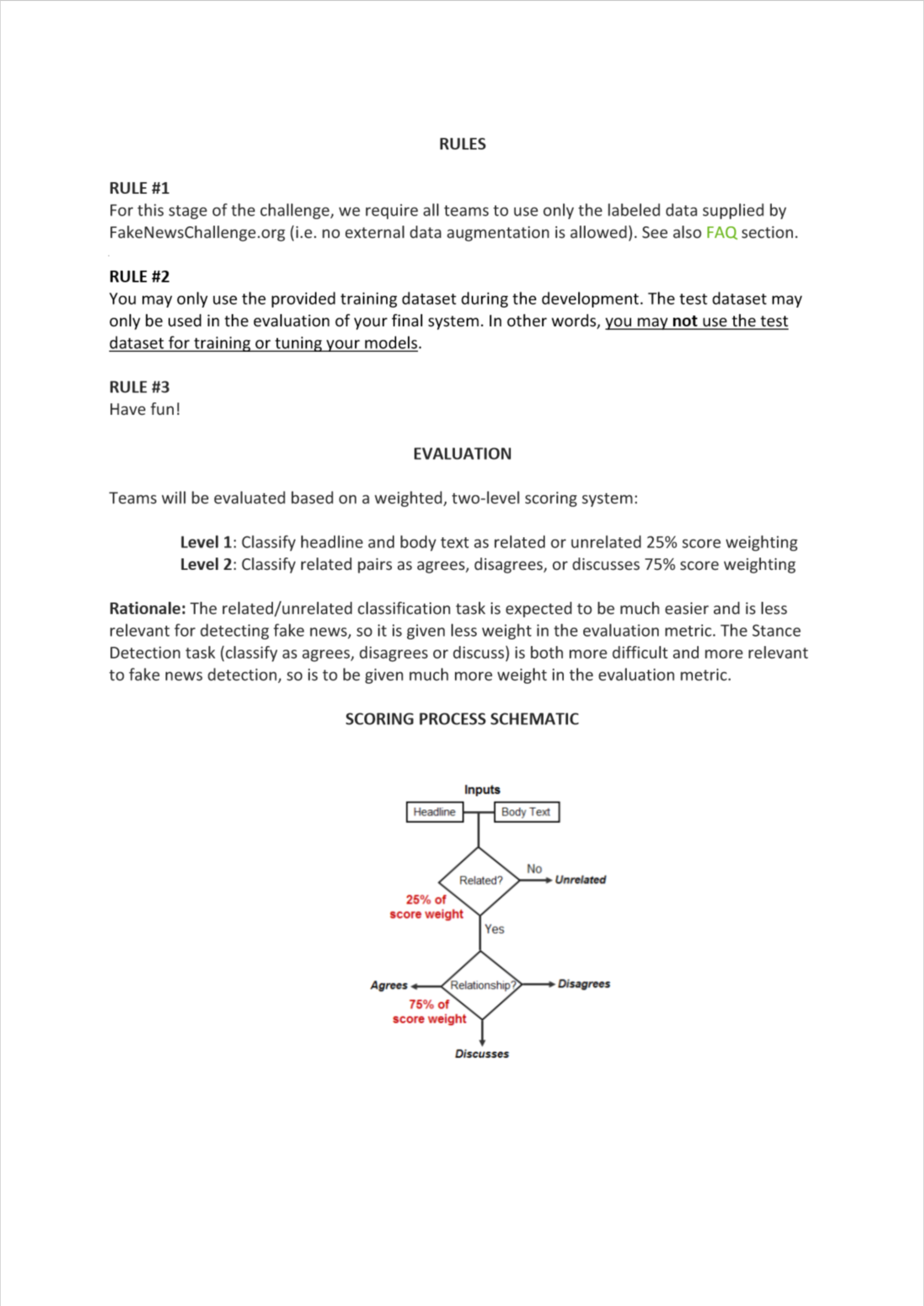
[HEADLINE, BODY TEXT]

Pairs of headline and body text without class labels used to evaluate systems.

DETAILS

Data: The dataset and a brief description of the data is provided at the FNC-1 github.

Source: The data is derived from the Emergent Dataset created by Craig Silverman.



RULES

RULE #1
For this stage of the challenge, we require all teams to use only the labeled data supplied by FakeNewsChallenge.org (i.e. no external data augmentation is allowed). See also FAQ section.

RULE #2
You may only use the provided training dataset during the development. The test dataset may only be used in the evaluation of your final system. In other words, <u>you may **not** use the test dataset for training or tuning your models</u>.

RULE #3
Have fun!

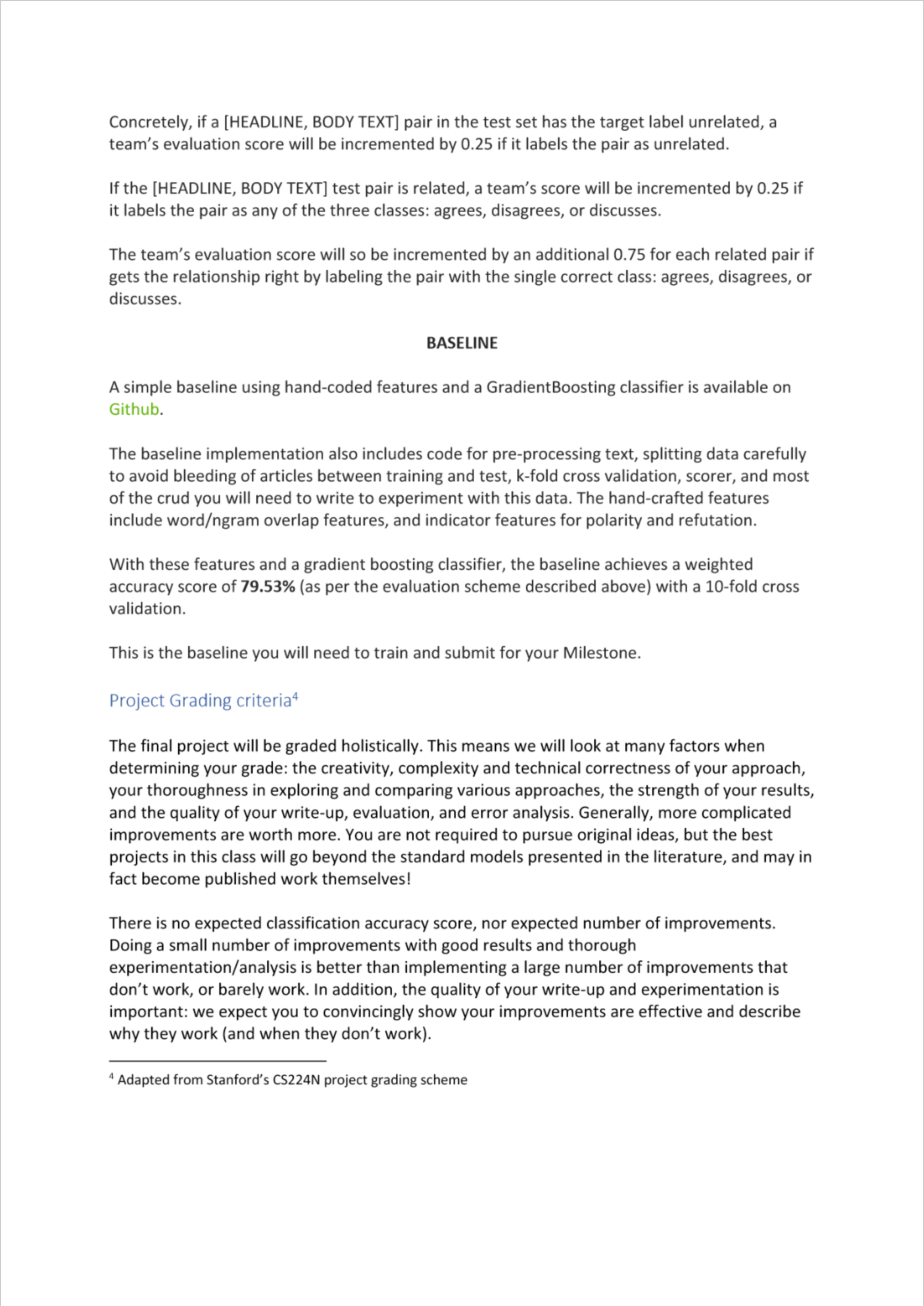
EVALUATION

Teams will be evaluated based on a weighted, two-level scoring system:

> **Level 1**: Classify headline and body text as related or unrelated 25% score weighting
> **Level 2**: Classify related pairs as agrees, disagrees, or discusses 75% score weighting

Rationale: The related/unrelated classification task is expected to be much easier and is less relevant for detecting fake news, so it is given less weight in the evaluation metric. The Stance Detection task (classify as agrees, disagrees or discuss) is both more difficult and more relevant to fake news detection, so is to be given much more weight in the evaluation metric.

SCORING PROCESS SCHEMATIC



Concretely, if a [HEADLINE, BODY TEXT] pair in the test set has the target label unrelated, a team's evaluation score will be incremented by 0.25 if it labels the pair as unrelated.

If the [HEADLINE, BODY TEXT] test pair is related, a team's score will be incremented by 0.25 if it labels the pair as any of the three classes: agrees, disagrees, or discusses.

The team's evaluation score will so be incremented by an additional 0.75 for each related pair if gets the relationship right by labeling the pair with the single correct class: agrees, disagrees, or discusses.

BASELINE

A simple baseline using hand-coded features and a GradientBoosting classifier is available on Github.

The baseline implementation also includes code for pre-processing text, splitting data carefully to avoid bleeding of articles between training and test, k-fold cross validation, scorer, and most of the crud you will need to write to experiment with this data. The hand-crafted features include word/ngram overlap features, and indicator features for polarity and refutation.

With these features and a gradient boosting classifier, the baseline achieves a weighted accuracy score of **79.53%** (as per the evaluation scheme described above) with a 10-fold cross validation.

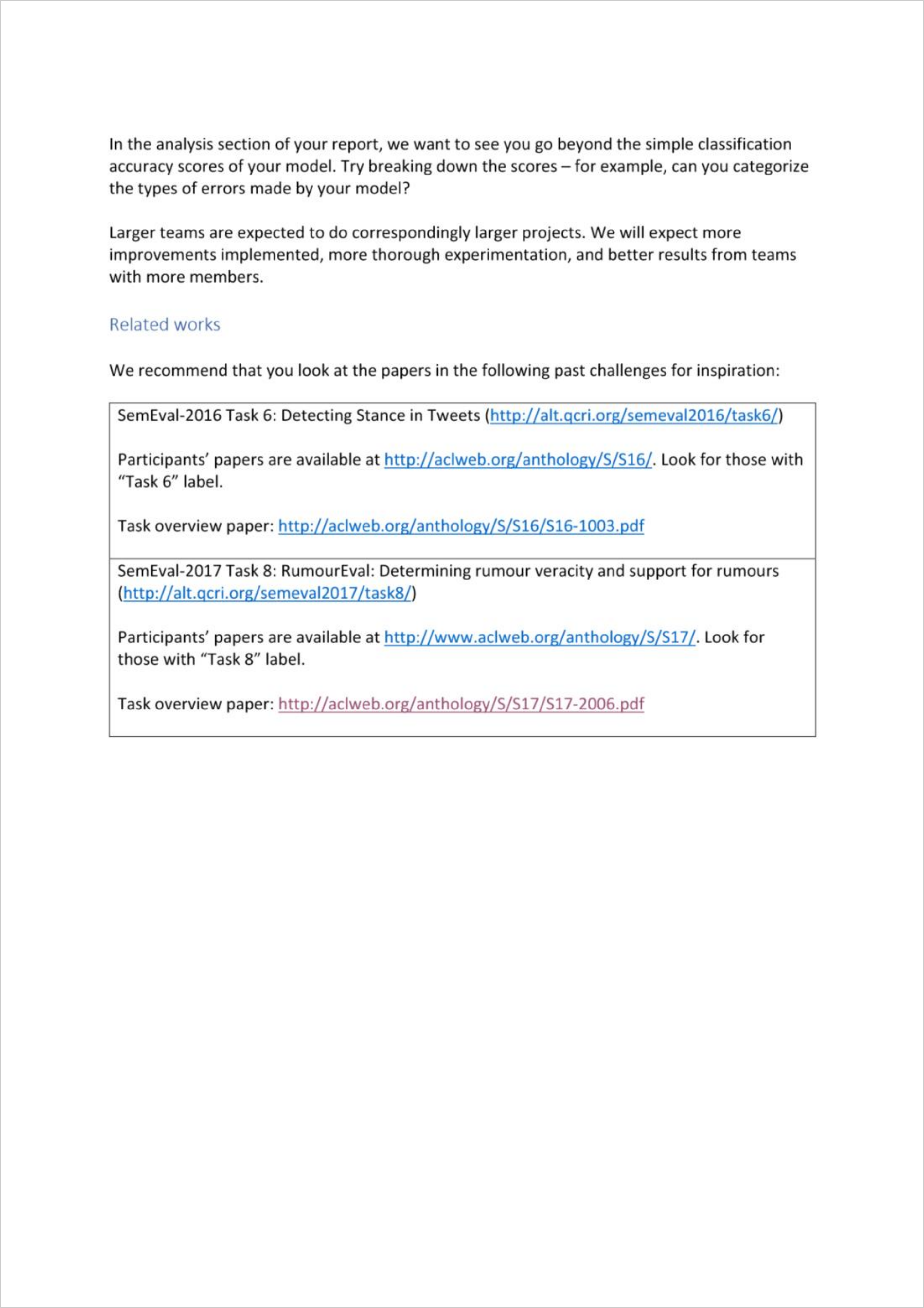
This is the baseline you will need to train and submit for your Milestone.

Project Grading criteria[4]

The final project will be graded holistically. This means we will look at many factors when determining your grade: the creativity, complexity and technical correctness of your approach, your thoroughness in exploring and comparing various approaches, the strength of your results, and the quality of your write-up, evaluation, and error analysis. Generally, more complicated improvements are worth more. You are not required to pursue original ideas, but the best projects in this class will go beyond the standard models presented in the literature, and may in fact become published work themselves!

There is no expected classification accuracy score, nor expected number of improvements. Doing a small number of improvements with good results and thorough experimentation/analysis is better than implementing a large number of improvements that don't work, or barely work. In addition, the quality of your write-up and experimentation is important: we expect you to convincingly show your improvements are effective and describe why they work (and when they don't work).

[4] Adapted from Stanford's CS224N project grading scheme



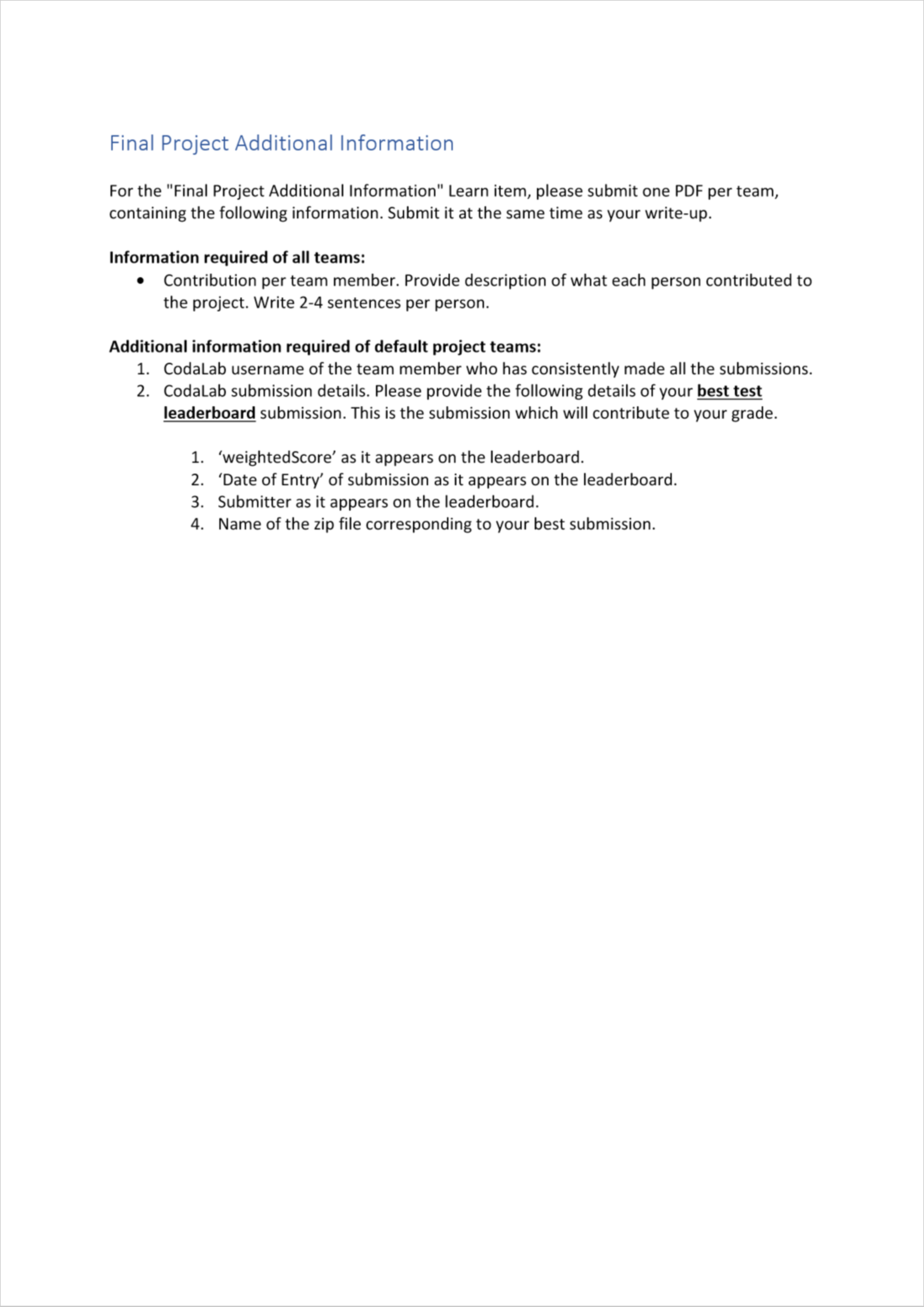
In the analysis section of your report, we want to see you go beyond the simple classification accuracy scores of your model. Try breaking down the scores – for example, can you categorize the types of errors made by your model?

Larger teams are expected to do correspondingly larger projects. We will expect more improvements implemented, more thorough experimentation, and better results from teams with more members.

Related works

We recommend that you look at the papers in the following past challenges for inspiration:

SemEval-2016 Task 6: Detecting Stance in Tweets (http://alt.qcri.org/semeval2016/task6/) Participants' papers are available at http://aclweb.org/anthology/S/S16/. Look for those with "Task 6" label. Task overview paper: http://aclweb.org/anthology/S/S16/S16-1003.pdf
SemEval-2017 Task 8: RumourEval: Determining rumour veracity and support for rumours (http://alt.qcri.org/semeval2017/task8/) Participants' papers are available at http://www.aclweb.org/anthology/S/S17/. Look for those with "Task 8" label. Task overview paper: http://aclweb.org/anthology/S/S17/S17-2006.pdf



Final Project Additional Information

For the "Final Project Additional Information" Learn item, please submit one PDF per team, containing the following information. Submit it at the same time as your write-up.

Information required of all teams:

- Contribution per team member. Provide description of what each person contributed to the project. Write 2-4 sentences per person.

Additional information required of default project teams:

1. CodaLab username of the team member who has consistently made all the submissions.
2. CodaLab submission details. Please provide the following details of your **best test leaderboard** submission. This is the submission which will contribute to your grade.

    1. ‘weightedScore’ as it appears on the leaderboard.
    2. ‘Date of Entry’ of submission as it appears on the leaderboard.
    3. Submitter as it appears on the leaderboard.
    4. Name of the zip file corresponding to your best submission.