

Webscience Project 2 - Review

Assessment of:

449e3d6f886deec705c7842d2f5bdbbc.zip

April 2, 2016

1. **Does the report answer correctly the project questions?**

Yes, both a naive solution as well as a deep learning solution have been tested using 3-fold cross-validation.

2. **Does the report break the project guidelines?**

No, the report fully lived up to the guidelines.

3. **Are there portions of the report unrelated to the project questions?**

The report is very concise and at no point diverges from the project questions.

4. **Are there sufficient examples to support the author's points?**

All results are listed and referenced clearly, no points are made without sufficient data.

5. **Is the overall organization of the report clear and effective?**

Yes, tables are clearly defined and referred, all sections contain only information relevant to the section.

6. **What are the report's main strengths?**

It is very clear and concise, references are spot on, nice with data in appendix and references to this, a great quick overview of the different methods used. Readability is very high.

7. **What are the report's main weaknesses?**

While it is noted that there is overwhelmingly many positive labels in our data, how many percent of these is not explicitly written. Writing that it is approximately 55% positive labels, diminishes the result of 60% and 67% hit rate, as they functionally only exceed by very little as opposed to just answering positive each time.

8. **Recommendations concerning the revision of this project**

A description of the data in the beginning would be nice, this would help to show the significance of the results. Sections should never be empty as with **2. METHODOLOGY** and **2.2 Sentiment Evaluation**, a short intro here would be great. Remove "TEMPLATE"/"Keep anonymous" from title. Appendix is referred to as Appendix A, there is no Appendix A.

9. **Summarization of**

(a) **Technical quality**

The code seems very well documented and very well organized, the imports are a bit of a mess though. All file names are hard coded, this is bad practise as it makes the code less reusable. I like that reformatted data is saved by pickling. Impressive that it automatically downloads data, great stuff.

(b) **Presentation quality**

The report presents the problem very well, as well as describes the way to attack the problem. Small mishaps as mentioned above

with the title, appendix and empty sections. But these are in my opinion very minor.

(c) **Adequacy of citation**

Everything that needed citation is cited, great job.