

CE807 – Assignment 2 - Final Practical Text Analytics and Report

Student id: XXX

Abstract

Say what you are doing in short.

You must write report using ACL format. Note the citation method like OLID (Zampieri et al., 2019) dataset. The report needs to be in 4-6 pages, excluding References.

You must refer any Table Figure in the report in the text. For e.g., in Table 1, we show the dataset statistics, we observed that blah blah

1 Materials

You must provide the clickable link to the Google Colab code and Recorded presentation. For example, this is my link: [LaTeX-Tutorial](#). Make sure that link is clickable.

- [Code](#)
- [Google Drive Folder](#) containing models and saved outputs
- [Presentation](#)

2 Model Selection (Task 1)

2.1 Summary of 2 selected Models

2.2 Critical discussion and justification of model selection

- It should contain the Figure 1 of both models

3 Design and implementation of Classifiers (Task 2)

In this section, you should add

- data set details as a Table and discuss
- Model implementation details like Hyperparameters
- You should compare the selected model's performance

- provide interesting examples in a Table 4 , and explain the model's output. See Table 7 for inspiration at <https://peerj.com/articles/cs-559/>

Dataset	Total	% OFF	% NOT
Train	xxx	xxx	xxx
Valid	xxx	xxx	xxx
Test	xxx	xxx	xxx

Table 1: Dataset Details

Model	F1 Score
Model 1	xxx
Model 2	xxx
SoA (?)	xxx

Table 2: Model Performance

4 Data Size Effect (Task 3)

In this section, you should add

- data set details for different data sizes as a Table and discuss
- Did you change any Hyperparameters in the model
- Critical Discussion of performance compared to different data size
- You must report performance on both validation and test as plots (You can take inspiration from the “Data hunger of the model” part in <https://peerj.com/articles/cs-559/> see Figure 2, 3 4.)
- provide interesting examples in a Table 5 and 6 for both models and explain the model's output. See Table 7 for inspiration at <https://peerj.com/articles/cs-559/>

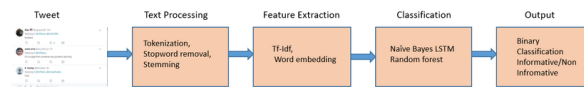


Figure 1: Diagram Explaining your pipeline and models. Note that this is a sample pipeline, you need to modify it to fit with your model. Only saying LSTM is not enough, you need to provide a bit more details. And diagram should be visible, this one is too small.

Data %	Total	% OFF	% NOT
25%	xxx	xxx	xxx
50%	xxx	xxx	xxx
75%	xxx	xxx	xxx
100%	xxx	xxx	xxx

Table 3: Train Dataset Statistics of Different Size

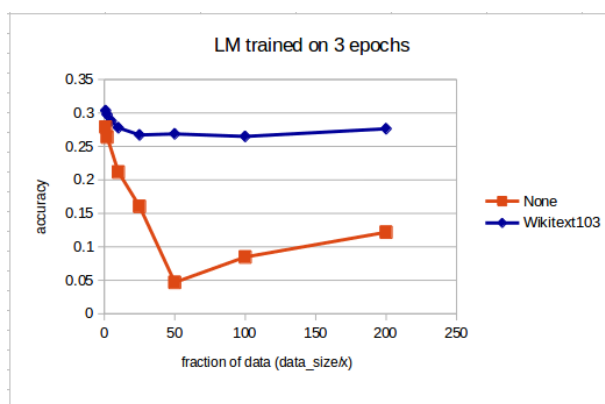


Figure 2: Comparison of Models based on Different data sizes. Note that you have 2 models, and you should report both validation and testing set performance.

You need to use Table 5 6 and 4 to compare the model's output and provide exciting insights. Note that you can't have same examples in all tables.

5 Summary (Task 4)

5.1 Discussion of work carried out

5.2 Lessons Learned

6 Conclusion

If you want to have.

References

Marcos Zampieri, Shervin Malmasi, Preslav Nakov, Sara Rosenthal, Noura Farra, and Ritesh Kumar. 2019. Predicting the Type and Target of Offensive Posts in Social Media. In *Proceedings of NAACL*.

Example %	GT	M1(100%)	M2(100%)
Example 1	xxx	xxx	xxx
Example 2	xxx	xxx	xxx
Example 3	xxx	xxx	xxx
Example 4	xxx	xxx	xxx
Example 5	xxx	xxx	xxx

Table 4: Comparing two Model's using 100% data: Sample Examples and model output using Model 1 & 2. GT (Ground Truth) is provided in the test.csv file.

Example %	GT	M1(25%)	M1(50%)	M1(75%)	M1(100%)
Example 1	xxx	xxx	xxx	xxx	xxx
Example 2	xxx	xxx	xxx	xxx	xxx
Example 3	xxx	xxx	xxx	xxx	xxx
Example 4	xxx	xxx	xxx	xxx	xxx
Example 5	xxx	xxx	xxx	xxx	xxx

Table 5: Comparing Model Size: Sample Examples and model output using Model 1 with different Data Size

Example %	GT	M2(25%)	M2(50%)	M2(75%)	M2(100%)
Example 1	xxx	xxx	xxx	xxx	xxx
Example 2	xxx	xxx	xxx	xxx	xxx
Example 3	xxx	xxx	xxx	xxx	xxx
Example 4	xxx	xxx	xxx	xxx	xxx
Example 5	xxx	xxx	xxx	xxx	xxx

Table 6: Comparing Model Size: Sample Examples and model output using Model 2 with different Data Size