

Computational Behavioural and Social Science (PS941)

Practical report 1

Feedback for 5637058

Mark: 63%

Grade: B

Feedback on your submission

Q1: Your code appears to be correct although you needed to submit raw R script for me to test it. Without it, I cannot test your scraper with different categories. Nonetheless, the code appears functional, and I am glad that you included custom headers and delays between requests. One additional way in which you could make it more robust would be to make sure that you catch various types of errors and error codes (e.g., 4xxx, 5xxx). Finally, although you provide good ethical considerations, you also needed to explain how to build a scraper in your tutorial. So while you included meaningful annotations, I was looking for a more detailed explanation of how one goes about scraping this particular website. Q2: Your implementation of the three models accurately shows underfitting, overfitting, and good fit (but unfortunately you did not include raw R script). You also correctly note the difference between MSEs between test and train data for the overfitting model. Your visualizations are useful, but are limited since they only capture one variable (a full picture would require a 3d plot). Nonetheless, good effort in achieving the main objective of this activity. Q3: You have correctly implemented a top-down and bottom-up model for classification. However, you could do more to explain the meaning of your results. You could also do more to visualize your results. For example, rather than just showing the tokens with higher tfidf scores, you could plot the tokens with highest coefficients from your Ridge regression model. Also, you could include ROC curves and confusion matrix, along with their interpretation.

Additional Note: You commented that Tabula did not accept .zip files, but most students were able to successfully submit a zip file. You must contact me and the main office as soon as you encounter an issue with Tabula, otherwise you face a risk of obtaining a lower grade for not including all necessary materials.