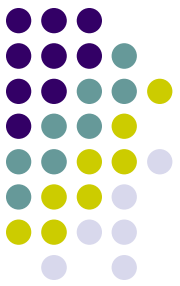


Big Data Analytics

Fourth Assignment



Deadline: 8 February 2021



Let's employ the movie review database corpus
and let's keep the first and the last 500
reviews

```
data("data_corpus_moviereviews", package =  
      "quanteda.textmodels")  
  
corp <- tail(data_corpus_moviereviews,  
             500)+head(data_corpus_moviereviews, 500)
```

To extract the actual texts of the movie reviews
and adding them as docvars in the corpus
you just extracted you can write:

```
docvars(corp, "texts") <- texts(corp)
```

Deadline: 8 February 2021



Apply a topic analysis; justify the number of topics extracted and then try to label them

Suggestion: I would probably avoid to stem the texts (but that's up to you!)