

Report nr.3

Project plan

Group: D15

Repo: <https://github.com/martentyrk/FakeNewsDetection>

Tools:

For our project we will be using Jupyter notebook, thanks to its easy GUI and the constant feedback the user gets from their code. Jupyter is also great for visualizing data.

Tasks:

- 1) Writing the documents needed for getting a better understanding of our project and also to finish homework 10.
 - a) We decided to do this collectively so that we would all be on the same page when starting the project. It took us around 2.5h to finish all the needed documents.
- 2) Create a .py template, where we will be writing our code.
 - a) Each team member will have to know the project structure so that would ideally be taking an estimated 0.6h from each of us.
- 3) Read in the data and clean it, removing any unnecessary cells and values. Fixing whitespace issues, removing null values, etc.
 - a) This task will take 2h collectively.
- 4) Research different methods for creating the best classifier.
 - a) This will probably take the most time, because we don't yet know how we will approach the problem. We will probably have everyone do their own research and then present their findings to the rest of the group.
 - b) Time: 5h from each member
- 5) Find the most popular words used in fake news and real news articles accordingly regarding one of our project goals.
 - a) Time: 2h
- 6) Start training the model and analysing the initial results.
 - a) 3h
 - b) Word2vec possibly
- 7) Visualize the performance of the models used regarding another one of our project initial goals.
 - a) 4h
- 8) Create a confusion matrix and a heatmap to show off the performance of the model
 - a) 3h

b)