

Computational Communication Science 2

Week 7 - Lecture

»Rule-based text classification and an introduction to supervised machine learning«

Marthe Möller

Anne Kroon

A.M.Moller@uva.nl, @MartheMoller

a.c.kroon@uva.nl, @annekroon

April, 2022

Digital Society Minor, University of Amsterdam

Today

Introducing...

...the people

An example with a reference

An example with some Python code...

All course materials can be found at...

<https://github.com/annekroon/CCS-2>

Introducing...

Introducing...

...the people

Introducing...
...the people

Introducing...



dr. Marthe Möller
Assistant Professor

- interested in...

@marthemoller | A.M.Moller@uva.nl, | <https://www.uva.nl/profiel/m/o/a.m.moller/a.m.moller.html>

Introducing...



dr. Anne Kroon

Assistant Professor Corporate Communication

- Research focus on biased AI in recruitment, and media bias regarding minorities
- text analysis using automated approaches, word embeddings

@annekroon | a.c.kroon@uva.nl | <http://www.uva.nl/profiel/k/r/a.c.kroon/a.c.kroon.html>

Introducing...



Your name?

Your background?

Your reason to follow this course?

Do you have a dataset you are working on?

An example with a reference

Why program your own tool?

Vis (2013)

“[...] these [commercial] tools are often unsuitable for academic purposes because of their cost, along with the problematic ‘black box’ nature of many of these tools.”

**An example with some Python
code...**

A simple bottom-up approach

```
1 from collections import Counter
2 texts = ["I really really really love him, I do", "I hate him"]
3
4 for t in texts:
5     print(Counter(t.split()).most_common(3))
```

```
1 [('really', 3), ('I', 2), ('love', 1)]
2 [('I', 1), ('hate', 1), ('him', 1)]
```

References

References



Vis, F. (2013). A critical reflection on Big Data: Considering APIs, researchers and tools as data makers. *First Monday*, 18(10), 1–16. <https://doi.org/10.5210/fm.v18i10.4878>