



## Text Analysis with Python (S006) February 03, 2023

Course location: Utrecht Science Park, Sjoerd Groenman Building room C024

Course Director: Dr. Ayoub Bagheri  
E-mail: ms.summerschool@uu.nl

Day	Time	Type	Description
Friday	09:00 – 10:00	Lecture	Introduction to text mining <ul style="list-style-type: none"><li>• Preprocessing</li><li>• Vector space model</li></ul>
	10:00 – 10:55	Computer lab	Jupyter notebook & Google colab Python environment (NLP packages in Python) Text preprocessing Preparing text data
	11:00 – 11:20	Plenary discussion	We will discuss the solutions to the computer lab
	11:30 – 12:30	Lecture	Text classification <ul style="list-style-type: none"><li>• Bag-of-words (tf, tfidf, n-grams)</li><li>• Classification (binary vs multiclass)</li><li>• Evaluation</li></ul>
	12:30-13:30	Lunch break	
	13:30 – 14:25	Computer lab	Document-term matrices Classification task on a text dataset Evaluation
	14:30 – 14:50	Plenary discussion	We will discuss the solutions to the computer lab
	15:00 – 16:00	Lecture	Word embedding <ul style="list-style-type: none"><li>• Distributional hypothesis</li><li>• CBOW vs Skip-gram</li></ul>
	16:00 – 17:00	Computer lab and discussion	Keras Embedding layer