Universiteit Utrecht

summerschool UTRECHT

# Data Science: Applied Text Mining, from Foundations to Advanced (S42)

## 13 – 17 July 2026

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

| Day | Time | Type | Description | Location |
|---|---|---|---|---|
| Monday | 09:00 – 10:30 | Lecture | Introduction to Text Mining and Natural Language Processing (NLP)<br>Preprocessing text<br>Feature extraction | Koningsberger building<br>Room 224 |
| | 10:50 – 12:00 | Computer Lab | Python & NLP in Google Colaboratory | |
| | 12:00 – 12:30 | Plenary Discussion | Students and teachers discuss and present their solutions to the computer lab | |
| | 12:30 – 14:00 | Lunch | | |
| | 14:00 – 15:20 | Lecture | Text classification:<br>- Binary classification<br>- Multi-class classification | |
| | 15:30 – 16:30 | Computer Lab | Document-term matrix<br>Text classification & its evaluation | |
| | 16:30 – 17:00 | Plenary Discussion | Students and teachers discuss and present their solutions to the computer lab | |

| Day | Time | Type | Description | Location |
|---|---|---|---|---|
| Tuesday | 09:00 – 10:30 | Lecture | Feature selection in text | Koningsberger building Room 224 |
| | 10:50 – 12:00 | Computer Lab | Filter-based feature selection<br>Embedded feature selection<br>Dimension reduction | |
| | 12:00 – 12:30 | Plenary Discussion | Students and teachers discuss and present their solutions to the computer lab | |
| | 12:30 – 14:00 | Lunch | | |
| | 14:00 – 15:20 | Lecture | Text clustering & topic modeling | |
| | 15:30 – 16:30 | Computer Lab | K-Means clustering<br>Topic modelling<br>Latent Dirichlet allocation | |
| | 16:30 – 17:00 | Plenary Discussion | Students and teachers discuss and present their solutions to the computer lab | |

| Day | Time | Type | Description | Location |
|---|---|---|---|---|
| Wednesday | 09:00 – 10:30 | Lecture | Word embeddings<br>Distributional hypothesis | Koningsberger building Room 224 |
| | 10:50 – 12:00 | Computer Lab | Word embeddings<br>Pre-trained embeddings | |
| | 12:00 – 12:30 | Plenary Discussion | Students and teachers discuss and present their solutions to the computer lab | |
| | 12:30 – 14:00 | Lunch | | |
| | 14:00 – 15:20 | Lecture | Feed-forward neural networks<br>Recurrent neural networks | |
| | 15:30 – 16:30 | Computer Lab | Neural networks<br>The embedding layer<br>Pretrained word embeddings | |
| | 16:30 – 17:00 | Plenary Discussion | Students and teachers discuss and present their solutions to the computer lab | |

| Day | Time | Type | Description | Location |
|---|---|---|---|---|
| Thursday | 09:00 – 10:30 | Lecture | Convolutional Neural Networks<br>Transformers | Koningsberger building<br>Room 224 |
| | 10:50 – 12:00 | Computer Lab | Recurrent neural networks<br>Convolutional Neural Networks<br>RNN vs CNN<br>Hyperparameter Optimization | |
| | 12:00 – 12:30 | Plenary Discussion | Students and teachers discuss and present their solutions to the computer lab | |
| | 12:30 – 14:00 | Lunch | | |
| | 14:00 – 15:20 | Lecture | Sentiment analysis & other applications | |
| | 15:30 – 16:30 | Computer Lab | Sentiment classification and analysis | |
| | 16:30 – 17:00 | Plenary Discussion | Students and teachers discuss and present their solutions to the computer lab | |

| Day | Time | Type | Description | Location |
|---|---|---|---|---|
| Friday | 09:00 – 10:30 | Lecture | LLMs pre-training, prompting, & learning from human feedback | Koningsberger building<br>Room 224 |
| | 10:50 – 12:00 | Computer Lab | Prompting LLMs | |
| | 12:00 – 12:30 | Plenary Discussion | Students and teachers discuss and present their solutions to the computer lab | |
| | 12:30 – 14:00 | Lunch | | |
| | 14:00 – 15:20 | Lecture | Responsible text mining<br>Bias and fairness in text mining | |
| | 15:30 – 16:30 | Computer Lab | NLP applications and transformers | |
| | 16:30 – 17:00 | Plenary Discussion | Students and teachers discuss and present their solutions to the computer lab | |

**For information about the Social Programme, please visit the Utrecht Summer School website!**