Data Science: Applied Text Mining, from Foundations to Advanced (S42)
14 – 18 July 2025

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) Preprocessing text Feature extraction	**Koningsberger building** 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	In *Vening Meinesz building*	
	14:00 – 15:20	Lecture	Text classification: - Binary classification - Multi-class classification	
	15:30 – 16:30	Computer Lab	Document-term matrix 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 Embedded feature selection 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	In *Vening Meinesz building*	
	14:00 – 15:20	Lecture	Text clustering & topic modeling	
	15:30 – 16:30	Computer Lab	K-Means clustering Topic modelling 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 Distributional hypothesis	**Koningsberger building** Room 224
	10:50 – 12:00	Computer Lab	Word embeddings 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	In *Vening Meinesz building*	
	14:00 – 15:20	Lecture	Feed-forward neural networks Recurrent neural networks	
	15:30 – 16:30	Computer Lab	Neural networks The embedding layer 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 Transformers	**Koningsberger building** Room Pangea
	10:50 – 12:00	Computer Lab	Recurrent neural networks Convolutional Neural Networks RNN vs CNN 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	In *Vening Meinesz building*	
	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** 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	In *Vening Meinesz building*	
	14:00 – 15:20	Lecture	Responsible text mining 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 <u>Utrecht Summer School website</u>!