Universiteit Utrecht

summerschool UTRECHT

Data Science: Applied Text Mining (S42)

26 – 29 July 2021

Course location:

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

Saturday 24 & Sunday 25 July 2021		
Time	**Activity**	**Description**
12.00 – 18.00	Key pick up	*You will find the exact key pick up location in the pre-departure information, which becomes available after you have paid the course fee.*

Day	Time	Type	Description	Location
Monday	09:00 – 10:30	Lecture	Introduction to text mining Preprocessing text Feature extraction	
	10:45 – 11:45	Computer Lab	Python in Google Colaboratory: - Cleaning text - NLTK, Gensim, spaCy	
	11:45 – 12:30	Plenary Discussion	Students and teachers discuss and present their solutions to the computer lab	
	14:00 – 15:15	Lecture	Text classification: - Binary classification - Multi-class classification Sentiment analysis	
	15:30 – 16:30	Computer Lab	Document-term matrix Sentiment classification News classification Classification 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: - Dimensionality reduction - Principal component analysis	
	10:45 – 11:45	Computer Lab	Pointwise mutual information PCA, t-SNE visualisation	
	11:45 – 12:30	Plenary Discussion	Students and teachers discuss and present their solutions to the computer lab	
	14:00 – 15:30	Lecture	Text clustering: - K-Means, MBK, DBScan - Non-negative matrix factorization - Topic modelling	
	15:45 – 16:30	Computer Lab	Text clustering Building an LDA model Visualise the topics-words distributions Optimal number of clusters Clustering evaluation	
	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 embedding: - Distributional hypothesis - Singular value decomposition - CBOW vs Skip-gram	
	10:45 – 11:45	Computer Lab	Embedding layer Retrieve Pre-trained embeddings Visualise the embeddings FastText (using subword information)	
	11:45 – 12:30	Plenary Discussion	Students and teachers discuss and present their solutions to the computer lab	
	14:00 – 15:30	Lecture	Convolutional neural networks Recurrent neural networks - LSTM, GRU - BERT	
	15:45 – 16:30	Computer Lab	Multi-class text classification	
	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	Text summarization Advanced deep learning with text: - Attention models - Text generation with deep learning	
	10:45 – 11:45	Computer Lab	Text summarization Simple attention model for text	
	11:45 – 12:30	Plenary Discussion	Students and teachers discuss and present their solutions to the computer lab	
	14:00 – 15:30	Lecture	Bias and fairness in text mining Responsible text mining	
	15:45 – 16:30	Computer Lab	A complete pipeline Advanced topics	
	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!