**Text Mining & NLP**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Module Coordinator | | Florian Ellsäßer | | | |
| Programme(s) | | FT – Master in Applied Data Science | | | |
| Term | | Semester 3 | | | |
| Module Duration | | 1 Semester | | | |
| Compulsory/ Elective Module | | Elective | | | |
| Credits | | 6 ECTS | | | |
| Frequency | | Annually,1 time | | | |
| Language of Instruction | | English | | | |
| Total Workload: | 150 | Contact hours | 44 | Independent  Learning | 100 |
| Prerequisites | | Introduction to machine learning and deep learning | | | |
| Content | | This module is focused on applying machine learning techniques to gain language understanding. Natural language processing is one of the main sub-fields of machine learning and has driven major algorithmic break-throughs in recent years. Language is a form of time series so break throughs in natural language processing such as LSTM networks have been closely connected to advances in machine learning in general.  The module is thus taking a twofold approach. On the one hand we will introduce general machine learning techniques that can deal with time series and show how they can be effectively applied to give computers language understanding. On the other hand, we will combine these techniques with domain specific applications such as word embedding, semantic distance and dependency tree parsing.  The module takes a practical approach combining theory with practice, so roughly 50% of the module will be theory and 50% will be practice. | | | |
| Intended Learning Outcomes | | After completion of this class students should be able to:   * Understand the latest machine learning techniques to gain language understanding through computational techniques. * Translate the knowledge gained on NLP algorithms to novel language processing problems. * Apply natural language processing techniques to business problems to better understand the sentiment of customers, their needs and how they may be persuaded. * Analyze the most advanced machine learning techniques such as LSTM networks in a domain specific context, in our case natural language processing. * Evaluate which model is most appropriate for a problem, based on accuracy and convergence metrics of the optimization. | | | |
| Forms of teaching, methods and support | | Most of the content that we are going to use will be in Jupyter notebooks. For each class, you will have to complete a small programming assignment in the Jupyter notebook. | | | |
| Type of Assessment(s) and performance points | | |  |  |  |  | | --- | --- | --- | --- | | **Type of Assessment** | **Duration** | **Performance**  **Points** | **Due Date oder Date of Exam** | | Individual assignments | 6 weeks | 60 | End of Class | | Continuous assignments | 2 weeks | 60 | Continuous | | | | |
| Recommended Literature | | There is no set text-book, but students are expected to read the recommended papers and texts for every class in advance of the class. | | | |
| Module Structure | | |  |  |  | | --- | --- | --- | | **Session** | **Topic** | **Preparation** | | **1** | Introduction | Read lecture material | | **2** | Part of Speech Tagging, Dependency Parsing | Read lecture material | | **3** | Semantics I | Read lecture material | | **4** | Semantics II and Transfer Learning | Read lecture material | | **5** | Sequence to Sequence Modelling | Read lecture material | |  |  |  | |  |  |  | | | | |
| Usability in other modules/programmes | | Frontiers of AI | | | |
| §” | | *Approval Date by Programme Director and publishing date by programme assistant.* | | | |