Student Research Project (Studienarbeit)
in Application Subject Medicine Computer Science & Mathematics

**BERT vs. SVM – Can deep learning beat classical machine learning in biomedical text classification?**

**Goal**
- Feasibility of BERT-based approach for biomedical text classification
- Comparison to established Support Vector Machine (SVM)

**Background**
- Texts and their semantics are essential for biomedicine: e.g. in terms of classifying medical documents in terms of diagnosis or innovation-driven topic analysis.
- Besides established simple word representation (TF-IDF) and word embeddings with static vectors (word2vec), now it seems to be a new milestone is reached: Deep Learning transformer models with contextualized work representations and pre-trained language models (e.g. BERT).
- There is a lot of attention and movement in this field (e.g. fine-tuned BERT models such as BioBERT or ClinicalBERT). What are the requirements and limitations? How can we use them for biomedical text classification? Can it beat classical machine learning approaches?

**Content**
- Literature research about BERT (+ fine tuned models such as ClinicalBERT and BioBERT).
- Feasibility analysis with implementation of a BERT-based approach for biomedical text classification (training and test data will be provided).
- Comparison to established Support Vector Machine (SVM) with TF-IDF feature representation.

**Requirements:**
- Flexible start – just contact us!
- 6 credit module in the application subject medicine for Master Computer Science & Mathematics.
- 20 pages (English preferred) + final presentation.
- Independent and trusting work with support of our interdisciplinary team.

➤ Expandable to bachelor, master or other types of theses.