Stellingen

Text Mining to Support Knowledge Discovery from Electronic Health Records

1. The performance of systems that identify contextual properties of clinical concepts is impacted heavily by the type of clinical documents (this thesis).

2. Medical record databases with narratives benefit from word-clustering based normalization approaches to reduce feature dimensions prior to text mining (this thesis).

3. Propensity score models developed with unstructured textual information require a covariate filtering approach based on the association of the covariates with the outcome (this thesis).

4. Automated case-detection algorithms can facilitate large-scale epidemiological studies based on electronic healthcare databases (this thesis).

5. The set expansion approach can be used to detect cases with more specificity in case of high disease prevalence in electronic health care databases (this thesis).

6. The validation of clustering structures is the most difficult and frustrating part of cluster analysis. Without a strong effort in this direction, cluster analysis will remain a black art accessible only to those true believers who have experience and great courage (Jain and Dubes).

7. You shall know a word by the company it keeps (John Rupert Firth).

8. Like a double-edged sword, perfectionism and working-hard cuts both way.

9. We are going through the process where software will automate software (Mark Cuban).

10. Selecting right training data is important to avoid systematic bias and unwanted discriminatory behavior in machine learning models.

11. Artificial intelligence is no match for natural stupidity.