Auftrag Consultancy Project MScBA

Thema	**Deep Learning for the automatic detection of emotional content in text documents.**
Auftrag	**Background** **The Integrity Risk Management project adapts methods from the field of computer science to the domain of integrity research. Data scientists and integrity experts design methods that automatically identify integrity-related issues in text corpora (e.g., media and corporate publications). Media coverage of such issues is often accompanied by emotions like outrage (i.e. anger), sadness, disgust and surprise. Methods capable of automatically identifying the emotions from Robert Putchik's wheel of emotion (anger, fear, sadness, disgust, surprise, anticipation, trust, and joy) are, therefore, powerful tools that can aid in detecting integrity- related issues.** **Objectives:** **The proposed consultancy project aims at creating components capable of automatically identifying emotional content in text documents, by researching (i) methods, (ii) existing systems, and (iii) corpora for emotion detection.** **Project roadmap:** **1. A comprehensive literature review should provide an overview of relevant lexicons, gold standard corpora and methods.** **2. Based on the literature review the students extend/create an English and/or German evaluation corpus used for training and evaluating their emotion detection method.** **3. The following approaches for detecting emotional content are implemented:** **◦ a lexicon-based method that detects the mentioned emotions in text messages considering the grammatical structure provided by a third-party tool.** **◦ A deep-learning-based approach that uses embeddings (see literature) rather than emotion lexicons for identifying emotional content.** **4. Evaluate the created classifier based on the evaluation corpora from step (2) using standard evaluation metrics (precision, recall, F1).** **Project setup:** **The Integrity Risk Management (IRM) project will provide the project team with (a) access to a text pre-processing and grammar parsing Web service, (b) a corruption specific test corpus, and (c) access to a server from DAVis (Center for Data Analytics, Visualization and Simulation) that is suitable for performing deep learning tasks.** **Literature:** **- Bravo-Marquez, Felipe, Eibe Frank, Bernhard Pfahringer, and Saif M. Mohammad. "AffectiveTweets: A Weka Package for Analyzing Affect in Tweets."** ***Journal of Machine Learning Research*** **20, no. 1–6 (2019).** **- Poria, Soujanya, Navonil Majumder, Rada Mihalcea, and Eduard Hovy. "Emotion Recognition in Conversation: Research Challenges, Datasets, and Recent Advances." http://arxiv.org/abs/1905.02947.** **- Gensim – Word embeddings with Python; https://radimrehurek.com/gensim/**

	Project Team Profile **- at least 1-2 students with good Python or Java programming skills**
Sprache	**Based on the preference of the students either German or English**
Auftraggeber	**Integrity Risks Management Projekt**
Referent	**Albert Weichselbraun**
Termine	**Datum Besprechung Auftrag mit den Auftraggebern** **Datum Besprechung Projektplan, Disposition mit Auftraggeber und Referent Datum Besprechung Methodik mit Referent** **Datum Besprechung Ergebnisse mit Referent, Auftraggeber Datum Besprechung Implikationen mit Referent, Auftraggeber Datum Besprechung Präsentation mit Referent, Auftraggeber Datum Abgabe Projektbericht an Administration** **Datum Feedback des Referenten**
Studierende	**Himmet Kaplan, Sean McErlean, Roger Barras (Prio 1),**