This mini-track is composed of three excellent papers involving practice of predictive analytics via big data. These three papers present a diverse set of approaches, demonstrating the variety of ways in which text mining and prediction models can be applied to today’s complex decision situations ranging from US politics to stock price prediction, where the datasets handled are large and diverse.

In the first paper by Wozniak and Cogburn, large volume of text-based data related to the all remarks given by the three Secretaries of State, namely Secretary Clinton, Secretary Rice, and Secretary Albright, are analyzed. The paper is using the ASSANA methodology-Accelerating Social science Analysis for a New Age, which is a tool developed by the authors to conduct analysis of large-scale, unstructured, text-based data. Two research questions were examined mainly in this research: what the most critical keywords of each secretary are during their tenure and what the similarities are amongst those keywords and phrases. After revealing answers for each of the two questions, authors point out that main limitation of the study to be further investigated in future research is to conduct a context analysis. Although some keywords and phrases are revealed to be important due to their number of times used, these researchers are skeptical about the fact that the context that they are used in may differ and hence deserves further investigation.

The second paper by Hoffman et al. reports on an automated classification process for crowd work. It applies three different machine-learning algorithms to retrieve patent information and to assign levels of readability of them. The authors use complex patent abstracts simplified by laymen to be classified in two dimensions, i.e. correctness and readability. With regard to three performance measures, accuracy, f-score, and AUC, this study reveals that support vector machines-based model outperforms multi-layer perceptron and kNN models. Authors are critically stating that before putting this classifier into practice it would be appropriate to further optimize the SVM model.

In the third paper, Hagenau et al. propose a text mining-based stock price prediction. The study achieves this by employing various aggregation time intervals in terms of weeks and it uses various datasets (e.g. Deutsche Gesellschaft fur Adhoc-Publizitat and Reuters) to validate the approach. The findings reveal that trading strategies would be profitable if news momentum of the past weeks is used. News can be used for both short- and medium-term investments based on the satisfactory accuracy levels received at around 70%. The researchers of this study reported that their study was limited to company specific news and therefore needs to be enhanced to incorporate economic news and news from other countries.