

Data Analytics using *KNIME* open source tool

Dörheit, Eric
TU Berlin
Berlin, Germany
eric.doerheit@campus.tu-berlin.de

Bode, Olga
TU Berlin
Berlin, Germany
olga.bode@mailbox.tu-berlin.de

Poljak, Dorothea
TU Berlin
Berlin, Germany
@mailbox.tu-berlin.de

ABSTRACT

We present our results of the evaluation of the open source tool *KNIME* which is used for data analytics and data mining. We choose anomaly detection as the subject to evaluate *KNIME* as many methods of data analytics such as clustering, classification, time series analysis and statistical techniques are applicable to anomaly detection [1]. As a data set for the analysis we use the data provided for the *DEBS Grand Challenge 2012* [2].

[2] Z. Jerzak, T. Heinze, M. Fehr, D. Gröber, R. Hartung, and N. Stojanovic. The debs 2012 grand challenge. In *Proceedings of the 6th ACM International Conference on Distributed Event-Based Systems*, DEBS '12, pages 393–398, New York, NY, USA, 2012. ACM.

1. INTRODUCTION

2. FUNDAMENTALS

2.1 The Open Source Tool KNIME

2.2 DEBS 2012 Grand Challenge

2.3 Anomaly Detection

3. ANOMALY DETECTION WITH KNIME

3.1 Clustering

3.2 Classification

4. EVALUATION

5. CONCLUSIONS

6. REFERENCES

- [1] V. Chandola, A. Banerjee, and V. Kumar. Anomaly detection: A survey. *ACM Comput. Surv.*, 41(3):15:1–15:58, July 2009.