

# 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].

## 1. INTRODUCTION

For the evaluation of *KNIME* we first investigated the tool by following several white-papers provided by *KNIME*<sup>1</sup>. We start with describing the functionalities and the usage of *KNIME*. Then we explain the data used for the evaluation. Subsequently we provide an overview on anomaly detection based on [1]. In Section 2 we evaluate *KNIME* through applying anomaly detection on the data set described in Section 1.2.

### 1.1 The Open Source Tool KNIME

### 1.2 DEBS 2012 Grand Challenge

### 1.3 Anomaly Detection

#### 1.3.1 Based on Classification

#### 1.3.2 Based on Nearest Neighbor

#### 1.3.3 Based on Classification

#### 1.3.4 Statistical Anomaly Detection Techniques

## 2. ANOMALY DETECTION WITH KNIME

---

<sup>1</sup>Explain contents of white-papers

## 2.1 Clustering

## 2.2 Classification

## 3. EVALUATION

## 4. CONCLUSIONS

## 5. REFERENCES

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