

**Tutorial Proposal**  
**IEEE 15<sup>th</sup> International Conference on Industrial Informatics INDIN' 2017**  
**-The Undergoing Industrial Informatics R-Evolution-**

**Tutorial Title:**

Balancing Small Samples and Big Data -  
An Introduction to Time Series Feature Extraction for Industrial Applications

**Tutorial presenter:**

Dr. Andreas W. Kempa-Liehr

**Affiliation:**

Department of Engineering Science, University of Auckland

**Email:**

a.kempa-liehr@auckland.ac.nz

**1. KEYWORDS**

Feature Engineering, Data Mining, Machine Learning, Industrial Applications, Internet of Things, Industrie 4.0

**2. AIMS AND LEARNING OBJECTIVES**

- Understanding the data science process and its support for decision automation in industrial applications
- Fundamentals of feature engineering and its challenges for time series analysis
- Fundamentals of the FRESH algorithm (Feature Extraction based on Scalable Hypothesis tests)
- Integration of automated time series feature extraction with machine learning pipelines
- Use cases for industrial applications like predictive maintenance and process optimization



# Tutorial Proposal

## IEEE 15<sup>th</sup> International Conference on Industrial Informatics INDIN' 2017 -The Undergoing Industrial Informatics R-Evolution-

### 3. SHORT SUMMARY OF CONTENTS

Promising fields of application for machine learning are the Internet of Things (IoT) and Industrie 4.0 environments. In these fields, machine learning models anticipate future device states by combining knowledge about device attributes with historic sensor time series. They permit the classification of devices into risk classes with respect to a specific defect.

This tutorial introduces a distributed and parallel time series feature extraction algorithm on basis of the recently published Python library tsfresh, which allows to balance small samples (e.g. predictive maintenance) with big data volumes from sensor time series and enterprise data on basis of scalable hypothesis tests.

The tutorial will explain the use cases both from the application and machine learning point of view as well as implications for the enterprise architecture, and will demonstrate the integration of the automated time series feature extraction into machine learning pipelines. Since its publication in October 2016, the respective Github project (<https://github.com/blue-yonder/tsfresh>) has been starred nearly 1,900 times and benefits from a growing international user group.

### 4. TARGET AUDIENCE

- Data Scientists
- Enterprise Architects
- Software Engineers
- IoT and Industrie 4.0 Experts

### 5. DURATION

80-90 min.

### 6. SHORT BIOGRAPHY OF SPEAKER (max. ½ page)

Dr. Andreas W. Kempa-Liehr is a Senior Lecturer at the Department of Engineering Science of the University of Auckland, New Zealand, and an Associate Member of the Freiburg Materials Research Center (FMF) at the University of Freiburg, Germany.

Andreas received his doctorate from the University of Münster in 2004 and continued his research as head of service group Scientific Information Processing at FMF.

From 2009 to 2016 he was a professional analyst in different data science roles at EnBW Energie Baden-Württemberg AG and Senior Data Scientist at Blue Yonder GmbH.

His research is focussed on the synergy of data science and operations research for engineering applications.

