

## INTRODUCTION TO TEXT MINING

Text mining is the analysis of data contained in natural language text. The purpose of Text Mining is to process unstructured information, extract meaningful numeric indices from the text, and, thus, make the information contained in the text accessible to the various data mining algorithms.

While text analytics is an emerging field, applying advanced analytics technique to text data is a powerful means to uncover information concealed in document collections.

This programme provides a solid understanding of general text analysis mining techniques. During the programme, many use cases are introduced at a practical level.

### **At Course Completion:**

Participants will gain an understanding of the possibilities of the analysis of unstructured text provides to companies. This is often an underused resource with great potential.

**Duration:** 2 Days

**Requires:** RapidMiner, SQL or any other relational database on PC.

### **Delivered by:**

Dr. Markus Hoffman  
ITB Lecturer with research interests in virtual learning environments,

business intelligence, data mining and web 2.0.

### **Who Should Attend:**

Anyone who is interested in using unstructured / semi-structured text data. Some data mining skills and understanding is desirable.

### **Key Topics:**

- General introduction
- Text Mining Terminology
- Loading Text
- Text Pre-processing
- Document Vectors
- Classification Algorithms
- Clustering Algorithms
- Visualisation of text throughout the modelling stages
- Evaluation
- Deployment of findings to unseen data