



# Tech Briefing:

## Customized Distance Functions in KNIME

Bernd Wiswedel

KNIME

# Outline

---

- Few slides on custom distance functions in KNIME
- Three use cases – as workflows

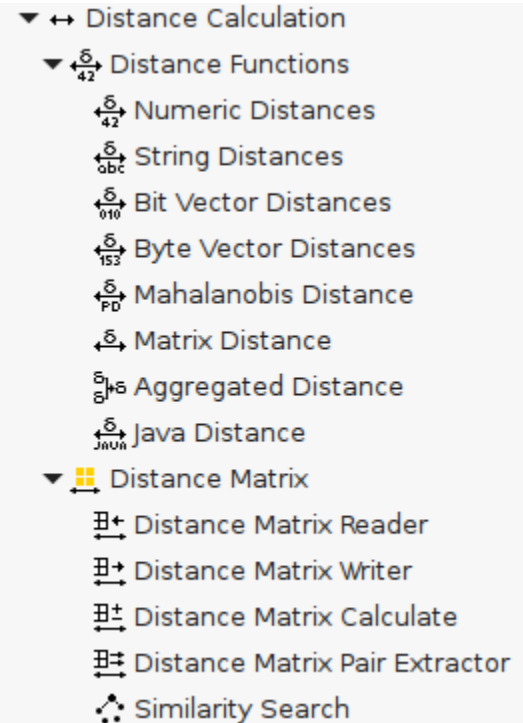
# Distance functions – why?

---

- Measure of how dissimilar two objects are
- Applications:
  - Classification: Distance-based methods, e.g. k-NN
  - Clustering: all methods rely on notion of distance
  - Querying: e.g. similarity search
  - ...

# Distance functions – in KNIME

- Custom port type describing a distance function
- Variety of nodes to define functions:
  - Numerical data (e.g. Euclidean)
  - Character sequences (e.g. string edit, n-grams)
  - Aggregates



# Three use cases

---

- Finding context similar terms in IMDB reviews (remember yesterday's use case?)
- Address de-duplication: Dealing with typos and other imprecision
- Random Forest Distance

The KNIME® trademark and logo and OPEN FOR INNOVATION® trademark are used by KNIME.com AG under license from KNIME GmbH, and are registered in the United States.