

# Wissensbasierte Systeme - ML-Übung #1

## "Nächste Nachbarn" und "Random Forest"

**Dozenten:** Dr. Sina Keller, M.sc. Felix Riese

**Datum:** 22.11.2017

**Datensatz:** EnMAP-Contest, [Download der Daten](#) mit [Beschreibung](#)

## Aufgabenstellung

Setzen Sie die folgenden Aufgaben in einem **Jupyter Notebook** (Vorlage finden Sie im Ordner "02 Übungen") um und dokumentieren Sie Ihre Schritte nachvollziehbar.

Das fertige Notebook senden Sie nach Beendigung der Aufgabe an: [sina.keller@kit.edu](mailto:sina.keller@kit.edu) und [felix.riese@kit.edu](mailto:felix.riese@kit.edu).

Wenn Sie möchten, können Sie die Aufgaben in einem *2er-Team* lösen.

1. **Einlesen** des Datensatzes mit `h5py`, z.B. in ein Dictionary
2. **Überblick** über den Datensatz erhalten:
  - Ein paar Zeilen (= Datenpunkte) anschauen mit `.head(5)` und `.tail(5)`
  - Größe des Datensatzes und Anzahl möglicher Klassen herausfinden
  - Datensatz geeignet visualisieren in mehreren Plots
  - Karte `Im` geeignet visualisieren mit entsprechender Klasse
3. **Vorprozessierung** implementieren:
  - Reduzierung der Bänder (z.B. Zusammenfassen von Bändern)
  - Glättung (Filterung) der Spektren
  - Möglicherweise schon Dimensionsreduzierung implementieren
4. **Datensatz splitten** in Trainings- und Testdatensatz
5. **Klassifikator** (Random Forest sowie Nächster Nachbar) implementieren und testen
6. **Qualitätsmaße** für den jeweiligen Klassifikator implementieren und ausgeben lassen
7. **Klassifikationsergebnis** auf der Karte `Im` ausgeben lassen
8. **Klassifikator optimieren**
  - Parameter tunen
  - bessere Vorprozessierung (s. 3.)
  - ...

Optional für alle, die super schnell sind... :-): Testen Sie, ob die Klassifikation mit einem MLP auf dem EnMAP-Datensatz funktioniert!