

DHBW Karlsruhe, Angewandte Informatik  
Vorlesung: Wissensbasierte Systeme

Dr. Sina Keller ([sina.keller@kit.edu](mailto:sina.keller@kit.edu))  
Felix M. Riese ([felix.riese@kit.edu](mailto:felix.riese@kit.edu))



**Abgabe:** bis 15.04.2018, 23:15 Uhr MESZ

04. April 2018

# Übungsblatt 01

## Klassifikation mit RF, K-NN und ANN

Alle Programmierungen sollen in einem Jupyter Notebook erfolgen und entsprechend nachvollziehbar dokumentiert werden. Die Dokumentation sowie die Beantwortung der Fragen kann direkt im Notebook im Markdown-Format verfasst werden. Dieses Notebook reichen Sie bis spätestens zum Abgabedatum (**15.04.2018, 23:15 Uhr MESZ**) über das Moodle-Portal zur Bewertung ein. Das Übungsblatt ist Pflicht und bildet einen Teil Ihrer Note für das Teilmodul „Wissensbasierte Systeme“.

**Datensatz:** LUCAS-Datenbank (Download, Dokumentation)

### 1 Visualisierung der Daten

- Machen** Sie sich mit dem Paket pandas und den Daten **vertraut**. **Geben** Sie Teile des Datensatzes **aus** mit `.head(5)` und `.tail(5)`, `.describe()` und `.info()`.
- Finden** Sie die Größe des Datensatzes und die Anzahl der tatsächlich verwendeten Klassen **heraus**.
- Visualisieren** Sie den Datensatz geeignet in mehreren Plots (Histogramm, Pairplot, Heatmap...).

## 2 Vorprozessierung (BONUS)

- a) **Implementieren** Sie einen Filter zur Glättung der Spektren.
- b) **Implementieren** Sie eine Dimensionsreduzierung mit
  - i) der Zusammenlegung von Bändern.
  - ii) Hauptachsentransformation (*principal component analysis*, PCA).
  - iii) *t-distributed stochastic neighbor embedding* (t-dsne).
- c) **Implementieren** Sie eine Methode zur Skalierung (z.B. MinMax) der Daten.

## 3 Klassifikation

Sie verwenden in dieser Aufgabe das Python-Paket scikit-learn.

- a) **Splitten** Sie den Datensatz in Trainings- und Testdatensatz im Verhältnis 1:9.
- b) **Implementieren** Sie den Klassifikator Random Forest.
- c) **Implementieren** Sie den Klassifikator k-Nächste-Nachbarn.
- d) **Implementieren** Sie ein künstliches Neuronales Netz (ANN) als Klassifikator.
- e) BONUS: **Implementieren** Sie ein künstliches Neuronales Netz (ANN) in Keras (statt scikit-learn).
- f) **Trainieren** Sie beide Klassifikatoren auf dem Trainingsdatensatz und **predizieren** Sie auf den Testdatensatz.

## 4 Evaluierung und Visualisierung der Ergebnisse

- a) **Finden** Sie geeignete Qualitätsmaße bzw. Fehlermaße einer Klassifikation. **Implementieren** Sie diese Maße.
- b) **Evaluieren** Sie die Vorhersagen (Prediktionen) aus Aufgabe 3 anhand dieser Qualitätsmaße. **Interpretieren** Sie die Ergebnisse.
- c) **Visualisieren** Sie die Klassifikationsergebnisse in geeigneter Form.
- d) BONUS: **Werten** Sie den Einfluss der Vorprozessierungen aus Aufgabe 2 aus.