



Big Data Anwendungen

## Praxisaufgabe (Klassifikation und Clustering)

Kobe Bryant beendete seine Basketball Karriere am 12. April 2016. In seinem letzten Spiel für die NBA warf er 60 Punkte für die Los Angeles Lakers. Kobe Bryant begann seine Karriere bei der NBA im Alter von 17 und war einer der erfolgreichsten Basketballspieler der letzten Jahre.

Der vorliegende Datensatz umfasst 20 Jahre von Daten von Kobe Bryants Treffern und Fehlwürfen. Ihre Aufgabe ist es vorherzusagen, ob ein Wurf ein Treffer oder ein Fehlwurf ist.

Download: <https://www.kaggle.com/c/kobe-bryant-shot-selection>

- (a) Installieren und öffnen Sie ihr Datenanalyse-Tool.
- (b) Lesen Sie die Daten ein.
- (c) Klassifizieren Sie die Daten mit einem Entscheidungsbaum.  
Für alle folgenden Schritte: Prüfen Sie, ob der Schritt das Ergebnis verbessert.
- (d) Identifizieren Sie fehlende Werte und behandeln Sie diese.
- (e) Entfernen Sie alle Attribute, die auf Grund ihrer inhaltlichen Bedeutung nicht zum Ergebnis beitragen.
- (f) Finden Sie Ausreißer und entfernen Sie diese.
- (g) Wandeln Sie die Attribute mit dem Datentyp Objekt so um, dass sie vom Klassifikationsalgorithmus genutzt werden können.
- (h) Überlegen Sie, ob `minutes_remaining` und `seconds_remaining` zu einem sinnvollen gemeinsamen Attribut kombiniert werden kann und führen Sie ggfs. die Kodierung durch.
- (i) Versuchen Sie das bisherige Ergebnis durch Anpassung der Parameter des Entscheidungsbaums zu verbessern.
- (j) Versuchen Sie das Ergebnis mit Hilfe einer Support Vector Machine zu verbessern.
- (k) Versuchen Sie das Ergebnis mit Hilfe eines Neuronalen Netzwerks zu verbessern.