

# Programmmentwurf

## Data Science Code Prototyp am Beispieldatensatz mit Fahrzeugen

Es ist ein Fahrzeugdatensatz gegeben, in dem verschiedene Merkmale von Fahrzeugen genannt werden.

**Teil 1:** Dieser Fahrzeugdatensatz ist mit Hilfe statistischer Standardfunktionen zu analysieren, zu interpretieren und ggfs. zu bereinigen. Nutzen Sie visuelle, möglichst in die Programmierumgebung eingebettete Darstellungsarten in verschiedenen Diagrammen, um wichtige Zusammenhänge zu visualisieren. Nutzen Sie Kommentare, um zu beschreiben, was jeweils Interessantes zu erkennen ist.

**Teil 2:** Verwenden Sie zwei verschiedene Vorhersageverfahren, um die PS-Zahl anhand eines gegebenen Datensatzes ohne PS-Zahl vorherzusagen.

**Teil 3:** Nutzen Sie zwei Klassifikationsverfahren, um ein Fahrzeug bei sonstigen gegebenen Daten einer Zylinderzahl zuzuordnen.

**Für Teile 2 und 3:** Validieren und testen Sie Ihre Ergebnisse jeweils mit Evaluationsmethoden wie diese bei Data Science Projekten üblich sind. Stellen Sie jeweils die Ergebnisse Ihrer Verfahren vergleichend optisch dar und beschreiben Sie das jeweilige Endergebnis. Diskutieren Sie, für welche Lösung Sie sich entscheiden. Optimieren Sie in einer bis maximal drei Iteration(en) die gewählte Lösung. Kommentieren Sie ggfs., welche weiteren Möglichkeiten zur Verbesserung Sie noch anwenden könnten. Erstellen Sie eine Methode, mit der man die Daten, für die eine Vorhersage getroffen werden soll, manuell eingeben kann, und diese einem das Resultat liefert und auch optisch sinnvoll anzeigt.

**Als Eingangsdaten liegen vor:** CSV-Datei mit Daten wie oben beschrieben.

**Programmiersprachen:** wahlweise Python (Version 3, IPython Notebook) oder R (auch lauffähig abzugeben, aktuelle Version)

### Bewertungskriterien

- 1. Fachliche Bewertung (25 Punkte):** Lösungsqualität und Umfang der Betrachtung, Umsetzung von Data Science in einem Code-Prototyp, korrekte Verwendung von wichtigen Funktionen / Bibliotheken, Güte der Endlösung, Nutzung der erworbenen Kenntnisse aus der Vorlesung
- 2. Dokumentation (15 Punkte):** Dokumentation des Vorgehens der Datenauswertung im Sinne von Data Science, Codekommentare wie in der Informatik üblich, Codequalität

### Abgabe

Bearbeitung in Gruppen mit jeweils **2 Personen** (beide müssen eine lauffähige Version des Codes installiert haben) bis zum **18. April 2019 einzureichen über das Moodle**

### Lernsystem.

### Abzugeben sind:

Programm (Quellcode, lauffähig), Dokumentation (enthalten z. B. im IPython Notebook / Beschriftungen direkt an Diagrammen, wenn möglich / Codekommentare zur Dokumentation / logisch und selbsterklärend ist immer vorzuziehen / optional: zusätzlich maximal 2.000 Wörter in einem externen pdf-Dokument)