

Als Prüfungsleistung soll eine Studie zu einem Data-Analyse Thema erstellt werden. Das Thema kann frei gewählt werden.

Die Daten sollen in tabellarischer Form vorliegen und sich für eine Explorative Datenanalyse sowie für eine Modellierung eines Regressions- oder Klassifikationsproblems eignen.

Anhand des HTML-Berichts wird ein Auszug der Studie der Lerngruppe präsentiert (keine Bewertung der Präsentation).

Bei vertraulichen Daten können Einzeltermine für Coaching und Präsentation vereinbart werden.

Die Struktur der Analyse soll sich am CRISP-DM Modell oder einer anderen sinnvollen Struktur ausrichten. Mindestumfang:

- Aufgaben- und Problemstellung (Thesen sowie Einordnung als Regressions- und/oder Klassifikationsproblem)
- Beschreibung der Datenquelle
- Untersuchung der Daten (z.B. auf NA, Ausreißer) und falls erforderlich Aufbereitung der Daten
- Untersuchung von mind. 2 Thesen mit Methoden der EDA (d.h. Visualisierung, Transformation)
- Anwendung mind. einer Machine-Learning Methode und Beurteilung
- Kritische Bewertung/ Ideen für weitere Analysen

Abweichungen hiervon (Gewichtung zwischen EDA und ML) können individuell abgestimmt werden.

Gruppenarbeit bis zu 3 Personen ist möglich.

Der Umfang erhöht sich entsprechend:

- Untersuchung von mind. 2 Thesen pro Gruppenmitglied; Thesen sollten sich gut voneinander abgrenzen.
- Anwendung und Evaluation mind. eines ML-Modells pro Gruppenmitglied.

Die anderen Inhalte werden zusammen erarbeitet und verantwortet.

Auch hier können Abweichungen individuell abgestimmt werden, z.B. wenn sich ein Mitglied um eine ausführliche Aufbereitung der Daten kümmern, könnten die Thesen hier wegfallen.

Abzugebende Dokumente sind:

- HTML-Bericht und Ausdruck als PDF
- Zugehöriger R-Code (QMD-Datei)
- Datensatz (falls nicht frei verfügbar, sonst Download-Link)

Das PDF kann auch direkt aus der QMD-Datei erzeugt werden, dies ist aber nicht zwingend erforderlich.

Bewertungsschema

	Kriterium	Gewicht in %	Bemerkung
1	Agenda	3	Ist die Studie übersichtlich strukturiert? Gibt es einen roten Faden?
2	Beschreibung des Datensatzes und des Problems	21	Datenquelle. Beschreibung der einzelnen Datenfelder, Wertebereich, etc. Ist die Aufgabenstellung und/oder These ausführlich beschrieben und nachvollziehbar?
3	Problemkomplexität	16	Ist das gewählte Thema anspruchsvoll oder eher einfach und mit wenigen Standardstatements zu lösen? Üblicherweise haben komplexere Themen einen größeren Umfang an Folien.
4	Herangehensweise und Darstellung	39	Ist das Vorgehen so, wie in der Vorlesung geschult oder unterliegt es einer anderen sinnvollen Struktur? Wurden Thesen aufgestellt und bewiesen oder widerlegt? Sind die Ergebnisse transparent und nachvollziehbar dargestellt, z.B. mit Grafiken und Tabellen?
5	Über die Vorlesung hinausgehende Leistung	6	Gibt es Methoden, Ansätze, die zum Thema passen, aber in der Vorlesung nicht geschult wurden?
6	Bewertung/ Nächste Schritte	10	Wurde das Ergebnis kritisch bewertet? Gibt es eine Idee für weitere Schritte? Üblicherweise reicht die Zeit nicht aus, ein Thema erschöpfend zu bearbeiten.
7	Demo-Code	5	Wurde der Code mitgeliefert und ist er ausführbar?
		100	

Beispiele für Prüfungsleistungen

Beispiel: Housing (sehr ausführlich)

Beispiel: Titanic (etwas knapper gehalten)

Sollten Sie sich an anderer Literatur/ Beispielen als in der Vorlesung verwendet orientieren, so geben Sie die Quelle(n) an.

Es ist ok, sich Ideen aus anderen Quellen (z.B. über Kaggle) zu holen. Die Abgabe muss sich jedoch deutlich von der Originalquelle unterscheiden.

Es muss deutlich werden, dass die vorliegende Arbeit ihre eigene Analyse ist und Sie verstehen, was dort gemacht wird. Es ist daher ratsam, viel erklärenden Text zu verwenden.