

Semesterarbeit: Erstellen eines Data Product

Schliessen Sie sich mit 4 bis 5 Studierenden zu einer Gruppe zusammen. Die Semesterarbeit besteht aus 3 Teilen mit 2 Zwischenabgaben für Feedbacks (Mail an tani@zhaw.ch). Die Semesterarbeit wird benotet und erhält ein Gewicht von 40% für die Modulendnote.

1. KW 40-41: Thema, Fragestellung Daten und Planung

- **Thema, Problemstellung und Daten:** Legen Sie fest, in welchem Themenbereich Sie Predictive Analytics betreiben wollen. Mögliche Themen sind:

- Kreditwürdigkeit
- Immobilienpreise
- Ozonwerte

Es sollte sich um ein Regressions- oder ein Klassifikationsproblem handeln. Suchen Sie dafür geeignete Daten und beschreiben Sie das Vorhersageproblem. Hier eine Auswahl an Datenportalen:

- <https://www.kaggle.com/datasets>
- <http://snap.stanford.edu/data/>
- <https://archive.ics.uci.edu/ml/>

- **Literatur:** Machen Sie eine Literatursuche zum Thema. Geben Sie mindestens 3 Quellen (wennmöglich wissenschaftlicher Art) an und beschreiben Sie den Zusammenhang mit Ihrer Arbeit.
- **Arbeitsplan:** Notieren Sie Ihre Überlegungen zum praktischen Vorgehen: Arbeitsschritte, potenzielle Vorhersagemethoden, Gestaltung des Data Products.

Erstellen Sie ein Dokument mit Beschrieb der Problemstellung, der Daten, der Literatursuche und des Arbeitsplans inkl. Namen der Gruppenmitglieder und schicken sie es mir als PDF-Datei.

Abgabe: KW 42, 16. Oktober 2023

2. KW 43-47: Datenanalyse

- **Datenanalyse** Plausibilisieren Sie zuerst die Daten mittels deskriptiven Methoden (Tabellen, Grafiken), führen Sie Datenbereinigungen durch und erstellen Sie einen Analysedatensatz. Wenden Sie dann die gewählte(n) Vorhersagemethode(n) an den Analysedatensatz an. Dokumentieren Sie die prädiktive Performanz und Robustheit der Methode. Vergleichen Sie dazu 2-3 Modellvarianten oder vergleichen Sie die Resultate mit der Literatur. Der R-Code sollte vollständig ausführbar (keine "Code-Leichen") und kommentiert sein.
- **Bericht:** Schreiben Sie einen kurzen Bericht zur Datenaufbereitung, der Anwendung der Vorhersagemethode(n), den Resultaten und Ihren Erkenntnissen.

Abzugeben ist der Bericht im PDF-Format inkl. Namen der Gruppenmitglieder, die R-Skripte und die Daten (TXT, CSV oder XLSX).

Abgabe: KW 48, 27. November 2023

3. KW 48-50 Data Product

- **Data Product:** Erstellen Sie das geplante Data Product mit Shiny. Schreiben Sie eine kurze Benutzeranleitung zur Bedienung Ihrer Anwendung.

- **Bericht:** Fassen Sie Ihre Arbeiten (alle 3 Teile) in einem Abschlussbericht zusammen. Berücksichtigen Sie dabei die Feedbacks zu den Abgaben 1 und 2.

Abzugeben sind der Abschlussbericht im PDF-Format inkl. Namen der Gruppenmitglieder, die R-Skripte (Analyse und Shiny-App) und die Daten (TXT, CSV oder XLSX).

Abgabe: KW 51, 18. Dezember 2023