

Künstliche Intelligenz und Big Data Analytics

Alexander Hinzer

HTW Berlin

Wintersemester 24/25

Semesterprojekt

Die Belegaufgabe entspricht 50% der Modulbewertung. Sie kann alleine oder in Gruppen bearbeitet werden¹.

Präsentationen können in jeder Übungsveranstaltung durchgeführt werden, letzter möglicher Termin: 15.01.2025. Der Code und Ergebnisse müssen bei Moodle hochgeladen/verlinkt werden bis: 15.01.2025, 23:59.

Aufgabe: Setzen Sie ein selbst gewähltes Machine Learning Projekt um und präsentieren Sie Ihre Arbeit und Ergebnisse. Erfüllen Sie **alle** Anforderungen 1-6 und wählen Sie **eine** der Anforderungen (2-7) für eine tiefergehende Bearbeitung².

Anforderungen:

1. **Problemstellung:** Suchen Sie sich ein Problem, welches Sie mittels KI lösen möchten. Formulieren Sie eine klare Zielstellung.
2. **Datenbeschaffung:** Beschaffen oder erheben Sie einen Datensatz, der sich eignet, um ein ML-Modell für die Lösung dieses Problems zu trainieren. Setzen Sie sich kritisch mit dem Datensatz auseinander.
3. **EDA:** Analysieren Sie den Datensatz und visualisieren Sie relevante Zusammenhänge.
4. **Feature Engineering:** Bereinigen, Extrahieren und Transformieren Sie Ihre Daten zu sinnvollen Features für das Training eines ML-Modells.
5. **Modelltraining** Wählen Sie einen passenden ML-Algorithmus und trainieren Sie ein Modell für die Lösung ihrer Problemstellung.
6. **Modellevaluierung:** Bewerten Sie die Performance ihres Modells in einer Kreuzvalidierung.
7. *(Bonusaufgabe:) Evaluieren Sie Verbesserungsansätze durch andere Modelle, Hyperparameter-Tuning (z.B. mit Gridsearch) oder Erweiterung mit anderen Datensätzen.*
8. **Präsentation:** Präsentieren Sie Ihre Arbeit und Ergebnisse (10 - 15 Minuten, Gruppen +5 Minuten je Mitglied):
 - (a) Stellen Sie Ihr Problem, Ihren Datensatz und Zielstellung vor.
 - (b) Erklären Sie Ihren Lösungsansatz und gehen Sie auf Ihr Vertiefungsgebiet ein.
 - (c) Präsentieren und bewerten Sie Ihre Ergebnisse und Verbesserungsansätze.

¹Für Gruppenarbeit: Um eine individuelle Bewertung zu ermöglichen, kennzeichnen Sie die Verantwortlichkeiten für die Teilaufgaben, bzw. reichen ein Repository mit nachvollziehbarem git blame ein.

²Für Gruppenarbeit: wählen Sie eine Vertiefungen je Person

Weitere Hinweise

Projektauswahl:

Wählen Sie ein Projekt nach Ihren Interessen und Vorlieben aus. Vielleicht gibt es ein Problem, welches Sie für persönliche Zwecke gerne angehen möchten oder ein Thema welches Sie für Ihre Bachelor-Arbeit ins Auge gefasst haben und schonmal erste Versuche unternehmen möchten.

Möglicherweise lassen sich mit Ihrem Projekt nicht alle Anforderungen erfüllen, dafür aber ganz andere Herausforderungen des Machine Learnings bearbeiten. Abweichungen von den Anforderungen sind vollkommen in Ordnung, wenn Sie vorher mit mir abgesprochen wurden.

Besonders komplexe oder wenig erfolgversprechende Problemstellungen können auch gerne bearbeitet werden. Die Bewertung ist nicht von Ihrer Modellperformance abhängig. Erfordert die Lösung zu viele Teilschritte oder mehrere aufeinanderfolgende Modelle, ist eine Teillösung auch ausreichend. Am besten definieren wir dann zusammen eine individuelle Zielsetzung.

tl;dr - haben Sie Unsicherheiten bezüglich Ihrer Projektwahl und -umsetzung, sprechen Sie mit mir oder schreiben mir.

Vertiefung:

In den meisten Fällen wird Ihre Problemstellung und ihr Datensatz individuelle Herausforderungen bereitstellen, welche die notwendige Vertiefung vorgeben.

- Sind Daten nur schwierig zu finden, müssen selber erhoben oder Webseiten gescraped werden, ist ihre Vertiefung 2.
- Haben Sie einen umfangreichen aber strukturierten Datensatz, möchten sie viele aussagekräftige Statistiken und Visualisierungen erforschen oder interaktive Visualisierungen umsetzen, kann ihre Vertiefung 3. sein.
- Verwenden Sie sehr komplexe und/oder unstrukturierte Daten, die kreative Ansätze zur Merkmalsextraktion erfordern, kann ihre Vertiefung 4. sein.
- Möchten Sie ein Deep Learning Modell auf unstrukturierten Daten trainieren, eine eigene Architektur aufbauen oder Finetuning von bestehenden Modellen umsetzen, kann ihre Vertiefung 5. sein.
- Möchten Sie das bestmögliche Modell für Ihr Problem ermitteln, vertiefen Sie 6. oder 7. und vergleichen die Ergebnisse verschiedener Ansätze.

Bewertung:

Fehlende oder kaum bearbeitete Anforderungen in der Implementierung führen zu Punktabzügen.

Ansonsten erfolgt die Bewertung Ihres Projekts zum größten Teil auf Grundlage Ihrer Präsentation. Stellen Sie sicher, dass Ihre Problemstellung und Herausforderungen klar herüber gebracht werden. Fokussieren Sie sich auf die Darstellung ihres Vertiefungsgebietes und wie es zum Gesamtergebnis beigetragen hat. Schließen Sie den Bogen mit der Darstellung der Ergebnisse im Kontext Ihrer Problemstellung und Vorschlägen für zukünftige Verbesserungen.

Slides/Powerpoint sind nicht unbedingt notwendig, es kann auch am Notebook präsentiert werden. Dieses sollte dann aber auch gut strukturiert und aufgeräumt sein (Tipp: Jupyter Notebooks können auch zu Slideshows umgewandelt werden).