

Leistungsbeurteilung 259

Anbieter-Version 1.0

Verordnung 2021

Autor nicola.colic@bbb Baden.ch

Experte

Datum

Klasse

Nachname

Vorname

Hilfsmittel

Folgende Hilfsmittel sind erlaubt:

- Alle Unterlagen
- Internet

Prüfungsdauer

Sie arbeiten über die Dauer des Moduls an einem eigenen Projekt, für welches Sie 8 Lektionen (360 Minuten) Zeit bekommen.

Notenskala

Linear (Note = $1 + 5 * \frac{\text{erreichte Punkte}}{\text{maximale Punkte}}$), auf halbe Noten gerundet.

1 ist die schlechteste, 6 die beste Note.

Hinweise

- Diese Prüfung ist eine Einzelarbeit.
- Betrugereien wie Plagiate o.Ä. oder Betrugsversuche führen zum Ausschluss aus der Prüfung. In diesem Falle gilt die Leistungsbeurteilung als absolviert, aber nicht bestanden.
- Falls Sie die Abgabe für einen Teil dieser Leistungsbeurteilung unbegründet verpassen oder zu spät einreichen, erhalten Sie für den entsprechenden Teil keine Punkte.

Handlungsziele

		Aufgabe
1	Identifiziert und bewertet technische Varianten aus dem Machine Learning-Umfeld für die Lösung eines Problems.	3.2
2	Setzt die Anforderungen des Datenschutzes bei der Sammlung und Nutzung von Daten um.	1.2
3	Bereitet Rohdaten für die maschinelle Verarbeitung auf, ordnet und analysiert diese im Hinblick auf die Verwendung im Machine Learning-Umfeld.	2.4
4	Klärt die sequenzielle Abhängigkeit von Aktivitäten der gewählten Machine Learning Technologie und definiert die zeitliche Reihenfolge.	1 2 3 4
5	Evaluiert ein performantes und passendes Modell aus dem Machine Learning Umfeld für die Entwicklung einer ICT-Lösung.	4.4
6	Nimmt Dienste und ICT-Lösungen zum Einsatz von Machine Learning-Technologien in Betrieb.	3
7	Entwickelt nach Vorgaben eine ICT-Lösung mit der geeigneten Machine Learning-Technologie.	1 2 3 4

Bewertung

Aufgabe	Gewichtung	Punkte	Bewertung
	1	3	Die administrativen Vorgaben sind eingehalten.
1.1	1	3	Die Datensätze sind gemäss Anforderungen ausgewählt und eignen sich für ML.
1.2	1	3	Der Datenschutz wurde berücksichtigt.
2.1	1	3	Es wurde ein für die Analyse zielführendes Feld ausgewählt.
2.2	2	3	Die Berechnungen sind korrekt durchgeführt worden.
2.3	1	3	Die Grafik ist korrekt erstellt und sinnvoll ausgesucht worden.
2.4	1	3	Die Daten wurden korrekt skaliert, dort, wo es nötig und sinnvoll ist. <i>oder</i> Die Argumentation, warum keine Skalierung nötig ist, wurde stimmig aufgestellt.
3.1	1	3	Die Aufteilung wurde korrekt vorgenommen.
3.2	2	3	Es wurde ein passender Algorithmus gewählt, und das Modell korrekt berechnet.
3.3	1	3	Die Überprüfung wurde korrekt und ausführlich durchgeführt.
4.1	1	3	Es wurden die wichtigen, aussagekräftigen Felder richtig erkannt.
4.2	2	3	Es wurde eine passende Messmetrik gewählt und diese korrekt berechnet.
4.3	1	3	Es sind geeignete Berechnungen erstellt worden.
4.4	1	3	Die Erklärungen wurden sinnvoll aufgestellt.

Total:

51

Einleitung

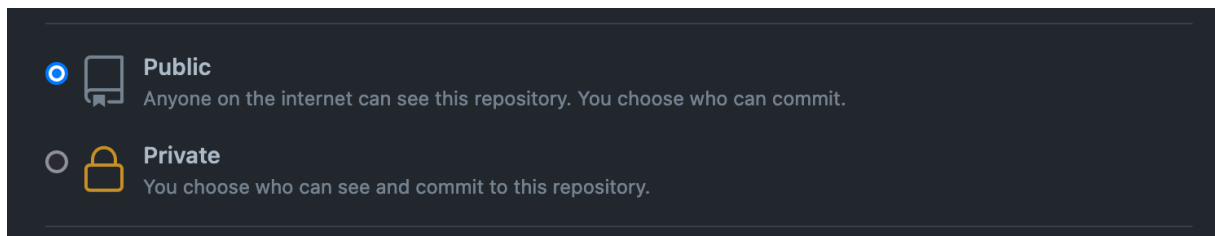
Es begleitet dieses Modul eine Projektarbeit, für welche Sie acht Lektionen Zeit bekommen, um daran zu arbeiten. Ziel ist es, dass Sie von Anfang bis Schluss eine einfache Lösung maschinellen Lernens implementieren; d.h. mittels eines Datensatzes ein Modell berechnen, mit welchem Sie Vorhersagen machen können, und dieses evaluieren. Sie sind dabei frei in der Wahl Ihres Datensatzes, Ihres Modells und Ihrer Frage; und werden sogar ermutigt, sich einen Themenbereich auszuwählen, der in Ihrem Interessensfeld liegt.

Die acht Projektarbeits-Lektionen finden verteilt über das Modul statt, sodass Sie das Gelernte an Ihrem eigenen Projekt umsetzen können.

Vorarbeit

Erstellen Sie ein neues *repository* auf github.com und geben Sie ihm einen Namen, der zu Ihrem Projekt passt. Sie werden die einzelnen Abgaben, die im Zuge dieses Projekts entstehen, dort hochladen und die Adresse Ihres *repository* jeweils abgeben. 5'

⚠ Stellen Sie sicher, dass Sie die Sichtbarkeit Ihres repository auf public setzen, damit Ihre Lehrperson Ihr Projekt bewerten kann.



Teil 1: Datenbeschaffung

Finden Sie sich einen geeigneten Datensatz, auf welchem Sie Ihr Projekt aufbauen. Beachten Sie dabei Folgendes:

- Er muss ausreichend groß sein, d.h. mindestens 10 000 Zeilen, Einträge o.ä. haben. Dies ist ein sehr grober Richtwert, welcher sich je nach Art des Datensatzes verändern kann; wichtig ist, dass Sie genug Daten haben, um aussagekräftige Modelle berechnen zu können.
- Er muss verschiedene Felder verfügen, welche auch im Alltag in einem Bezug zueinander stehen, d.h. gewisse Felder sollten in Abhängigkeit von anderen Feldern stehen, die aber nicht offensichtlich ist.
- Er muss frei zugänglich sein, d.h. unter einer CC-, MIT- oder vergleichbaren Lizenz veröffentlicht worden sein.

1.1 Laden Sie diesen Datensatz herunter, bereinigen Sie ihn und laden Sie ihn auf Ihr *repository* hoch. Falls der Datensatz zu groß ist, dann nehmen Sie die ersten 1000 Zeilen, Einträge o.ä. und laden Sie diese als Demonstration hoch. 40'

1.2 Verfassen Sie zusätzlich eine `README.txt`, in welcher Sie in 50 bis 100 40'

Wörtern Ihren Datensatz und seine Felder beschreiben. Beschreiben Sie in *weiteren* 50 bis 100 Wörtern, inwiefern der Datensatz Datenschutz-Richtlinien tangiert oder nicht, und welche Maßnahmen entweder von den Erstellern des Datensatzes oder von Ihnen getroffen wurden, um den Datenschutz zu gewährleisten. Fügen Sie dieses `README.txt` ebenfalls Ihrem *repository* hinzu.

1.3 Senden Sie die Adresse Ihres *repository* an Ihre Lehrperson. Sie finden hierzu auf Moodle ein Formular, über welches Sie die Abgabe tätigen können. 5'

💡 Sie finden eine gute Auswahl an Datensätzen beispielsweise auf [kaggle.com](https://www.kaggle.com) und huggingface.co. 😊

Teil 2

Machen Sie sich mit Ihren Daten vertraut und bereiten Sie sie auf die programmatische Verarbeitung vor. Wie dies konkret aussieht, hängt von Ihrem Projekt ab, aber in jedem Fall sollten Sie Ihren Code, Ihre Berechnungen und Ihre Erkenntnisse in einem Notizbuch namens `data_description.ipynb` zusammenfassen, welches Sie Ihrem *repository* hinzufügen. Mindestens sollten Sie dabei Folgendes bestimmen:

- | | |
|---|-----|
| 2.1 Für welches Daten-Feld möchten Sie Vorhersagen treffen können? | 20' |
| 2.2 Bestimmen Sie für jedes Feld einige relevanten statistischen Informationen wie beispielsweise Median und Standardabweichung. | 25' |
| 2.3 Erstellen Sie mindestens eine Grafik, beispielsweise ein Histogramm für eines der Datenfelder oder eine Regression für mehrere Felder. | 25' |
| 2.4 Skalieren Sie mindestens ein Datenfeld; oder argumentieren Sie in weiteren 50 bis 100 Wörtern in Ihrem Notizbuch, warum dies bei Ihrem Datensatz nicht sinnvoll ist. | 15' |
| 2.5 Falls sich die Adresse Ihres <i>repository</i> geändert haben sollte, so senden Sie es via Formular an Ihre Lehrperson. | 5' |

Teil 3

Erstellen Sie ein neues Notizbuch `model.ipynb`, in welchem Sie mindestens ein Modell berechnen und dieses benutzen, um eine Vorhersage zu treffen.

3.1 Teilen Sie Ihren Datensatz in einen *test*- und einen *train*-Satz ein. | 10'

3.2 Bestimmen Sie einen Algorithmus aus `sklearn`, der für Ihren Datensatz und Ihr Problem geeignet ist. Erklären Sie Ihre Entscheidung in 50 bis 100 Wörtern. Alternativ dazu können Sie auch verschiedene Algorithmen evaluieren und so bestimmen, welcher am besten geeignet ist. Berechnen Sie ein Modell mittels Ihres ausgesuchten Algorithmus auf Ihren *train*-Daten. | 50'

3.3 Testen Sie, wie gut Ihr Modell funktioniert, indem Sie einige Vorhersagen aus dem *test*-Datensatz generieren und diese manuell auf ihre Sinnhaftigkeit überprüfen. Fassen Sie Ihre Erkenntnisse in 50 bis 100 Wörtern zusammen. | 25'

3.4 Falls sich die Adresse Ihres *repository* geändert haben sollte, so senden Sie es via Formular an Ihre Lehrperson. | 5'

Teil 4

Erstellen Sie ein Notizbuch namens `evaluation.ipynb`:

- | | |
|---|-----|
| 4.1 Bestimmen Sie, welche Felder Ihrer Daten für Ihr Modell besonders aussagekräftig sind. | 20' |
| 4.2 Wählen Sie eine geeignete Messmetrik für Ihr Modell und berechnen Sie sie. | 20' |
| 4.3 Wählen Sie geeignete Bedingungen und erstellen Sie eine Wahrheitsmatrix für Ihr Modell. Berechnen Sie darüber hinaus Sensitivität und Spezifität. | 20' |
| 4.4 Fassen Sie in 50 bis 100 Wörtern zusammen, wie gut Ihr Modell funktioniert, und stellen Sie Hypothesen auf, warum. | 25' |
| 4.5 Falls sich die Adresse Ihres <i>repository</i> geändert haben sollte, so senden Sie es via Formular an Ihre Lehrperson. | 5' |