

Abgabe WBG-Projektabschluss

TH Brandenburg - FB Wirtschaft

Sommersemester 2024

!!! Abgabe am 19.7.2024 !!!

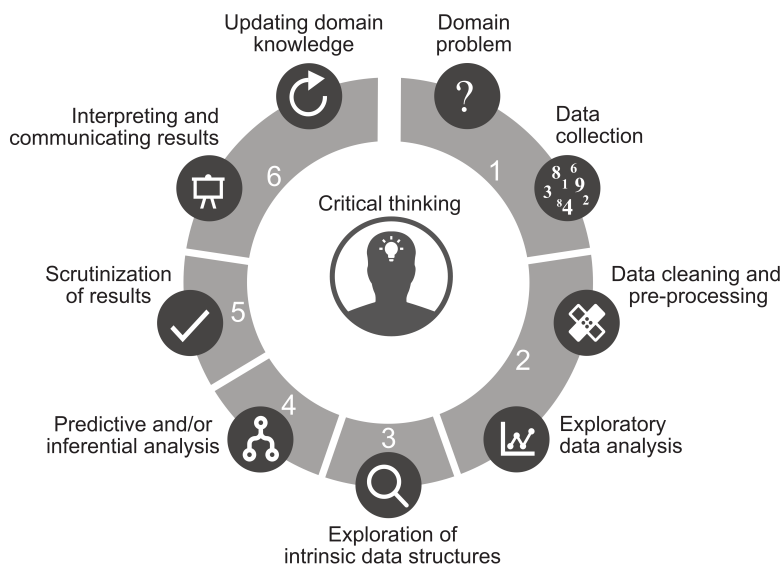
Kontext

Sie haben erfolgreich Ihre (bisherige) Arbeit im Rahmen des WBG-Projekts am 24.6.2024 präsentiert. Nun sollen Sie die endgültige Fassung abgeben, die die Ergebnisse Ihrer Arbeit dokumentiert und die Sie in den letzten Wochen weiterentwickelt haben.

Im Zentrum steht dabei die Anwendung von Verfahren des maschinellen Lernens aus den Bereichen des überwachten oder auch nicht-überwachten Lernens. **Grundlage bildet der Datensatz der WBG.**

Ein zentraler Aspekt der Abgabe ist der sogenannte **Data-Science Lifecycle**. Darunter versteht man den Prozess der Erstellung eines mathematischen Modells aus Daten, um Muster, Beziehungen und Gesetzmäßigkeiten in den Daten zu erkennen und für Vorhersagen zu nutzen. Die prädiktive Analyse umfasst Techniken wie Regression, Klassifikation und Clustering.

Orientieren Sie sich dabei an den folgenden Schritten:



Ihre Abgabe sollte alle zentralen Elemente dieses Zyklus umfassen. Dabei sollten Sie sich auf die **Datenaufbereitung**, die **explorative Analyse**, die **Modellierung**, **Evaluation** und **Interpretation** der Ergebnisse konzentrieren.

Abgabe -- entweder als Notebook, Text-Dokument oder Poster

Die finale Arbeit können Sie in Form eines **Python-Notebooks** oder eines **Text-Dokuments** abgeben. Sie können auch ein **Poster** erstellen, das die wichtigsten Ergebnisse Ihrer Arbeit zusammenfasst.

Die Vorlagen im ***.pptx** oder ***.otp** Format (für LibreOffice) für das Poster finden Sie auf der Moodle-Seite zu dieser Abgabe.

Umfang

Es ist keine Mindestanzahl an Seiten bzw. Länge der Notebooks vorgegeben. Achten Sie darauf, dass Sie alle Schritte und Entscheidungen, die Sie bei der Datenaufbereitung getroffen haben, klar und verständlich dokumentieren.

Hinweis zum Poster

Wenn Sie ein Poster abgeben sollte der **Anteil von Text und Abbildungen bzw. Tabellen ausgewogen** sein.

Nutzen Sie die **Formatvorlage**. Verändern Sie bitte nicht die **Schriftart und -größe**.

Einreichungsformat

Format der Einreichung als **zip-Datei** auf Moodle.

Die zip-Datei umfasst dabei:

- ausführbarer Python-Code oder pythonbasierte Notebooks (*.ipynb Dateien)
- Datendatei in einem geeigneten Format (z.B. csv, xlsx)
- ggf. weitere Datendateien, falls Sie weitere externe Daten verwendet haben
- Abgabe in Gruppen. Eine Einreichung pro Gruppe genügt!
- Text-Datei oder Poster als **PDF-Datei**
- Die Abgabe soll auf Deutsch erfolgen.

Bewertungskriterien

- Die **Fragestellung und Hypothese** sollten klar und verständlich sein und die Ergebnisse Ihrer Analysen angemessen darstellen. Achten Sie darauf, dass die Fragestellung und Hypothese gut lesbar und verständlich sind und dass Sie die Ergebnisse der Fragestellung und Hypothese klar und verständlich interpretieren.
- Die **Datenaufbereitung** sollte sorgfältig und nachvollziehbar dokumentiert sein. Achten Sie darauf, dass Sie alle Schritte und Entscheidungen, die Sie bei der Datenaufbereitung getroffen haben, klar und verständlich dokumentieren. Dies beinhaltet auch die Dokumentation von offenen Fragen und Unsicherheiten, die Sie bei der Datenaufbereitung hatten.
- Die **Visualisierungen** sollten klar und verständlich sein und die Ergebnisse angemessen darstellen. Achten Sie darauf, dass die Visualisierungen gut lesbar und verständlich sind und dass Sie die Ergebnisse der Visualisierungen klar und verständlich interpretieren.

- Die **Deskriptiven Statistiken** sollten klar und verständlich sein und die Ergebnisse Ihrer Analysen angemessen darstellen. Achten Sie darauf, dass die Statistiken gut lesbar und verständlich sind und dass Sie die Ergebnisse der Statistiken klar und verständlich interpretieren.
- Im **Fokus steht die Entwicklung** eines deskriptiven oder prognostischen Modells. Sie können Verfahren aus dem Bereich des überwachten oder auch nicht-überwachten Lernens verwenden. Verwenden Sie Verfahren aus dem Kurs. Wenn Sie Verfahren wählen, die im Kurs nicht behandelt wurden, wird dies positiv bewertet. Achten Sie jedoch darauf, dass Sie die Verfahren für Ihre Fragestellung anwendbar sind, und sie diese verständlich erklären und begründen!
- Die **Prognoseevaluation** sollte die üblichen Metriken umfassen. So z.B. für die Regressionsanalyse RMSE, MAE, MAPE, etc. Achten Sie darauf, dass Sie die Ergebnisse der Prognoseevaluation klar und verständlich darstellen und interpretieren.
- Die **Gesamte Präsentation** sollte klar und verständlich sein und die Ergebnisse Ihrer Analysen angemessen darstellen. Sie sollten Ihre Fragestellung im Auge behalten und die Ergebnisse nachvollziehbar interpretieren.

Individuelle Bewertung innerhalb der Gruppen

Für eine genauere Bewertung können Sie die individuellen Leistungsanteile innerhalb der Gruppe selbst verteilen (**ganz am Ende im Notebook**). Dafür werden von jeder Gruppe folgende Informationen benötigt:

Eine Prozentzahl von jedem Gruppenmitglied für jedes andere Gruppenmitglied.

Beispiel A

Anke hat einen Großteil der Arbeit erledigt und gleichzeitig alle anderen Gruppenmitglieder einbezogen und gleichzeitig individuelle Arbeitspakete verteilt und in das Gesamtergebnis integriert. Sie hat auch einen Großteil der Präsentation übernommen.

Jedem Mitglied in dieser 4er-Gruppe stehen 3 Punkte ($n - 1$) zur Verfügung. Anke verteilt ihre Punkte gleichmäßig auf alle anderen Mitglieder. Die Anderen haben Anke jeweils zwei Punkte gegeben und den dritten Punkt auf die anderen drei Mitglieder für ihre individuellen Beiträge.

==> Angedacht: ~20% Variation durch gegenseitige Bewertung der Gruppenmitglieder (z.B., 2,0 – 2,7 innerhalb einer Gruppe möglich)

Beispiel B

Werden diese Informationen nicht eingereicht, erhalten alle Gruppenmitglieder dieselbe Note.

Beispiel für die Bewertung einer Gruppenleistung (mit der Gruppennote vom obigen Beispiel) Fall A Person 1 nimmt an allen Veranstaltungen teil und Personen 2 Personen 3 und 4 lassen sich in der Vorlesung berieseln und nehmen nicht an den Übungen teil. => Alle Gruppenmitglieder erhalten eine 2,7