Data Exploration Project

Machine Learning Fundamentals

Prof. Dr. Maximilian Scherer

maximilian.scherer@dhbw-mannheim.de

SoSe 2024

Organisatorisches

Prüfungsleistung

- Jupyter Markdown Notebook (.ipynb)

Prüfungsleistung

- Jupyter Markdown Notebook (.ipynb)
  - Kombination aus Projektbericht, Quellcode und Kommentaren

Prüfungsleistung

- Jupyter Markdown Notebook (.ipynb)
  - Kombination aus Projektbericht, Quellcode und Kommentaren
  - Fließtext: 2-3 Seiten pP

Prüfungsleistung

- Jupyter Markdown Notebook (.ipynb)
  - Kombination aus Projektbericht, Quellcode und Kommentaren
  - Fließtext: 2-3 Seiten pP
  - Referenzen, Tabellen / Abbildungen

Prüfungsleistung

- Jupyter Markdown Notebook (.ipynb)
  - Kombination aus Projektbericht, Quellcode und Kommentaren
  - Fließtext: 2-3 Seiten pP
  - Referenzen, Tabellen / Abbildungen
  - Quellcode-Zellen mit Output

Prüfungsleistung

- Jupyter Markdown Notebook (.ipynb)
  - Kombination aus Projektbericht, Quellcode und Kommentaren
  - Fließtext: 2-3 Seiten pP
  - Referenzen, Tabellen / Abbildungen
  - Quellcode-Zellen mit Output
  - `https://www.kaggle.com/masumrumi/a-statistical-analysis-ml-workflow-of-titanic`

Prüfungsleistung

- Jupyter Markdown Notebook (.ipynb)
  - Kombination aus Projektbericht, Quellcode und Kommentaren
  - Fließtext: 2-3 Seiten pP
  - Referenzen, Tabellen / Abbildungen
  - Quellcode-Zellen mit Output
  - https://www.kaggle.com/masumrumi/a-statistical-analysis-ml-workflow-of-titanic
  - https://www.kaggle.com/c/titanic/code?competitionId=3136&sortBy=voteCount

Prüfungsleistung

- Jupyter Markdown Notebook (.ipynb)
  - Kombination aus Projektbericht, Quellcode und Kommentaren
  - Fließtext: 2-3 Seiten pP
  - Referenzen, Tabellen / Abbildungen
  - Quellcode-Zellen mit Output
  - `https://www.kaggle.com/masumrumi/a-statistical-analysis-ml-workflow-of-titanic`
  - `https://www.kaggle.com/c/titanic/code?competitionId=3136&sortBy=voteCount`
- One-Pager:Teamarbeit / Aufteilung darstellen

Prüfungsleistung

- Jupyter Markdown Notebook (.ipynb)
  - Kombination aus Projektbericht, Quellcode und Kommentaren
  - Fließtext: 2-3 Seiten pP
  - Referenzen, Tabellen / Abbildungen
  - Quellcode-Zellen mit Output
  - `https://www.kaggle.com/masumrumi/a-statistical-analysis-ml-workflow-of-titanic`
  - `https://www.kaggle.com/c/titanic/code?competitionId=3136&sortBy=voteCount`
- One-Pager:Teamarbeit / Aufteilung darstellen
- Abschlusspräsentation

Prüfungsleistung

- Jupyter Markdown Notebook (.ipynb)
  - Kombination aus Projektbericht, Quellcode und Kommentaren
  - Fließtext: 2-3 Seiten pP
  - Referenzen, Tabellen / Abbildungen
  - Quellcode-Zellen mit Output
  - `https://www.kaggle.com/masumrumi/a-statistical-analysis-ml-workflow-of-titanic`
  - `https://www.kaggle.com/c/titanic/code?competitionId=3136&sortBy=voteCount`
- One-Pager:Teamarbeit / Aufteilung darstellen
- Abschlusspräsentation
  - 15 - 25 Minuten (nach Gruppengröße)

Prüfungsleistung

- Jupyter Markdown Notebook (.ipynb)
  - Kombination aus Projektbericht, Quellcode und Kommentaren
  - Fließtext: 2-3 Seiten pP
  - Referenzen, Tabellen / Abbildungen
  - Quellcode-Zellen mit Output
  - `https://www.kaggle.com/masumrumi/a-statistical-analysis-ml-workflow-of-titanic`
  - `https://www.kaggle.com/c/titanic/code?competitionId=3136&sortBy=voteCount`
- One-Pager:Teamarbeit / Aufteilung darstellen
- Abschlusspräsentation
  - 15 - 25 Minuten (nach Gruppengröße)
  - 5 Minuten Fragen und Diskussion

Bewertungskriterien

- Wurde die Data-Science Pipeline sauber angewandt und dargestellt?

Bewertungskriterien

- Wurde die Data-Science Pipeline sauber angewandt und dargestellt?
- Konzept und Code

Bewertungskriterien

- Wurde die Data-Science Pipeline sauber angewandt und dargestellt?
- Konzept und Code
- Performance des ML Modells

Bewertungskriterien

- Wurde die Data-Science Pipeline sauber angewandt und dargestellt?
- Konzept und Code
- Performance des ML Modells
- Bericht

Bewertungskriterien

- Wurde die Data-Science Pipeline sauber angewandt und dargestellt?
- Konzept und Code
- Performance des ML Modells
- Bericht
  - Relevante Grundlagen erklären

Bewertungskriterien

- Wurde die Data-Science Pipeline sauber angewandt und dargestellt?
- Konzept und Code
- Performance des ML Modells
- Bericht
  - Relevante Grundlagen erklären
  - Anforderungen an wiss. Arbeiten bleiben auch beim Notebook Format

Bewertungskriterien

- Wurde die Data-Science Pipeline sauber angewandt und dargestellt?
- Konzept und Code
- Performance des ML Modells
- Bericht
  - Relevante Grundlagen erklären
  - Anforderungen an wiss. Arbeiten bleiben auch beim Notebook Format
  - Struktur, Literaturreferenzen / Quellenangaben

Bewertungskriterien

- Wurde die Data-Science Pipeline sauber angewandt und dargestellt?
- Konzept und Code
- Performance des ML Modells
- Bericht
  - Relevante Grundlagen erklären
  - Anforderungen an wiss. Arbeiten bleiben auch beim Notebook Format
  - Struktur, Literaturreferenzen / Quellenangaben
- Präsentation (Stil, Verständlichkeit, Motivation)

- Gruppenbildung: 2 bis 4 Studierende pro Gruppe

- Gruppenbildung: 2 bis 4 Studierende pro Gruppe
- Themenwahl und Themenausarbeitung

- Gruppenbildung: 2 bis 4 Studierende pro Gruppe
- Themenwahl und Themenausarbeitung
  - Vorstellung Themen-Ideen

- Gruppenbildung: 2 bis 4 Studierende pro Gruppe
- Themenwahl und Themenausarbeitung
  - Vorstellung Themen-Ideen
  - Mehrfachbelegung möglich

- Gruppenbildung: 2 bis 4 Studierende pro Gruppe
- Themenwahl und Themenausarbeitung
  - Vorstellung Themen-Ideen
  - Mehrfachbelegung möglich
  - Eigene Themen nach Absprache möglich

- Gruppenbildung: 2 bis 4 Studierende pro Gruppe
- Themenwahl und Themenausarbeitung
  - Vorstellung Themen-Ideen
  - Mehrfachbelegung möglich
  - Eigene Themen nach Absprache möglich
- Gruppe bilde, Thema wählen (automatische Verteilung)

Veranstaltungstermine

- heute: Themenvorstellung

Veranstaltungstermine

- heute: Themenvorstellung
- nächstes Mal: Konzeptvorstellung jeder Gruppe (Kurzpräsentation 5-10 Minuten), Feedback

Veranstaltungstermine

- heute: Themenvorstellung
- nächstes Mal: Konzeptvorstellung jeder Gruppe (Kurzpräsentation 5-10 Minuten), Feedback
- Hackathon

Veranstaltungstermine

- heute: Themenvorstellung
- nächstes Mal: Konzeptvorstellung jeder Gruppe (Kurzpräsentation 5-10 Minuten), Feedback
- Hackathon
  - 8:30 Uhr Tour de table: Präsentation des Zwischenstands / Nächste Schritte

Veranstaltungstermine

- heute: Themenvorstellung
- nächstes Mal: Konzeptvorstellung jeder Gruppe (Kurzpräsentation 5-10 Minuten), Feedback
- Hackathon
  - 8:30 Uhr Tour de table: Präsentation des Zwischenstands / Nächste Schritte
  - Weiterarbeiten am Projekt

Veranstaltungstermine

- heute: Themenvorstellung
- nächstes Mal: Konzeptvorstellung jeder Gruppe (Kurzpräsentation 5-10 Minuten), Feedback
- Hackathon
  - 8:30 Uhr Tour de table: Präsentation des Zwischenstands / Nächste Schritte
  - Weiterarbeiten am Projekt
  - Einzeltermine buchen (Moodle Forum)

Veranstaltungstermine

- heute: Themenvorstellung
- nächstes Mal: Konzeptvorstellung jeder Gruppe (Kurzpräsentation 5-10 Minuten), Feedback
- Hackathon
  - 8:30 Uhr Tour de table: Präsentation des Zwischenstands / Nächste Schritte
  - Weiterarbeiten am Projekt
  - Einzeltermine buchen (Moodle Forum)
- Weitere Besprechungstermine auf Anfrage

Veranstaltungstermine

- heute: Themenvorstellung
- nächstes Mal: Konzeptvorstellung jeder Gruppe (Kurzpräsentation 5-10 Minuten), Feedback
- Hackathon
  - 8:30 Uhr Tour de table: Präsentation des Zwischenstands / Nächste Schritte
  - Weiterarbeiten am Projekt
  - Einzeltermine buchen (Moodle Forum)
- Weitere Besprechungstermine auf Anfrage
- Abgabe / Abschlusspräsentationen im Juli

Projekt-Anforderungen

- Kontext erklären

Projekt-Anforderungen

- Kontext erklären
- Explorative Datenanalyse

Projekt-Anforderungen

- Kontext erklären
- Explorative Datenanalyse
- Auswahl und Vergleich verschiedener DS/ML Verfahren

Projekt-Anforderungen

- Kontext erklären
- Explorative Datenanalyse
- Auswahl und Vergleich verschiedener DS/ML Verfahren
- Grundlagen erklären und referenzieren

Projekt-Anforderungen

- Kontext erklären
- Explorative Datenanalyse
- Auswahl und Vergleich verschiedener DS/ML Verfahren
- Grundlagen erklären und referenzieren
- Konzept und Ergebnisdarstellung

Projekt-Anforderungen

- Kontext erklären
- Explorative Datenanalyse
- Auswahl und Vergleich verschiedener DS/ML Verfahren
- Grundlagen erklären und referenzieren
- Konzept und Ergebnisdarstellung
- Präsentation: Konzept und Ergebnisse

Moodle-Raum

- Moodle Kursraum:

  `https://moodle.dhbw-mannheim.de/course/view.php?id=11560`

- Schlüssel: **124**

Themen / Datensätze

Wine Quality

- https://www.kaggle.com/yasserh/wine-quality-dataset
- Qualität von Wein vorhersagen

Non-performing loans

- https://www.kaggle.com/yasserh/loan-default-dataset
- Vorhersage, wie wahrscheinlich der Ausfall eines Kredits ist

Stellar Classification

- https://www.kaggle.com/fedesoriano/stellar-classification-dataset-sdss17
- Vorhersage anhand von Spektralmessungen, ob es sich um eine Galaxie, Stern oder Quasar handelt

Song Popularity

- https://www.kaggle.com/yasserh/song-popularity-dataset
- Vorhersage wie beliebt / erfolgreich ein Musiktitel ist anhand vordefinierter Audiofeatures
- Keine direkte Audioanalyse notwendig

Medizinische Vorhersage

- https://www.kaggle.com/fedesoriano/stroke-prediction-dataset
- https://www.kaggle.com/datasets/kamilpytlak/personal-key-indicators-of-heart-disease/data
- Bestimmung des Schlaganfall-Risiko / Herzkrankheiten anhand sozio-ökonomischer und gesundheitlicher Merkmale

Fußballergebnisse

- https://www.kaggle.com/datasets/martj42/international-football-results-from-1872-to-2017
- Vorhersage von Toren / Ergebnis (WLD) von Länderspielen

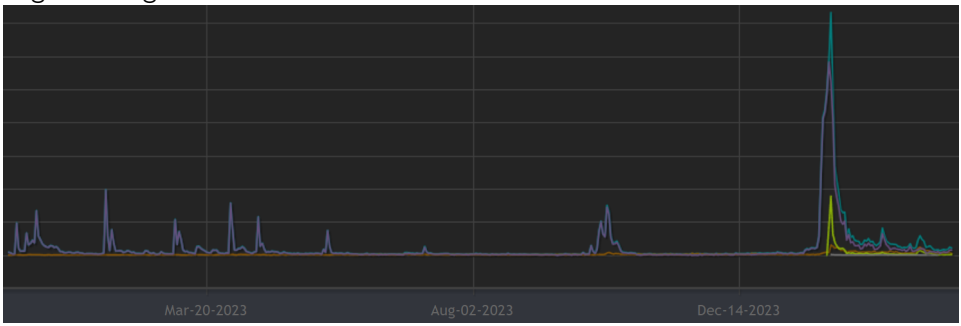
Wasserqualität

- https://www.kaggle.com/datasets/adityakadiwal/water-potability
- Wasserqualität (potable ja/nein) vorhersagen

Bildklassifikation

- Image Classification `https://www.kaggle.com/datasets/hasibalmuzdadid/shoe-vs-sandal-vs-boot-dataset-15k-images`

Steam Sales

- Views / Wishlists / Sales von Steam Spiel analysieren / vorhersagen / Treiber finden
- `https://store.steampowered.com/app/2216770/JOY_OF_PROGRAMMING__Software_Engineering_Simulator/`
- 

- Daten werden gestellt - vertraulich behandeln

Andere Datensätze

- Andere Datensätze von Kaggle oder weiteren Quellen
- Eigene Datensätze
- Bitte kurz absprechen

Gruppenbildung / Themenfindung

2-4 Studierende, max. 10 Gruppen

Gruppenbildung / Themenfindung

2-4 Studierende, max. 10 Gruppen

- Bei Fragen bitte auf mich zukommen

Gruppenbildung / Themenfindung

2-4 Studierende, max. 10 Gruppen

- Bei Fragen bitte auf mich zukommen
- Eigene Themenideen bitte abklären

Gruppenbildung / Themenfindung

2-4 Studierende, max. 10 Gruppen

- Bei Fragen bitte auf mich zukommen
- Eigene Themenideen bitte abklären
- Für Gruppe Thema / Themen wählen:

`https://moodle.dhbw-mannheim.de/mod/ratingallocate/view.php?id=331340`