

Eindopdracht Data Mining 24/25

In deze opdracht werk je in dezelfde groep als bij het vak Data Visualisatie. Je voert een data-analyse uit op een bestaande dataset. Deze dataset is dezelfde als die je gebruikt hebt voor Data Visualisatie.

Beantwoord een onderzoeksvraag of doe een voorspelling op basis van de data. Je past technieken toe uit de lessen Data Mining: data opschonen, analyse, visualisatie, en machine learning.

Wat lever je in?

1. Dataset

De dataset waarmee je hebt gewerkt.

2. Onderzoeksvraag

Een duidelijke onderzoeksvraag of voorspelling, afgestemd op je dataset.

3. Rapport (PDF)

Een concreet rapport waarin je het proces en de analyse beschrijft.

Het rapport bevat:

- Omschrijving van je onderzoeksvraag of doelstelling voorspelling
- Visualisaties van de rauwe data (mag uit Data Visualisatie komen)
- Beschrijving van de data preparatie
- Beschrijving van de data analyse
- Beschrijving en toelichting resultaten
- Toelichting op gebruikte machine learning technieken
- Toelichting gebruikte features
- Conclusie met antwoord op je onderzoeksvraag

4. Code (Jupyter Notebook / .py bestand)

Het code-bestand waarin de data analyse / ML is uitgevoerd.

Beoordeling

Het beoordelingsmodel is te vinden op Blackboard.

Je wordt beoordeeld op:

- De kwaliteit van je analyse en aanpak
 - Juiste toepassing van technieken uit de lessen
 - Overzichtelijke visualisaties
 - Duidelijke rapportage en verantwoording
 - Netheid en reproduceerbaarheid van je code
 - **Motivatie voor keuzes (preparatie/analyse/features)**
-

Inleveren

- Rapport in PDF
 - Dataset in .csv
 - Code in Jupyter Notebook of .py
 - Alles inleveren via Blackboard
-

Opmerking

Bij de huiswerkopdrachten en de eindopdracht is vooral de “waarom-vraag” belangrijk. We verlangen niet alleen een werkende code, maar vooral een onderbouwing van de keuzes:

Waarom heb je x-rijen verwijderd tijdens het opschonen?

Waarom heb je gekozen voor een lineaire regressie?

Waarom heb je deze variabelen gekozen?

Waarom is dit model nou goed of slecht?