

# Tentamen Collectieve Intelligentie

maandag 30 mei 2022 9:00-11:00

Vul hier je naam en studentnummer in vóór je begint:  .....	/ 43 p
---	--------

1. Je mag de vragen in Engels of Nederlands beantwoorden.
2. Dit is een "gesloten boek"-tentamen. Je mag voor het invullen je pen of potlood gebruiken, maar verder niets. Schrijf duidelijk en niet te groot.
3. Leg je studentenkaart (of ander ID met foto) klaar op je tafel. We komen langs om te kijken of je hierboven je naam hebt ingevuld en of deze klopt met je ID.
4. Vul je antwoorden in op dit tentamenblad zelf.
5. Laat het weten als je extra papier of kladpapier nodig hebt.
6. Als je vragen hebt over hoe we iets bedoelen, dan kunnen we dat waarschijnlijk niet beantwoorden zonder een deel van het antwoord weg te geven (maar voel je vrij om het te proberen!).
7. Tip: Het tentamen bevat vragen over veel verschillende onderwerpen. Kijk eerste even het hele tentamen door en begin met de onderwerpen die je het beste beheerst.

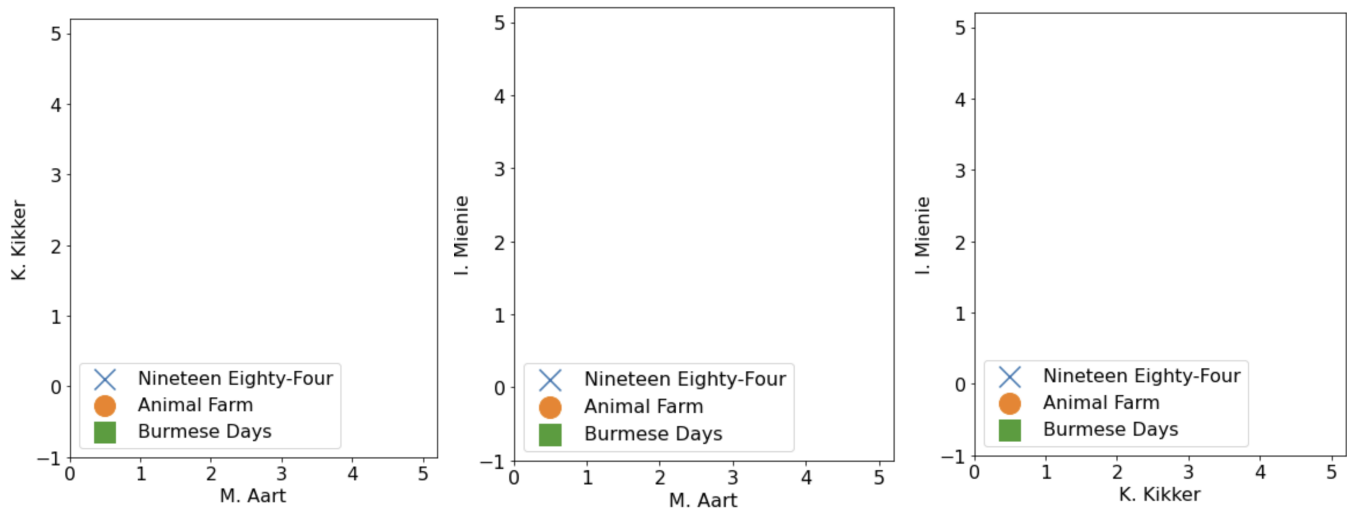
# Orwell

Je bent een *collaborative filtering*-algoritme aan het ontwikkelen voor het aanbevelen van boeken. Je hebt de volgende data:

	user	book	rating
0	M. Aart	Nineteen Eighty-Four	4
1	M. Aart	Animal Farm	4
2	M. Aart	Burmese Days	3
3	K. Kikker	Nineteen Eighty-Four	3
4	K. Kikker	Animal Farm	2
5	K. Kikker	Burmese Days	1
6	I. Mienie	Nineteen Eighty-Four	4
7	I. Mienie	Animal Farm	5
8	I. Mienie	Burmese Days	1

2p **Question 1.**

Vul aan de hand van deze data in de onderstaande figuren in:



4 p **Question 2.**

We hebben het vermoeden dat Nineteen Eighty-Four meer lijkt op Animal Farm dan op Burmese Days. Welke similarity-maat reflecteert dit het best? *manhattan similarity*, *euclidean similarity* of *cosine similarity*? Leg uit waarom.

Antwoord

---

---

---

---

---

---

---

---

---

---

---

4 p **Question 3.**

Voor deze data kan het beter zijn om de **adjusted** *cosine similarity* te gebruiken in plaats van de *cosine similarity*. Waarom is dat?

Antwoord

---

---

---

---

---

---

---

---

---

---

---

## Brownies en kip

3 p **Question 4.**

Je hebt een dataset met gerechten en ingrediënten. Je wilt bepalen welke van deze gerechten het meest op elkaar lijken (de hoogste *similarity* hebben). De *entries* voor de eerste drie gerechten uit je data zien er zo uit:

	ingredients
brownies	flour, dark chocolat, eggs, sugar
banana bread	bananas, flour, eggs, sugar, baking powder, oil
fried chicken	flour, egg, drumsticks, bread crumbs

Deze data kunnen we niet direct gebruiken om *similarities* te berekenen. Zet deze data om in een *one-hot encoding* zodat dat we daarna makkelijk *similarities* kunnen berekenen. **Teken een volledig uitgewerkte tabel!**

Antwoord

This image shows a blank sheet of white paper with horizontal ruling lines. The lines are evenly spaced and run across the width of the page. There are no margins, text, or other markings on the paper.

3 p **Question 5.**

Bereken de *jaccard similarity* tussen *brownies* en *fried chicken*. Ter herinnering, de jaccard similarity is gedefinieerd als:

$$J = \frac{M_{11}}{M_{01} + M_{10} + M_{11}}$$

Antwoord

---

---

---

---

---

---

---

---

---

---

## Zinnen

### 5 p Question 6.

We hebben de volgende drie zinnen:

---

Zin 1: Wij gaan weg.

Zin 2: Wij gaan naar huis.

---

Zin 3: Jan gaat naar Parijs.

---

Eén van de onderstaande tabellen geeft de juiste tf-idf vectorisatie van de bovenstaande zinnen. Welke? Leg uit waarom. Je hoeft dit niet uit te rekenen. Je kan het beredeneren.

Tabel A:

	parijs	wij	weg	gaat	naar	jan	gaan	huis
Zin 1:	0	1	1	0	0	0	1	0
Zin 2:	0	1	0	0	1	0	1	1
Zin 3:	1	0	0	1	1	1	0	0

Tabel B:

	parijs	wij	weg	gaat	naar	jan	gaan	huis
Zin 1:	0.0	0.41	1.1	0.0	0.00	0.0	0.41	0.0
Zin 2:	0.0	0.41	0.0	0.0	0.41	0.0	0.41	1.1
Zin 3:	1.1	0.00	0.0	1.1	0.41	1.1	0.00	0.0

Tabel C:

	parijs	wij	weg	gaat	naar	jan	gaan	huis
Zin 1:	0.00	1.1	0.41	0.00	0.0	0.00	1.1	0.00
Zin 2:	0.00	1.1	0.00	0.00	1.1	0.00	1.1	0.41
Zin 3:	0.41	0.0	0.00	0.41	1.1	0.41	0.0	0.00

This image shows a blank sheet of white paper with horizontal ruling lines. The lines are evenly spaced and run across the width of the page. There are no margins, text, or other markings on the paper.

# Matrices

## 2 p Question 7.

Bekijk de onderstaande matrices:

$$A = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}, B = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}$$

We hebben de volgende vergelijking:

$$A \cdot B = C$$

Wat is de matrix  $C$ ?

Antwoord

---

---

---

---

---

---

---

---

---

---

---

## 6 p Question 8.

Bekijk de onderstaande matrices en matrixvergelijking:

$$X = \begin{bmatrix} 1 & 1 \\ 2 & 3 \\ 3 & 1 \end{bmatrix}, Y = \begin{bmatrix} 4 \\ 10 \\ 12 \end{bmatrix}$$

$$X \cdot B = \hat{Y}$$

We willen een matrix  $B$  zo vinden dat de *mean squared error* (MSE) tussen  $Y$  en  $\hat{Y}$  minimaal is.

- Beschrijf de stappen van het *gradient descent*-algoritme waarmee we dit voor elkaar kunnen krijgen.
- Waarom zal de MSE nooit precies nul worden?



[illegible]

## Gastcolleges

De onderstaande vragen zijn gebaseerd op de gastcolleges.

### Nicolas Mattis

2 p **Question 9.**

In het college van Nicolas Mattis noemt hij het belang van een gevarieerde nieuwsconsumptie voor een democratie. Hij noemt een aantal manieren om diversiteit te meten. Eén daarvan is *Alternative voices*. Leg in één of twee zinnen uit wat daarmee wordt bedoeld.

Antwoord

---

---

---

---

---

---

### Min Fang

5 p **Question 10.**

Het is belangrijk om een *recommender system* zowel offline als online te evalueren. Waarom zijn deze beide methodes nodig? Leg uit hoe deze methodes elkaar aanvullen aan de hand van hun voor en nadelen.

Antwoord

---

---

---

---

---

---

---

## David Graus

### 4 p Question 11.

In het gastcollege van David Graus noemt hij het belang van de omgang met technologie als ondersteuning van het menselijke proces, niet de basis. Om bias te voorkomen binnen het *recommender system* van Randstad worden bepaalde features zoals geslacht, nationaliteit en leeftijd niet meegenomen. Op welke manier zouden deze features juist wel kunnen worden gebruikt om bias tegen te gaan?

Antwoord

---

---

---

---

---

---

---

---

---

---

