Big data



Docentenhandleiding

Inhoudsopgave

[**1. Inleiding**](#_hqqh8hv28gaj) **4**

[**2. De module in een oogopslag**](#_74xo8x721rb2) **5**

[**3. Leerdoelen per hoofdstuk**](#_51ck5c2m7rx6) **7**

[3.1 Leerdoelen van hoofdstuk 1](#_6d0a88oascyg) 7

[3.2 Leerdoelen van hoofdstuk 2](#_nuy2t5my9e26) 7

[3.3 Leerdoelen van hoofdstuk 3](#_ws2wsmbgkx7o) 8

[**4. Contexten**](#_ig7s0hmpzj7b) **9**

[**5. Concepten per hoofdstuk**](#_66aselno0xv1) **10**

[5.1 Concepten van hoofdstuk 1](#_y0ukiirdomz0) 10

[5.2 Concepten van hoofdstuk 2](#_r1fn0kr35xts) 10

[5.3 Concepten van hoofdstuk 3](#_dpj9fwyibopo) 11

[**6. Vaardigheden**](#_2wpjz22qoaao) **12**

[**7. Vakkennis**](#_6jj77jfjx6x6) **13**

[7.1 Wiskunde](#_yg0xzgh0y4qo) 13

[7.2 Natuurkunde](#_7rwy96uwrpv7) 13

[7.3 Informatica](#_e39qxnp84igr) 13

[**8. Leerplan**](#_yv4wwekqrmki) **14**

[8.1 Beschrijving per hoofdstuk](#_po8fl2chgxg9) 14

[8.1.1 Beschrijving van hoofdstuk 1](#_3b26k49eblvl) 14

[8.1.2 Beschrijving van hoofdstuk 2](#_2dsthq3kwzd1) 15

[8.1.3 Beschrijving van hoofdstuk 3](#_4l8itq9outhy) 17

[8.2 Differentiatie per hoofdstuk](#_k22qxy0kp5v) 17

[8.1.1 Differentiatie in hoofdstuk 1](#_jk7sye122374) 18

[8.1.2 Differentiatie in hoofdstuk 2](#_nou8rx6xa93h) 18

[8.1.3 Differentiatie in hoofdstuk 3](#_9bqxdf7jbsa1) 18

[**9. Toelichting bij leerlingopdrachten**](#_owzvav10e9gp) **19**

[9.1 Hoofdstuk 1](#_gzfls04enw2g) 19

[9.2 Hoofdstuk 2](#_rkq1oac3jr5b) 19

[9.3 Hoofdstuk 3](#_r3l4ejg7crxp) 19

[**10. Bronnen bij leermateriaal**](#_nfwdwgbwfr2p) **20**

[10.1 Hoofdstuk 1](#_y13zus2se3md) 20

[10.2 Hoofdstuk 2](#_ttinv428l87d) 23

[10.3 Hoofdstuk 3](#_glws54p3b56f) 26

[**11. Toetsing**](#_ysvi0d4348du) **35**

[11.1 Hoofdstuk 1](#_n9e3v6she8ml) 35

[11.2 Hoofdstuk 2](#_n1900ihvyh2k) 35

[11.3 Hoofdstuk 3](#_nffuc2uk7uac) 35

[**12. Suggesties en extra opdrachten**](#_1ektpmrpnfwh) **36**

[12.1 Hoofdstuk 1](#_8qjrriime3kw) 36

[12.2 Hoofdstuk 2](#_uz055tpdnwb4) 36

[12.3 Hoofdstuk 3](#_f8pdgfke047t) 36

[**13. Achtergrondinformatie**](#_6gjyx75c36qq) **38**

[13.1 Contexten en concepten](#_mxs0tm3q2bzr) 38

[13.2. Studieinfo](#_sai2rbxzj1ko) 38

[**Bijlage A - SO**](#_lgkhm2yqcci3) **40**

[**Bijlage B - Toets**](#_w8tvhfjx751z) **41**

[**Bijlage C - Antwoorden practicum**](#_w6xqt0jrq5z8) **46**

[**Bijlage D - Uitwerking antwoorden 3.2**](#_b1w0v8qe4m7h) **59**

# 1. Inleiding

Dit document bevat de docentenhandleiding van de NLT-module ‘Big Data’. De module is ontwikkeld door drie studenten van de lerarenopleiding Master Science Education and Communication aan de TU Delft. Hierbij is gebruik gemaakt van de inhoudelijke expertise van Martin Bruggink. De module is ontwikkeld onder supervisie van vakcoach Wim Sonneveld.

De module is bedoeld voor VWO 5 leerlingen. We hebben gekozen voor een bedrijfskunde context, namelijk het behalen van winst door het analyseren van data. Deze aanpak is vooral relevant voor de leerlingen die een E&M en N&T hebben gekozen. Om ook de C&M en N&G leerlingen te activeren is gekozen voor de volgende aanpak. Tijdens de lessen over toepassingen betreffen deze toepassingen de publieke sector. In de publieke sector wordt namelijk Big Data zelden toegepast. terwijl er wel veel winst mee valt te behalen, ook in de publieke sector.

Tijdens het ontwikkelen van de module is de nadruk geweest op het gebruik van zoveel mogelijk verschillende werkvormen. Hierdoor wordt de leerstof voor leerlingen relevanter. Ook zijn er zo meer mogelijkheden om kennis, vaardigheden en leerstijlen van de leerlingen te activeren. Deze module bevat ook een computerpracticum zodat leerlingen de theorie in de praktijk kunnen beoefenen.

Larbi et Yaakoubi

Julian de Groot

Bert Lobbezoo

Doelgroep: VWO 5

NLT-domeinen: Domein A, B, C en E.

# 2. De module in een oogopslag

|  |  |  |
| --- | --- | --- |
| Titel | Big Data |  |
| Code | nltx-qxxx |  |
| Status | testversie |  |
| Havo/vwo | vwo |  |
| Beoogd leerjaar | 5 vwo |  |
| Omschrijving | Een reis vanuit wat data precies is naar wat data ‘Big’ maakt met een scala aan toepassingen en wiskundige technieken. |  |
| Context | Door recente technologische ontwikkelingen is het mogelijk geworden veel informatie op te slaan. Bedrijven zijn geïnteresseerd in het hebben van veel informatie omdat dit geld kan opleveren. Doordat de hoeveelheid informatie harder stijgt ten opzichte van de technologische mogelijkheden voor het opslaan van deze data zijn nieuwe technieken nodig om de data te doorzoeken. | [hfst. 4](#_ig7s0hmpzj7b) |
| Concepten | Definitie   * Bits * Representatie * Grootte van data   Toepassingen   * Big Data in publieke & private sector * Big Data analytics * Privacy & ethiek   Technieken   * Associatietechnieken * Clustertechnieken * Algoritmiek/Scripting | [hfst. 5](#_66aselno0xv1) |
| Vaardigheden | Bijna alle vaardigheden uit domein A van het examenprogramma  komen in deze module aan bod. Verder komen de aspecten uit domein E aan bod:  Technologische ontwikkelingen kunnen benoemen en processen hierbij kunnen beschrijven/analyseren en toepassen op interdisciplinaire (ontwerp)vraagstukken aan de hand van wiskundige concepten.  Verder komt ook theorie uit subdomein C2 aan bod: (Duurzaamheid)  het duurzaam gebruik van grondstoffen, energie en ruimte. | [hfst. 6](#_2wpjz22qoaao) |
| Voorkennis TF | Voorkennis is nodig van het vak Wiskunde | [hfst. 7](#_6jj77jfjx6x6) |
| Studielast | 20 slu | [hfst. 8](#_yv4wwekqrmki) |
| Structuur | Lineair | [hfst. 8](#_po8fl2chgxg9) |
| Leerling-activiteiten | Spellen  Debat  Pitch  Flipping the classroom  MOOC cursus  Practicum | [hfst. 8](#_yv4wwekqrmki) / [hfst. 10](#_nfwdwgbwfr2p) |
| Faciliteiten | Voor deze module zijn de volgende faciliteiten nodig:   * Twee practicumlessen in computerlokaal * Installatie Geogebra op computer in het computerlokaal | [hfst. 9](#_owzvav10e9gp) |
| Toetsvormen en weging | 30% schriftelijke overhoring  10% professionaliteitscijfer  20% beoordeling score MOOC  30% beoordeling schriftelijke toets  10% beoordeling practicum | [hfst. 11](#_ysvi0d4348du) |

# 3. Leerdoelen per hoofdstuk

Dit hoofdstuk bevat een overzicht van de leerdoelen die aan het einde van elk hoofdstuk van de module behaald moeten zijn.

## 3.1 Leerdoelen van hoofdstuk 1

Na dit hoofdstuk

* weet je welke waarden een bit kan hebben
* kun je een tekst kunt omzetten in bits
* kun je muziek omzetten in bits
* kun je een afbeelding omzetten in bits
* kun je een video omzetten in bits
* kun je berekenen hoeveel opslag je nodig hebt voor een bepaalde hoeveelheid informatie
* kun je berekenen hoe lang het versturen dan wel ontvangen duurt van een bepaalde hoeveelheid informatie
* kun je berekenen hoe lang het duurt om een bepaalde berekening uit te voeren op de data
* kun je bepalen of data uit verschillende bronnen consistent is
* kun je bepalen welke bron minder betrouwbaar is
* kun je bepalen in hoeverre data van verschillende bronnen zijn te combineren

## 3.2 Leerdoelen van hoofdstuk 2

Na dit hoofdstuk kun je

* minstens drie voorbeelden geven van manieren waarop Big Data in de praktijk wordt toegepast
* opnoemen binnen welke gebieden van de publieke en private sector Big Data een grote invloed heeft of zal gaan hebben
* de rol van het Internet of Things (IoT) in combinatie met Big Data benoemen
* de stappen uit de feedback cyclus benoemen binnen de context van Big Data
* voorbeelden geven op wat voor manier Big Data bijdraagt aan de verduurzaming van onze maatschappij
* de randvoorwaarden benoemen als het gaat om de inzet van persoonsgegevens bij Big Data
* de ethische kanten van Big Data benoemen
* de primaire noodzaak om te veranderen benoemen voor organisaties in de publieke- en private sector.
* vijf organisatieprincipes benoemen die een steeds grotere rol zullen gaan spelen in de toekomst.
* benoemen welke vier rollen nodig zijn voor het model van de voorspellende analyse
* voorbeelden geven van beroepen en opleidingen die horen bij Big Data
* via online cursusmateriaal de basisconcepten van een data scientist jezelf eigen maken

## 3.3 Leerdoelen van hoofdstuk 3

Na dit hoofdstuk

* kun je in eigen woorden het doel en de taak van associatieanalyse bij datamining toelichten.
* kun je de begrippen verzamelingenleer, associatieanalyse, associatieregel uitleggen.
* weet je de toepassingen van associatieanalyse, bijvoorbeeld Market basket analyse
* ben je in staat praktische voorbeelden van associaties opnoemen.
* kun je Apriori algoritme gebruiken bij het vinden van de associatieregel in een eenvoudig probleem.
* gebruiken bij het bepalen van Doorsnede en de vereniging van twee of meer verzamelingen.
* gebruiken bij het berekenen van Support van een deelverzameling
* gebruiken bij het berekenen van Confidence van een associatieregel van een deelverzameling
* als rekenmiddel voor de uitvoering van de benodigde berekeningen bij elke iteratie van Apriori algoritme.
* weet je het doel en de taak van clusteranalyse bij datamining.
* kun je een paar toepassingen van clustering noemen;
* kun je de begrippen datapunt, object, cluster, centrum, centroide, Euclidische afstand, Manhattan afstand; centrum uitleggen .
* kun je de afstand tussen zowel twee objecten als twee clusters met behulp van afstandsmaten zoals Euclidische afstand en Manhattan afstand berekenen.
* kun je de centrum of centroide tussen zowel twee objecten als twee clusters met behulp van van afstandsmaten zoals Euclidische afstand en Manhattan afstand berekenen.
* ben je in staat de clusteringenmethoden, zoals de -means methode op passende problemen toepassen.

# 4. Contexten

Grote bedrijven zoals Facebook en Google verzamelen enorme hoeveelheden data. Met deze data kan veel geld verdient worden. Hierdoor is er een nieuw vakgebied ontstaan, namelijk het onderzoeken van data om er geld mee te kunnen verdienen of om achter belangrijke informatie te komen. Deze hoeveelheid data is zo groot geworden dat er nieuwe technieken ontwikkeld moeten worden om de data op een gebruiksvriendelijke manier te kunnen onderzoeken. Om deze reden is de term ‘Big Data’ in het leven geroepen.

Veel bedrijven zijn actief bezig met het vergaren van data: data mining. Sensoren en allerlei netwerkapparatuur gebruikt men om de gegevens snel binnen te halen en apparatuur met elkaar te laten communiceren. Wij noemen dit tegenwoordig ook wel het Internet of Things (IoT). De analyse van alle vergaarde data, ook wel Big Data analytics genoemd, is de expertise van data scientists die samen met business managers, analisten en IT system managers, een bedrijfsvoering gericht op Big Data, op gang houden. Op basis van uitgebreide analyse is het mogelijk om zelfs voorspellingen te doen en bedrijven zodoende van constante feedback te voorzien in de bedrijfsvoering. Het verzamelen en verwerken van al deze gegevens brengt ook vragen met zich mee omtrent privacy en ethiek. Kennis van Big Data is nodig om de maatschappelijke aanverwante vragen te kunnen beantwoorden.

In onze samenleving zien we dat men binnen de publieke en private sector anders om gaat met Big Data en de vooruitgang die hier uit voorkomt. Zo valt er in alle gevallen veel winst te behalen en bewuster met grondstoffen om te gaan. Duurzaamheid is ook ‘Big’ bij Big Data.

In deze module is voor twee vormen van datamining gekozen: associatieve technieken en clustertechnieken.

De associatie analyse zoekt naar verbanden tussen verschillende groepsgegevens. Deze technieken maken het mogelijk om na te gaan wat winkelende klanten zoal in hun boodschappenmandje stoppen en of hier bepaalde patronen of combinaties in te vinden zijn. Door het toepassen van deze technieken is ondermeer ontdekt dat sommige klanten (vooral jonge volwassenen) vrijdagmiddag vlak voor sluitingstijd de supermarkt binnenlopen om luiers, chips en bier te kopen. Omdat deze inkoopcombinatie vaak voorkomt, spelen sommige supermarkten hier slim op in door daar selectief hun kortingsacties of assortiment op af te stemmen.

Met clusteranalyse probeert men daarbij categorieën te vinden. Hierbij gaat het om het vinden van eigenschappen van gegevens om nieuwe groepen met tot nu toe onbekende gemeenschappelijke kenmerken te ontdekken. Wanneer bijvoorbeeld informatie over mensen die een bepaalde speelfilm hebben gezien onderzocht wordt, kan clusteranalyse uitwijzen dat, gezien de leeftijd van de kijkers, de populatie in twee leeftijdsgroepen ingedeeld kan worden zoals een groep 4 tot 10-jarigen en een groep 25 tot 40-jarigen.

# 5. Concepten per hoofdstuk

Hieronder worden alle concepten opgesomd die behandeld worden in de verschillende hoofdstukken van de module Big Data.

## 5.1 Concepten van hoofdstuk 1

De concepten die aan de orde komen in hoofdstuk 1 zijn:

* Bit
* Representatie
* Pixels
* Subpixels
* Kleuren mengen
* Grootte van data
* Verzendtijd van bits
* Bewerkingssnelheid van data
* Consistentheid van bronnen
* Betrouwbaarheid van bronnen
* Combineren van bronnen

## 5.2 Concepten van hoofdstuk 2

De concepten die aan de orde komen in hoofdstuk 2 zijn:

* Publieke sector
* Private sector
* Data mining
* IoT (Internet of Things)
* Big Data analytics
* Learning analytics
* De feedbackcyclus
* Duurzaamheid via Big Data
* Big Social Data
* Privacy en ethiek
* Noodzaak om te veranderen (toekomst voor organisaties)
* Organisatieprincipes
* The predictive analytics cycle
* Business manager
* Business analist
* IT system manager (Big Data manager)
* Data scientist (data analist)
* Big Data opleidingen
* Big Data online cursussen (zelfstudie)

## 5.3 Concepten van hoofdstuk 3

De concepten die aan de orde komen in hoofdstuk 3 zijn:

* Associatie-analyse
* Verzamelingenleer
* Verzameling
* Deelverzameling
* Doorsnede
* Vereniging
* Algoritme
* Iteratie
* Dataset
* Itemsets
* Market basket analyse
* Associatieregel
* Apriori algoritme
* Support
* Confidence
* Frequent
* Clusteranalyse
* Datapunt
* Object
* Cluster
* Centrum
* Afstandsmaten
* Euclidische afstand
* Manhattan afstand
* K-means algoritme

# 6. Vaardigheden

Dit hoofdstuk bevat een overzicht van de vaardigheden die aan bod komen in deze module. Voor alle genoemde vaardigheden, met uitzondering van degene onder domein E, geldt dat ze gedeeltelijk bekend worden verondersteld, maar in deze module verder worden aangeleerd en geoefend.

**Domein A Vaardigheden**

- Algemene vaardigheden (profieloverstijgend niveau)

Subdomein A01: Informatievaardigheden gebruiken

Subdomein A02: Communiceren

Subdomein A03: Reflecteren op leren

Subdomein A04: Studie en beroep

- Natuurwetenschappelijke, wiskundige en technische vaardigheden (bètaprofielniveau)

Subdomein A05: Onderzoeken

Subdomein A06: Ontwerpen

Subdomein A07: Modelvorming

Subdomein A08: Natuurwetenschappelijk instrumentarium

Subdomein A09: Waarderen en oordelen

- NLT-specifieke vaardigheden

Subdomein A10: Interdisciplinaire vraagstukken in studie- en beroepspraktijk

Subdomein A11: Redeneren

Subdomein A12: Rekenkundige en wiskundige vaardigheden

Subdomein A13: Samenwerken

**Domein B Exacte wetenschappen en technologie**

Subdomein B01: Interdisciplinariteit

Subdomein B02: Wisselwerking tussen natuurwetenschap en technologie

**Domein C Aarde, natuur en heelal**

Subdomein C02: Duurzaamheid

**Domein E Technologische ontwikkeling**

Subdomein E01 Methoden en technieken van technologische ontwikkeling

Subdomein E02 Processen en producten

# 7. Vakkennis

## 7.1 Wiskunde

De wiskundige inhouden van dit hoofdstuk kunnen eenvoudig geïntroduceerd worden bij leerlingen in de bovenbouw omdat alle wiskundige voorkennis van de onderbouw nodig is: de stelling van Pythagoras, het middelpunt, het zwaartepunt, het gemiddelde, de frequentie en het percentage.

Hierbij speelt in mindere mate de voorkennis een rol, maar meer het abstractievermogen van een leerling. Er worden aanzienlijk veel nieuwe begrippen behandeld. De leerlingen zijn ook niet gewend aan dit soort wiskunde waarbij voornamelijk algoritmiek een rol speelt.

## 7.2 Natuurkunde

De voorkennis van natuurkunde:

* De leerling kan omgaan met elektrische schakelingen
* De leerling weet hoe een geluidsgolf werkt
* De leerling kent het voorvoegsel ‘kilo’ en ‘mega’.

## 7.3 Informatica

Er is geen noodzakelijke voorkennis vereist van het vak informatica. Wel speelt mee dat leerlingen die het vak volgen een voorsprong hebben op de leerlingen die het vak niet volgen.

# 8. Leerplan

## 8.1 Beschrijving per hoofdstuk

### 8.1.1 Beschrijving van hoofdstuk 1

Een toepasbare planning voor dit hoofdstuk is:

|  |  |
| --- | --- |
| Contactuur | Onderwerp |
| 1 | paragraaf 1.1 |
| 2 | paragraaf 1.2 **[toets]** |
| 3 | paragraaf 1.3 **[kaarten]** |

Voor een voorbeeld van een gedetailleerd lesplan zie onderstaande tabel.

|  |  |  |  |
| --- | --- | --- | --- |
| **Onderwijsfunctie** | **Activiteit docent** | **Activiteit leerling** | **Tijd** |
| **Contactuur 1** | | | |
| Uitleg | Concepten uitleggen:  Bit  Representatie  Pixels  Subpixels  Kleuren mengen | Luisteren en eventueel vragen stellen | 20 min |
| Zelfwerkzaamheid | Begeleiden | Vragen en opdrachten maken | 30 min |
| **Contactuur 2** | | | |
| Uitleg | Concepten uitleggen:  Grootte van data  Verzendtijd van bits  Bewerkingssnelheid van data | Luisteren en eventueel vragen stellen | 20 min |
| Oefenen | Begeleiden | Oefenen met opgaven | 10 min |
| Toetsen |  | Schriftelijke overhoring maken | 20 min |
| **Contactuur 3** | | | |
| NIeuwe kennis verwerken | Spel 1 | Spel 1 | 10 min |
| Uitleg | Vragen naar aanleiding van spel 1 beantwoorden | Luisteren en eventueel vragen stellen | 10 min |
| Samenvatten en verwerken | Spel 2 | Spel 2 | 10 min |
| Uitleg | Vragen naar aanleiding van spel 2 beantwoorden | Luisteren en eventueel vragen stellen | 10 min |
| Oefenen | Spel 3 | Spel 3 | 10 min |

### 8.1.2 Beschrijving van hoofdstuk 2

De leerlingen werken grotendeels zelfstandig aan de opdrachten van paragraaf 1 t/m 3, waarbij het grote Big Data debat bij paragraaf 2 wordt georganiseerd en begeleid door de docent(en). De leerlingen sluiten individueel het hoofdstuk af met de online cursus van bigdatauniversity waarbij de docent kan besluiten de score mee te laten wegen in het tussencijfer voor dit hoofdstuk.

Een toepasbare planner voor het hoofdstuk:

|  |  |
| --- | --- |
| Contactuur | Onderwerp |
| 1 | paragraaf 2.1 |
| 2 | paragraaf 2.2 **[het debat]** |
| 3+4 | paragraaf 2.3 **[bigdatauniversity course]** |

Meer in detail:

|  |  |  |  |
| --- | --- | --- | --- |
| **Onderwijsfunctie** | **Activiteit docent** | **Activiteit leerling** | **Tijd** |
| **Huiswerk vooraf opgeven (flipping the classroom)** | | | |
| Ter orientatie | Delen theorie hoofdstuk 3 en inventariseren vragen voor de introductie | Doorlezen hoofdstuk 3  1 vraag formuleren over de theorie, toe laten sturen via een Google Form of ander online formulier | 45 minuten |
| **Contactuur 1** | | | |
| Introductie | Vertellen, doornemen/tonen vragen + toelichten concepten:  publieke sector  private sector  data mining  IoT (Internet of Things)  Big Data analytics  Learning analytics  de feedbackcyclus | Luisteren en evt aansluitende vragen stellen | 15-20 minuten |
| Zelfwerkzaamheid + pitch als pair | Begeleiden  Zodra groepen toe zijn gekomen aan de pitch opdracht kunnen ze het beste een andere ruimte opzoeken om de leerlingen die zelfstandig werken niet te storen | Vragen en opdrachten maken | 35-40 minuten |
| **Huiswerk vooraf opgeven (flipping the classroom)** | | | |
| Voorbereiding debat en opdrachten | Delen URL documentaire: https://www.vpro.nl/programmas/tegenlicht/kijk/afleveringen/2013-2014/persoonlijke-data.html , aankondigen debat en alvast debatregels delen (afhankelijk van keuze docent debatvorm) | Documentaire bekijken | 50 minuten |
| **Contactuur 2** | | | |
| Introductie | Vertellen, doornemen/tonen vragen + toelichten concepten:  duurzaamheid via Big Data  Big Social Data  privacy en ethiek | Luisteren en evt aansluitende vragen stellen | 10 minuten |
| Zelfwerkzaamheid | Begeleiden / debat setting voorbereiden (regel desnoods een ander lokaal) | Vragen en opdrachten maken | 15 minuten |
| Discussie / beargumenteren | Refereren aan debatregels, groep organiseren (waar nodig) en begeleiden | Debatteren | 25 minuten |
| **Huiswerk vooraf opgeven (flipping the classroom)** | | | |
| Voorbereiding afsluiting hoofdstuk | Delen link bigdatauniversity en ander materiaal voor geinteresseerden. Leg uit hoe ze zich kunnen registeren en waar ze alvast kunnen beginnen, refereer aan de laatste opdracht | Registeren op de site + maken eerste opdrachten/onderdelen van data science 101 (zie opdracht 4) | 1 uur |
| **Contactuur 3 en 4** | | | |
| Introductie | Vertellen + toelichten concepten:  noodzaak om te veranderen (toekomst voor organisaties)  organisatieprincipes  The predictive analytics cycle  business manager  business analist  IT system manager (Big Data manager)  Data scientist (data analist)  Big Data opleidingen  Big Data online cursussen (zelfstudie) | Luisteren en evt aansluitende vragen stellen | 15 minuten |
| Zelfwerkzaamheid | Begeleiden | Vragen en opdrachten maken (MOOC cursus afronden) | 35 minuten |
| **Huiswerk achteraf (als de cursus niet af is, volgende les scores inventariseren)** | | | |

### 8.1.3 Beschrijving van hoofdstuk 3

Dit hoofdstuk gaat over datamining technieken die worden gebruikt bij Big data analyse. In dit hoofdstuk komen voornamelijk de twee bekendste en belangrijkste methoden van data mining aan bod: associatie en clustering. Met dit hoofdstuk gaan leerlingen basiskennis vergaren en met basis algoritmen oefenen, zodat zij de eerste stappen in datamining kunnen zetten.  
In paragraaf 1 wordt de associatie analyse behandeld. In deze paragraaf kijken we eerst naar de begrippen van verzamelingenleer. Daarbij leggen we de wiskundige basis voor het begrip associatie en de bijbehorende begrippen Support en Confidence. De datasets worden wiskundig gezien beschouwd als verzamelingen of een collectie van verzamelingen waarin de theorie van verzamelingsleer van toepassing is. Vervolgens wordt gezocht naar betrouwbare associatie d.m.v. een bekend Apriori-algoritme.  
Paragraaf 2 is een vervolg van paragraaf 1. Deze ICT-paragraaf kan gezien worden als het praktische deel met als doel de behandelde concepten en de algoritmen toegankelijker en makkelijker te maken voor de leerlingen. Ook biedt deze paragraaf de leerlingen de mogelijkheid om met bijbehorende applets uitgebreid te oefenen met problemen bij grote datasets. Daarnaast kunnen de leerlingen een start in het programmeren met Geogebra maken.  
Paragraaf 3 staat in het licht van de clusteranalyse met als leerdoel het verbeteren van clusters. Aan het begin van deze paragraaf wordt het begrip afstand uitgebreid geïntroduceerd t.o.v. een gewoon meetkundige begrip. Dit zodanig dat de nieuwe afstand toegepast kan worden bij het clusteren van objecten van datasets. Verder worden de meetkundige begrippen middelpunt en zwaartepunt algemeen gedefinieerd als centrums bij data analyse.

**Planning**

|  |  |  |  |
| --- | --- | --- | --- |
| **les** | **onderdeel** | **opmerkingen** | **applet** |
| 1 | Paragraaf 3.1 | Theorie : Associatie Analyse  Verwerkingsopdrachten |  |
| 2 | Paragraaf 3.2 | Practicum Associatie met geogebra | Associatie.ggb |
| 3+4 | Paragraaf 3.3 | Theorie : Clusteranalyse  Verwerkingsopdrachten |  |

Zie onderstaande tabel voor een gedetailleerde uitwerking van mogelijke lessen:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Onderwijsfunctie** | **Activiteit docent** | **Activiteit leerling** | **Tijd** | | |
|  | | | | | |
| **Contactuur 1** | | | | | |
|  |  |  |  | 3 min | |
| Introductie van associatie analyse  en oriëntatie op begrippen m.b.t het apriori algoritme. | Vertellen in het kort over het doel en de taak van de associatietechnieken bij Big data analyse  Prikkelen van de nieuwsgierigheid van de leerlingen  Activeren de voorkennis door vragen te stellen. | Luisteren en evt aansluitende vragen stellen | 5 min |
| Begeleide oefening met concepten | Samen starten met Klassikale bespreking van een paar korte opdrachten.    Toelichten van begrippen: verzameling, support en confidence aan de hand van simpele voorbeelden | Actief meedoen en antwoorden op de gestelde vragen.      Luisteren en evt aansluitende vragen stellen | 5min | 15 min | |
| Begeleiding van de leerlingen bij het het maken van verwerkingsopgaven |  | 10 min |
| Klassikale uitleg over het Apriorie algoritme | Uitleggen hardop hoe het apriori algoritme toegepast kan worden bij een eenvoudig probleem.    Begeleiding van de leerlingen bij het het maken van verwerkingsopgave. | Luisteren en evt aansluitende vragen stellen      Oefenen met deze opdrachten | 10 min        10 min | 20 min | |
| Zelfwerkzaamheid  . | Begeleiden de leerlingen bij zelfwerkzaamheden  Geven van feedback | Werken aan groepsopdrachten van drietallen of viertallen. |  | 15 min | |
| Evaluatie | Het geven van corrigerende feedback | Actief meedoen aan het geven van een samenvatting |  | 2 min | |
| **Contactuur 2 (Practicum)** | | | | | |
| Algemene instructie  over het practicum | Instructie geven over onder andere  de organisatie, de doelen, de werkvorm en de opdrachten van dit practicum. | Luisteren en evt aansluitende vragen stellen | 5 min | 5 min | |
| Uitleg /demo’s | Laten zien m.b.v. een demo hoe ze deze opdrachten m.b.v. Geogebra kunnen uitvoeren. |  | 10min | 10 min | |
| Zelfwerkzaamheid | Begeleiden de leerlingen bij het uitvoering van de praktische opdrachten en beoordelen van hun werk. | Werken In groepen aan de praktische opdrachten | 40min | 40 min | |
| **Contactuur 3 (Clusteranalyse)** | | | | | |
| Introductie van cluster analyse.  Oriëntatie op begrippen m.b.t. cluster anayse | Vertellen over het doel en de taak van de associatietechnieken bij Big data. | Luisteren en evt aansluitende vragen stellen | 5 min | | 3 min |
| Begeleide  oefening met concepten | Samen starten met Klassikale bespreking van een paar korte opdrachten.  Toelichten van begrippen:clusters. centrums, afstandsmaten… | Actief meedoen en antwoorden op de gestelde vragen.      Luisteren en evt aansluitende vragen stellen | 5min | | 15 min |
| Begeleiding van de leerlingen bij het het maken van verwerkingsopgaven | Maken opdrachten , zo mogelijk samen met anderen | 10 min | |
| Klassikal uitleg over het k-mean algoritme. | Uitleggen hardop van het k-mean algoritme aan de hand van een eenvoudig probleem.  Bespreken klassikale van een opdracht | Actief meedoen en antwoorden op de gestelde vragen.    Actief meedoen en vragen beantwoorden. | 10 min      5 min | | 15 min |
| Zelfwerkzaamheid | Begeleiden de leerlingen bij zelfwerkzaamheden .  Geven van feedback | Werken aan groepsopdrachten in in drietallen of viertallen. |  | | 15 min |
| Evaluatie | Het geven van corrigerende feedback | Actief meedoen aan het geven van een samenvatting |  | | 2 min |

## 8.2 Differentiatie per hoofdstuk

Deze paragraaf gaat in op differentiatie mogelijkheden binnen de module. Het mooie van deze module is dat het onderwerp Big Data op veel manier schaalbaarheid van de concepten meebrengt. Hierdoor zijn er altijd veel mogelijkheden tot differentiatie.

### 8.1.1 Differentiatie in hoofdstuk 1

In dit hoofdstuk zijn veel mogelijkheden tot differentiatie ingebouwd. In de eerste paragraaf zijn er drie niveaus in de opdrachten die de leerlingen kunnen maken. Zo kunnen ze zelf de meest uitdagende stof uitzoeken.

Bij de tweede paragraaf hoort een schriftelijke overhoring. Doordat in deze paragraaf veel wiskunde ook bod komt, is het mogelijk om de opgaven moeilijker te maken naar gelang er leerlingen zijn die het niveau te gemakkelijk vinden. De manier van berekenen die gebruikt wordt in deze paragraaf maakt het ook mogelijk om de opgaven te versimpelen, dit om de minder begaafde leerlingen ook de lesstof te laten leren.

De derde paragraaf biedt mogelijkheid voor discussie in de klas. Het is aan de docent om tijdens deze discussie het niveau te verdiepen of juist het niveau gemakkelijker te maken zodat alle leerlingen meekomen en de leerdoelen behalen.

### 8.1.2 Differentiatie in hoofdstuk 2

Dit hoofdstuk leent zich uitstekend voor het differentiëren naar voorkennis en/of interesse. Zo kan een snelle leerling naast het uitwerken van de theorie en het werken in pairs op eigen tempo de afsluitende opdracht van het hoofdstuk naar voren halen en wanneer er sprake is van genoeg interesse zelfs verschillende cursussen van de bigdatauniversity afwerken en daarbij bijvoorbeeld een studiepad aanhouden om meer basiskennis te vergaren aansluitend op de technieken die in hoofdstuk 4 verder wiskundig worden uitgewerkt. Zie voor een voorbeeld studiepad:

<https://bigdatauniversity.com/learn/data-science/>

De docent kan er dan bijvoorbeeld ook voor kiezen om daarbij door middel van een kort assessment de succesvolle afronding van dergelijke ‘extra’ cursussen terug te laten komen als een soort bonuspunten.

### 8.1.3 Differentiatie in hoofdstuk 3

In dit hoofdstuk wordt rekening gehouden met de verschillen tussen leerlingen, zodat de leerstof voor alle leerlingen toegankelijk wordt.

Dit hoofdstuk bevat opgaven met een geleidelijk stijgende moeilijkheidsgraad waarbij de opdrachten op maat worden aangeboden. Makkelijke opdrachten bieden zwakkere leerlingen de mogelijkheid om succeservaringen op te doen omdat ze de vragen kunnen beantwoorden opbouwend naar de moeilijkere opdrachten. Aangezien de hoge orde abstracte theorie bij dit hoofdstuk hoort wordt de leerstof zoveel mogelijk beeldend gemaakt. De betreffende theorieën worden uitgebreid uitgelegd met veel voorbeelden en analogieën. De uitleg krijgt ondersteuning door middel van visualisatie en schematisering. Er worden veel visuele schema’s gebruikt om de theorie van paragraaf 1 en paragraaf 3 te verduidelijken, zodat de zwakke leerlingen minder moeite hebben om de leerstof te begrijpen en goede leerlingen zelfstandig kunnen werken.

# 9. Toelichting bij leerlingopdrachten

## 9.1 Hoofdstuk 1

**Spel 1**

De leerlingen werken voor deze opdracht in groepjes van drie.

Het doel van het spel is de kaarten op de juiste stapel te leggen.

Er zijn drie stapels:

1. Er is geen consistentheid tussen bronnen
2. Er is geen betrouwbaarheid tussen bronnen
3. Er is geen (nuttige) combinatie mogelijk tussen bronnen.

|  |  |  |
| --- | --- | --- |
| Bron 1  06-nummer  Bron 2  +316-nummer | Bron 1  Een persoon met een 06-nummer.  Bron 2  Dezelfde persoon met een ander 06-nummer. | Bron 1  Eén persoon zonder 06-nummer.  Bron 2  Een 06-nummer van een ander persoon. |
| Bron 1  Postcode  Bron 2  Adres horend bij een andere postcode. | Bron 1  Postcode  Bron 2  Onjuist geschreven adres maar valt wel binnen deze postcode. | Bron 1  Postcode  Bron 2  Een andere postcode |
| Bron 1  Twee personen zijn vriend.  Bron 2  Dezelfde twee personen zijn geen vrienden. | Bron 1  Twee personen hebben een relatie.  Bron 2  Dezelfde twee personen zijn getrouwd met elkaar. | Bron 1  Twee personen zijn geen vrienden.  Bron 2  Twee andere personen zijn getrouwd. |

**Spel 2**

De leerlingen werken voor deze opdracht in groepjes van drie.

Het doel van het spel is met onderstaande kaarten een domino te vormen.

|  |  |  |
| --- | --- | --- |
| Bron 1  06-nummer  Bron 2  +316-nummer | Bron 1  Een persoon met een 06-nummer.  Bron 2  Dezelfde persoon met een ander 06-nummer. | Bron 1  Eén persoon zonder 06-nummer.  Bron 2  Een 06-nummer van een ander persoon. |
| Betrouwbaarheid | Combineren | Consistentie |

|  |  |  |
| --- | --- | --- |
| Bron 1  Postcode  Bron 2  Adres horend bij een andere postcode. | Bron 1  Postcode  Bron 2  Onjuist geschreven adres maar valt wel binnen deze postcode. | Bron 1  Postcode  Bron 2  Een andere postcode |
| Combineren | Consistentie | Betrouwbaarheid |

|  |  |  |
| --- | --- | --- |
| Bron 1  Twee personen zijn vriend.  Bron 2  Dezelfde twee personen zijn geen vrienden. | Bron 1  Twee personen hebben een relatie.  Bron 2  Dezelfde twee personen zijn getrouwd met elkaar. | Bron 1  Twee personen zijn geen vrienden.  Bron 2  Twee andere personen zijn getrouwd. |
| Consistentie | Betrouwbaarheid | Combineren |

**Spel 3**

De leerlingen werken voor deze opdracht in groepjes van drie.

Het doel van het spel is memory te spelen met te kaarten. Hierdoor wordt het gemakkelijker voor de leerlingen de relevante begrippen te onthouden.

|  |  |  |
| --- | --- | --- |
| Bron 1  06-nummer  Bron 2  +316-nummer | Bron 1  Een persoon met een 06-nummer.  Bron 2  Dezelfde persoon met een ander 06-nummer. | Bron 1  Eén persoon zonder 06-nummer.  Bron 2  Een 06-nummer van een ander persoon. |
| Consistentie | Betrouwbaarheid | Combineren |
| Bron 1  Postcode  Bron 2  Adres horend bij een andere postcode. | Bron 1  Postcode  Bron 2  Onjuist geschreven adres maar valt wel binnen deze postcode. | Bron 1  Postcode  Bron 2  Een andere postcode |
| Consistentie | Betrouwbaarheid | Combineren |
| Bron 1  Twee personen zijn vriend.  Bron 2  Dezelfde twee personen zijn geen vrienden. | Bron 1  Twee personen hebben een relatie.  Bron 2  Dezelfde twee personen zijn getrouwd met elkaar. | Bron 1  Twee personen zijn geen vrienden.  Bron 2  Twee andere personen zijn getrouwd. |
| Consistentie | Betrouwbaarheid | Combineren |

## 9.2 Hoofdstuk 2

Gedurende dit hoofdstuk maken leerlingen bij het maken van de opdrachten gebruik van internet waardoor het noodzakelijk is dat er een computerlokaal beschikbaar is. Het is ook mogelijk, wanneer de docent dat aandurft, om leerlingen het via een mobiele telefoon te laten opzoeken. Voor de uitvoering van de MOOC in de laatste paragraaf is een scherm van redelijk formaat vereist.

## 9.3 Hoofdstuk 3

Voor de uitvoering van ICT paragraaf 2 dient een computerlokaal beschikbaar te zijn. Bij de uitwerking van de betreffende opdrachten is het Geogebra programma nodig. De uitvoering van de installatie van Geogebra moet met het systeembeheer worden besproken. Verder hebben leerlingen applets nodig bij het maken van een aantal opdrachten in paragraaf 3.2.

Het programma Geogebra is

# 10. Bronnen bij leermateriaal

## 10.1 Hoofdstuk 1

**Paragraaf 1.1**

Niveau 1

1a) 2

1b) 4

2a) 00000001

2b) 00000010

3) 0001011000

4) 100101

5) 100101110101111101111111

Niveau 2

1) 256

2a) 000000010000001000000011

2b) 0000000110000000000000101000000000000011

3) 001010100100111111100100

4) 111011010111001100101101

5) 111001110010111010110110111111111011101010101110

Niveau 3

1) 65.536

2a) De decimale getallen zijn: 125 055 142 157 143 150 164

De binaire getallen zijn: 01111101 00110111 10001110 10011101 10001111 10010110 10100100

2b) De decimale getallen zijn: 061 053 061 075 062

De binaire getallen zijn:

2c) De leestekens staan niet in de tabel

De decimale getallen zijn: 114 000 000 000 124 000 153 000 000 163

De binaire getallen zijn: 01110010 00000000 00000000 00000000 01111100 000 00000 10011001 00000000 00000000 10100011

3) De gelezen waarden zijn:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Tijd** | **Waarde** | **Decimaal** | **Absoluut** | **Afgerond** | **Binair** |
| 0 | 0.5 | 4 | 12 | 12 | 1100 |
| 0.5 | 0 | 0 | 8 | 8 | 1000 |
| 1 | -1 | -8 | 0 | 0 | 0 |
| 1.5 | 0.5 | 4 | 12 | 12 | 1100 |
| 2 | 0.2 | 1.6 | 9.6 | 10 | 1010 |
| 2.5 | 0.1 | 0.8 | 6.4 | 6 | 0110 |
| 3 | 0 | 0 | 0 | 0 | 0000 |
| 3.5 | -0.6 | -4.8 | 3.2 | 3 | 0011 |
| 4 | 0.9 | 7 | 15 | 15 | 1111 |
| 4.5 | 0 | 0 | 8 | 8 | 1000 |
| 5 | -0.5 | -4 | 4 | 4 | 0100 |

4) 000 000 000

010 000 000

100 000 000

111 000 000

000 111 111

111 111 000

111 000 111

111 111 111

5)

000 000 000

111 111 000

010 000 000

111 111 000

100 000 000

111 000 111

111 000 000

111 111 111

000 000 111

111 111 000

000 111 000

111 111 111

000 000 000

111 111 111

010 000 000

111 111 111

100 100 100

010 010 000

010 000 000

010 010 000

100 000 000

010 000 000

111 000 000

000 010 000

**Paragraaf 1.3**

**Spel 1**

Stapel 1 - consistentheid

|  |  |  |
| --- | --- | --- |
| Bron 1  06-nummer  Bron 2  +316-nummer | Bron 1  Een persoon met een 06-nummer.  Bron 2  Dezelfde persoon met een ander 06-nummer. | Bron 1  Eén persoon zonder 06-nummer.  Bron 2  Een 06-nummer van een ander persoon. |

Stapel 2 - Betrouwbaarheid

|  |  |  |
| --- | --- | --- |
| Bron 1  Postcode  Bron 2  Adres horend bij een andere postcode. | Bron 1  Postcode  Bron 2  Onjuist geschreven adres maar valt wel binnen deze postcode. | Bron 1  Postcode  Bron 2  Een andere postcode |

Stapel 3 - Combineren

|  |  |  |
| --- | --- | --- |
| Bron 1  Twee personen zijn vriend.  Bron 2  Dezelfde twee personen zijn geen vrienden. | Bron 1  Twee personen hebben een relatie.  Bron 2  Dezelfde twee personen zijn getrouwd met elkaar. | Bron 1  Twee personen zijn geen vrienden.  Bron 2  Twee andere personen zijn getrouwd. |

**Spel 2**

Zie de stapels hierboven welke kaartjes bij welk concept horen.

**Spel 3**

Zie de stapels bij spel 1 welke kaartjes bij elkaar horen.

## 

## 

## 10.2 Hoofdstuk 2

**Paragraaf 2.1**

1a ) Data mining

1b ) Big Data Analytics

2a ) sensoren en IoT

2b ) enerzijds omdat het om grote hoeveelheden data gaat en anderzijds omdat vaak op basis van realtime data aanpassingen moeten worden gedaan, dat kan alleen als het over een netwerk verzameld en verzonden wordt.

2 c)

Hier zijn vele antwoorden mogelijk, maar het volgende is bijvoorbeeld een aardige uitwerking:

meten:

kijkgedrag bijhouden van de gebruikers

inzicht:

welke series/films zijn het meest populair onder gebruikers

kansen:

voor de plaatsing van een nieuwe serie/film kan met vrij grote nauwkeurigheid op basis van het verkregen inzicht bepaald worden of het veel bekeken zal worden

actie:

door extra content te plaatsen dat aansluit op de trends en/of huidige populaire content of optimalisaties voor beschikbaarheid door te voeren op de meest populaire content zodat het systeem de minste fouten zal vertonen richting de gebruikers op de voor hen meest belangrijke content

3 a) Door middel van Big Data kan een docent meer maatwerk leveren, sneller in kaart brengen wat/wel niet werkt bij de leerlingen, waar nog extra uitleg over nodig is en waarop eventueel actie moeten worden ondernomen.

3 b) Het verwerken en daadwerkelijk structureel inzetten van gegevens over de leerlingen in een school vereist de nodige expertise en digitale vaardigheden. Verder speelt ook privacy een rol; wie beheert precies al die 'privé' gegevens van leerlingen en wat wordt er allemaal mee gedaan?

4 a) Op het punt van actie (de laatste stap in de cyclus) zal er minder snel volledige automatisering plaatsvinden. Als het gaat om het nemen van beslissingen, wat dan vaak de actie is, is de mens namelijk nog erg huiverig om dat volledig aan een computer over te laten. Denk bijvoorbeeld aan bepaalde operaties in een ziekenhuis. Er kan dan bijvoorbeeld maatwerk worden verricht aan de hand van bepaalde besluitvorming waarbij de computer wel een rol heeft gespeeld, maar niet als enige. Voor innoveren zal voor langere tijd de computer een 'ondersteunende' rol hebben waarbij de innovatie gezien kan worden als een vorm van geautomatiseerd experimenteren om te zien of iets 'totaal nieuws/anders' tot grote onvoorziene verbeteringen kan leiden.

4 b) Omdat steeds meer onderdelen van onze maatschappij door middel van automatisering (op basis van Big Data) aan het veranderen is wordt het ook tegelijkertijd steeds belangrijker dat we begrijpen hoe een computer algoritme bij deze automatisering tot beslissingen komt en op welke data het algoritme dat baseert.

Heel concreet: als we als mens niet meer kunnen herleiden of enig begrip hebben van de manier waarop machines namens ons de leefomgeving verbeteren/beïnvloeden dan kan dat grote gevolgen hebben voor zaken als gezondheid, veiligheid en privacy.

5 ) Leerlingen moeten een schets maken waarbij binnen een cirkelachtige vorm de volgende elementen achtereenvolgens aan bod komen: Data - Context - Kansen - Actie

Een voorbeeld:



**Paragraaf 2.2**

1a ) Door Big Data in te zetten voor het analyseren van het gebruik van allerlei (schadelijke) grondstoffen/chemicaliën binnen de organisatie is het mogelijk om naast kostenbesparing ook minder het milieu te belasten en dus 'duurzamer' bezig te zijn. Zie het voorbeeld van de waterzuivering in de paragraaf.

1b ) Door Big Data in te zetten voor het grootschalig analyseren van de effecten van bepaalde ingrepen/medicijnen is het mogelijk om verbanden te leggen, ziektebronnen vroeger te detecteren en zelfs te voorkomen.

Verder mogelijke bron om aan te halen:

<https://www.greycampus.com/opencampus/big-data-developer/applications-of-big-data>

2 a) Extraversie tegenover introversie, Service gerichtheid tegenover interesse, zorgvuldigheid tegenover onzorgvuldigheid, emotionele stabiliteit tegenover emotionele instabiliteit, openheid tov nieuwe ervaringen tegenover gesloten koppigheid

2 b) Informatie over jouw persoonlijkheid maakt het mogelijk voor bedrijven om voorspellingen te doen over zaken als aankoopgedrag, de dingen waar je op klikt e.d.

Als je dat soort zaken kunt voorspellen weet je bij voorbaat welk product in welke mate zal verkopen.

2 c) Het is onduidelijk in hoeverre HR-afdelingen deze tools al gebruiken, maar het ligt voor de hand dat een bedrijf deze informatie wel degelijk tegen jou kan gebruiken en 'onbewust' nadelig kan zijn bij een aankomende of lopende sollicitatie.

3. Het is aan de docent om de debatvorm te bepalen; zo kan er gekozen worden voor het opsplitsen van de klas in 2 teams (voor- en tegen de stelling, met eventueel een voorzitter) of een debat volgens de methode van het lagerhuis, zie de site: <http://www.debatindeklas.nl/debatvormen/lagerhuis/>.

**Paragraaf 2.3**

1 a) In de private sector zien we dat nieuwe bedrijven, de ‘grote jongens’ inhalen doordat zij meteen vanaf het begin hun organisatiestructuur hebben opgebouwd aan de hand van de huidige technologische mogelijkheden. Zij zijn daardoor in staat om op korte termijn veranderingen door te voeren en tegen lagere kosten betere diensten te leveren. Klanten zijn hier vaak erg van gecharmeerd omdat het bedrijf daarmee als het ware meebuigt, snel aanpassingen maakt en dat ook nog tegen een redelijke prijs.

Als je dan als bestaand bedrijf niet verandert prijs je je vanzelf uit de markt en ga je failliet

1 b) Wegens economische druk moet men naar naar manieren zoeken om kosten te drukken zodat er enerzijds financiële ruimte is voor vernieuwing maar ook lastenverlichting in de vorm van belasting aan de kant van de burger.

2 a) Doordat de focus ligt op een vernieuwing elke paar jaar wordt het lerend vermogen van de organisatie over het hier en nu sterk belemmerd. Het lerende vermogen is dus het grootste pijnpunt.

2 b) De data leidend maken, zoveel mogelijk databronnen als input gebruiken voor analyse, een korte feedbackcyclus in stand houden, korte lijntjes en minder lagen in de organisatie aanbrengen en meer opdrachtgestuurd werken aan de hand van de data

3 a) Business manager, Business analist, IT system manager en data scientist

3 b) Business manager

3 c) Data scientist

3 d) IT system manager

4 ) nvt

## 

## 

## 10.3 Hoofdstuk 3

**Paragraaf 3.1**

**1.**

a.

A∪B ={1,2,3,4,5}∪{4,5,6,7}={1,2,3,4,5,6,7}

A∪C ={1,2,3,4,5}∪{5,8,11}={1,2,3,4,5,8,11}

B∪C ={4,5,6,7}∪{5,8,11}={4,5,6,7,8,11}

b.

A∩B ={1,2,3,4,5}∩{4,5,6,7}={4,5}

A∩C ={1,2,3,4,5}∩{5,8,11}={5}

B∩C ={4,5,6,7}∩{5,8,11}={5}

c.

A∪B∪C={1,2,3,4,5}∪{4,5,6,7}∪{5,8,11}={1,2,3,4,5,6,7,8,11}

**2.**

a.

T1∩T2 ={chips,zeep,appels}∩{ chips,zeep }={chips,zeep

T1∩T3 ={chips,zeep,appels}∩{ chips,zeep,bananen}={chips,zeep}

T4∩T3 ={ chips,zeep,bananen}∩{ zeep,bananen}={zeep,bananen}

b.

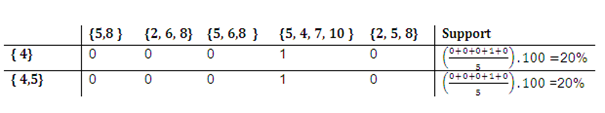
T1∪T2 ={chips,zeep,appels}∪{ chips,zeep }={chips,zeep,appels}

T1∪T3 ={chips,zeep,appels}∪{ chips,zeep,bananen}={chips,zeep,appels,bananen}

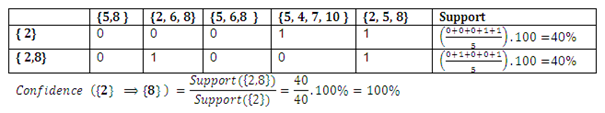
T4∪T3 ={ chips,zeep,bananen}∪{ zeep,bananen}={chips,zeep,bananen}

**3.**

**a.**

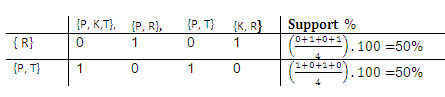


**b.**

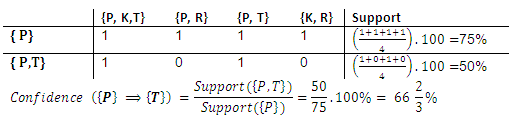


**4.**

a.

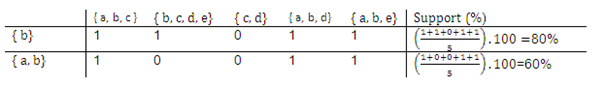


b.

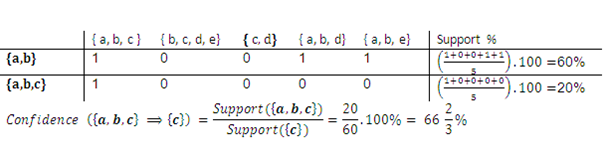


**5.**

a.

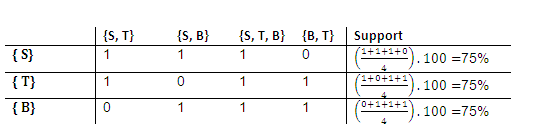


b.

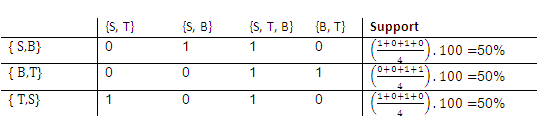


**6.**

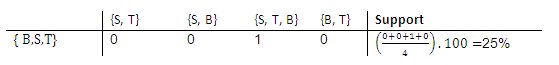
**a.**



**b.**

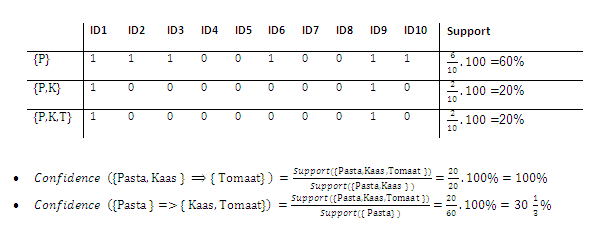


**c.**



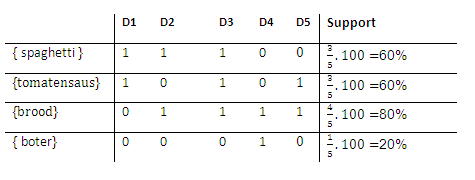
**7.**

* Stel P=Pasta , K =Kaas en T=Tomaat

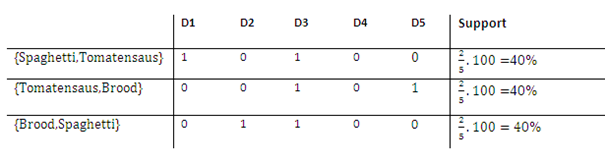


**8.**

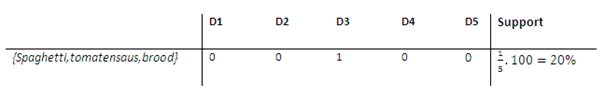
a.



b.



c.



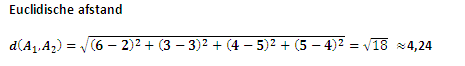
**Paragraaf 3.2**

Voor de leesbaarheid van deze docentenhandleiding zijn de uitwerkingen van deze paragraaf verplaatst naar Bijlage D.

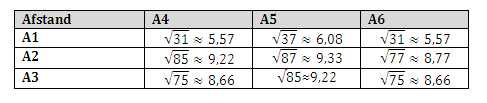
**Paragraaf 3.3**

**1.**

a



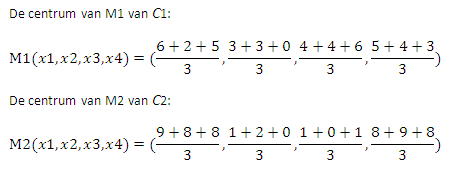
b.



c.

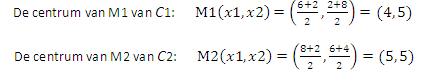


d.

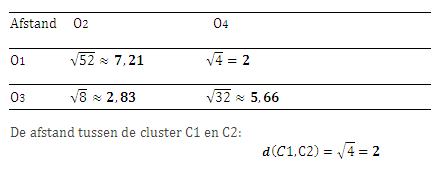


**2**.

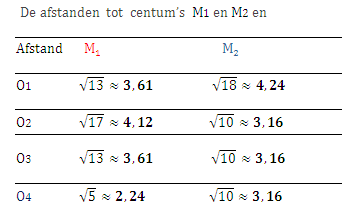
a.



b.

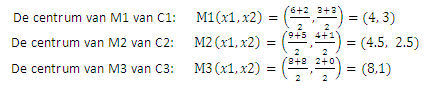


c.

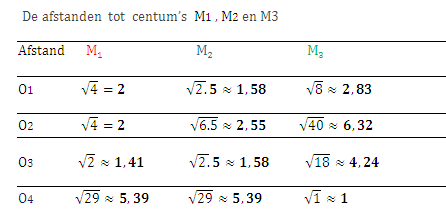


**3.**

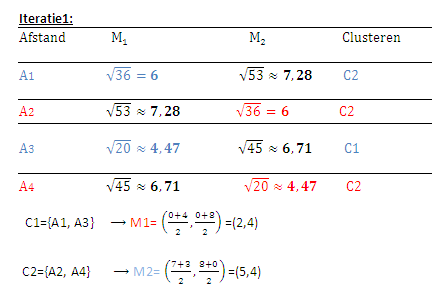
a.



b.

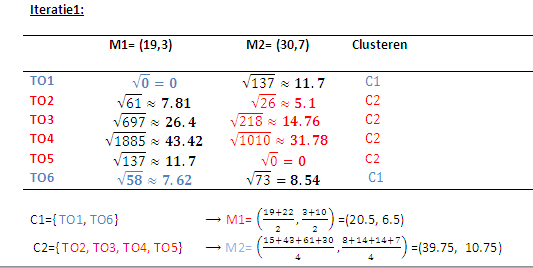


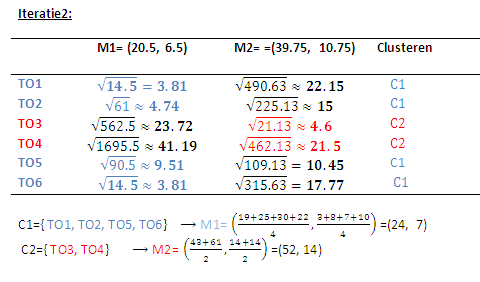
**4**.

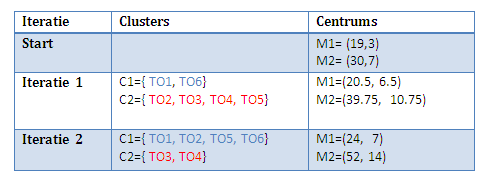


# 

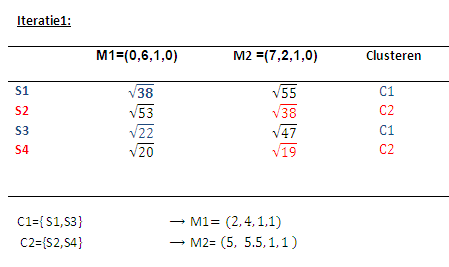
**5**.

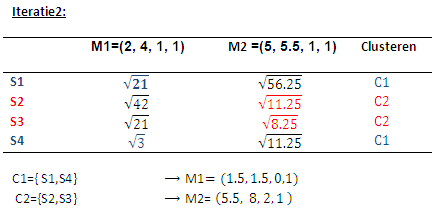


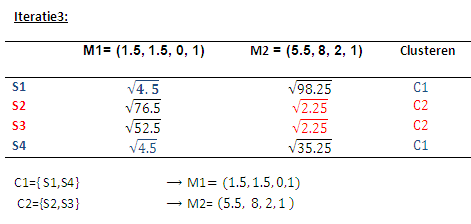


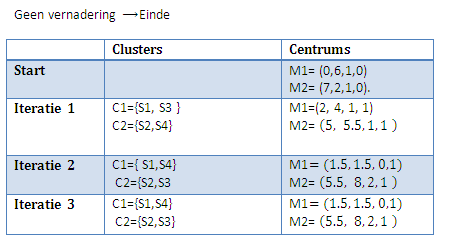


**6.**









# 

# 

# 11. Toetsing

## 11.1 Hoofdstuk 1

Na afloop van het behandelen van de tweede paragraaf wordt er een SO gegeven. Een voorbeeld SO staat beschreven in Bijlage A. Deze SO telt voor 30% mee met het eindcijfer.

De leerlingen kunnen een bonuspunt behalen op dit SO als ze een klein verslag (maximaal 2 A4) inleveren bij de docent. Dit verslag moet de antwoorden van paragraaf 1.1 en paragraaf 1.3 bevatten.

## 11.2 Hoofdstuk 2

Professionaliteitscijfer (Huiswerk elke keer gemaakt, pitch ok e.d.) (10%)

Beoordeling score MOOC (20%)

**De Big Data skills mindmap**

Optioneel kan een cijfer worden gegeven voor deze uitgebreide mindmap dat meetelt voor het eindcijfer van de gehele module.

## 11.3 Hoofdstuk 3

Dit hoofdstuk wordt afgesloten met een toets die uit twee delen bestaat:

Het eerste deel is een schriftelijke toets.

Het tweede deel is een practicum waarbij de leerlingen opdrachten gaan uitwerken met Geogebra applets. Aan het einde van de toets is er een verslag met de uitwerking van de gestelde vragen.

Beoordeling een schriftelijke toets(30%) en een praktische toets met Geogebra(10%)

# 12. Suggesties en extra opdrachten

Dit hoofdstuk bevat suggesties bij de leerdoelen, voor de uitleg of bij de opdrachten. Ook worden er in dit hoofdstuk een aantal extra (keuze)opdrachten ter beschikking gesteld.

## 12.1 Hoofdstuk 1

Tijdens de eerste les (paragraaf 1.1) kunnen leerlingen zelf opgaven maken voor elkaar. De opgaven zijn vrij gemakkelijk te maken maar het kost wel veel tijd om de juiste antwoorden te vinden.

Voor de derde les (paragraaf 1.3) zou de docent de leerlingen kunnen vragen om een tweetal bronnen met data mee te nemen of te bedenken. Tijdens de les kunnen de leerlingen nagaan of de bronnen te combineren zijn. Door de leerlingen zelf een onderwerp te laten aandragen wordt de lesstof relevanter voor ze. Ook werkt dit activerend.

## 12.2 Hoofdstuk 2

2.3.3

4) De MOOC is zo opgezet dat leerlingen direct worden beoordeeld op basis van hun antwoorden. Als het engels voor een leerling te moeilijk is kan het verstandig zijn om leerlingen te koppelen en pairs te maken waarbij de ene leerling beter in het engels is dan de ander zodat de andere leerlingen zich kan optrekken.

Keuzeopdracht als extra ter afsluiting van hoofdstuk 2:

**De Big Data skills mindmap**

Laat leerlingen in tweetallen op een A3 vel een uitgebreide mindmap maken.

Zij kunnen daarvoor online profielen en vacatures bekijken (tip: laat ze op engels zoeken en zoeken naar functieomschrijvingen die verschillende naam hebben.

Criteria:

De in paragraaf 2.3.3 genoemde typen experts zijn de basis en de vaardigheden en verbindingen/overeenkomsten tussen elk type expert moeten duidelijk zijn.

Het idee is om hier zoveel mogelijk de aangereikte theorie uit hoofdstuk 1 te gebruiken en bronnen op het internet om het geheel uit te breiden.

## 12.3 Hoofdstuk 3

**Suggesties**

* Voor de leerlingen die met Geogebra nauwelijks ervaring hebben is het van tevoren verstandig om Geogebra handleidingen van het internet door te nemen.

URL: <https://wiki.geogebra.org/nl/Hoofdpagina>

* Het is verder nuttig om een nieuwe ICT paragraaf te ontwerpen, die goed bij paragraaf 3 aansluit. Deze extra paragraaf zou dan als doel hebben het extra oefenen met de theorie van clusteranalyse en bijbehorende algoritmen.

**Keuze opdrachten**

* Laat leerlingen zelf zoeken naar andere variant van het Apriori-algoritme op internet. Daarna gaan ze het toepassen bij één van de opdrachten van 6 tot en met 8 van paragraaf 2 en de resultaten met die van de eerste variant. Op deze manier vergaren leerlingen meer kennis en leren nieuwe theorie te koppelen aan de eerder gegeven lesstof.
* Laat leerlingen uitgebreid oefenen met extra opdrachten, met het k-mean algoritme in het geval dat k groter is dan twee.
* Laat leerlingen in drietallen een Geogebra programma (Script) ontwerpen voor

berekening van afstanden van zowel twee punten en als twee gegeven

verzamelingen. Het voornaamste doel is om hier zoveel mogelijk van de geleerde

theorie uit hoofdstuk 3 te gebruiken en daarbij programmeer vaardigheid te

ontwikkelen.

# 13. Achtergrondinformatie

Dit hoofdstuk bevat twee paragrafen met achtergrondinformatie. In paragraaf 13.1 geven we een aantal bronnen waar meer achtergrondinformatie te vinden is over de contexten en concepten uit de module. In paragraaf 13.2 worden een aantal opleidingen besproken die specifiek zich richten op een aantal concepten uit de module.

## 13.1 Contexten en concepten

Hieronder staan een aantal bronnen met achtergrondinformatie over de context(en) en concepten uit de module.

<https://bigdatauniversity.com/>

<http://www.duurzaambedrijfsleven.nl/>

<https://www.mediawijsheid.nl/bigdata/>

<https://www.kaggle.com/>

## 13.2. Studieinfo

In hoofdstuk 2 bij de paragraaf “Nieuwe organisaties, nieuwe banen” komen specifiek de bijbehorende studies als volgt aan bod:

**Data scientists**

Data scientists hebben vaak een vooropleiding gehad binnen de volgende velden: (toegepaste) wiskunde, statistiek, technische informatica en econometrie.

In Delft bestaat hier bijvoorbeeld de studie Technische Informatica en Technische Wiskunde voor en in Rotterdam de studie Econometrie. Er bestaan tegenwoordig echter ook al zeer gerichte opleidingen op het gebied van data science. Zo biedt de universiteit in eindhoven de opleiding Data Science aan.

**Business analists & business managers**

Business analists en business managers hebben vaak een vooropleiding gehad in de vorm van business management met een combinatie van IT. In Rotterdam bestaat hier bijvoorbeeld de opleiding Business Information Management voor.

Er zijn ook vaak aansluitende masters of minors binnen de opleiding voor Big Data te vinden waardoor iemand zich vanuit het bedrijfsperspectief al kan richten op het werken met Big Data.

**IT system manager**

IT system managers hebben vaak een vooropleiding gehad in de vorm van informatica of bedrijfskundige informatica. Dergelijke opleidingen worden tegenwoordig door het hele land aangeboden en geven de mogelijkheid om heel breed uit te stromen.

Door het grote aanbod van opleidingen die in meer of mindere mate te maken hebben met Big Data is de afgelopen jaren de nadruk komen liggen op zelfstudie.

Werkgevers vragen vaak om aantoonbare kennis in het gebruik van bepaalde software pakketten en/of resultaten van online sites zoals <https://bigdatauniversity.com/> en <https://www.kaggle.com/>.

# Bijlage A - SO

## Toets

Naam: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Klas: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. Hoeveel waarden kunnen 2 bits bevatten?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

2. Zet de volgende tekst om in bits. Gebruik hiervoor de tabel:

|  |  |  |  |
| --- | --- | --- | --- |
| **Teken** | **Bits** | **Teken** | **Bits** |
| a | 0001 | A | 1001 |
| b | 0010 | B | 1010 |
| d | 0011 | D | 1011 |
| g | 0100 | G | 1100 |
| i | 0101 | I | 1101 |
| t | 0110 | T | 1110 |
| ENTER | 0111 | SPATIE | 1111 |

a. “big”

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

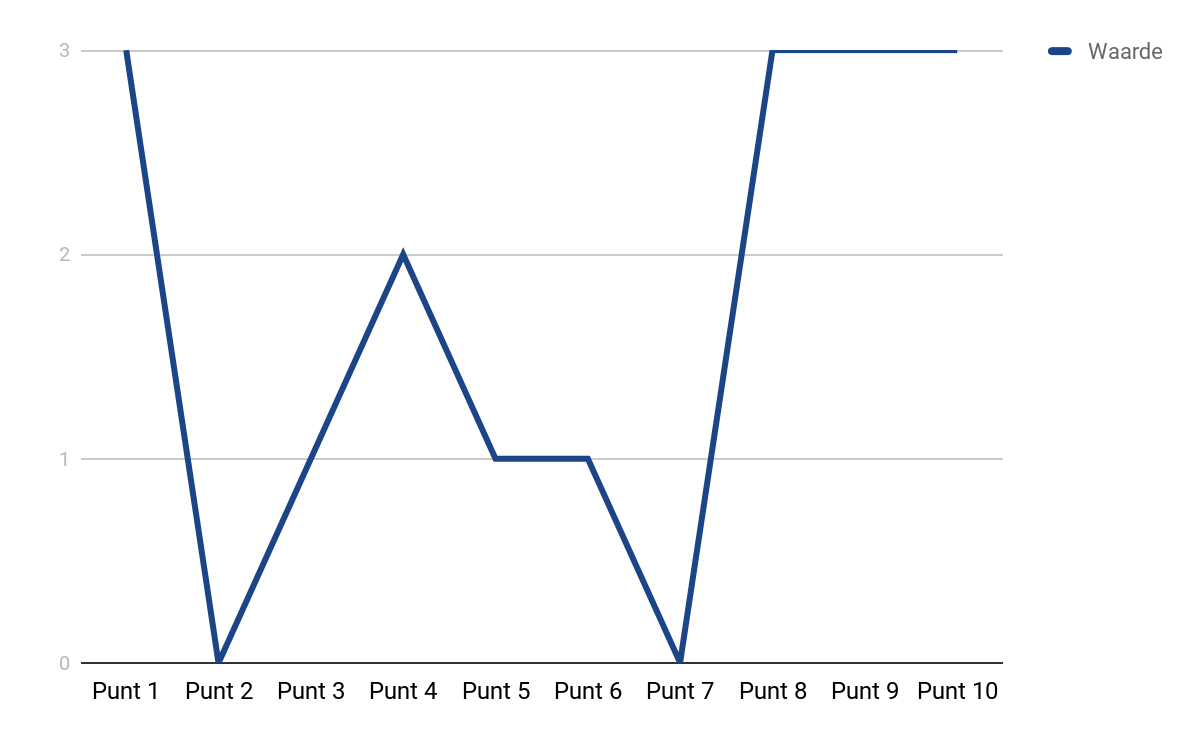
b. “data”

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

c. “Big Data”

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

3. Zet de volgende grafiek om in bits. Gebruik twee bits per punt.



\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

4. Zet de volgende afbeelding om in bits. Gebruik twee bits per pixel.

Hint: zwart heeft waarde 0.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

Schrijf de bits van elke rij pixels op een aparte regel.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

## A.2 Antwoordmodel

1. 4 bits

2a. 0010 0101 0100

2b. 0011 0001 0110 0001

2c. 1010 0101 0100 1101 1011 0001 0110 0001

3. 11 00 01 10 01 01 00 11 11 11

4. 01 11 10 11

11 00 11 01

01 00 10 00

## A.3 Becijfering

|  |  |  |
| --- | --- | --- |
| Vraag | Punten | Opmerkingen |
| 1 | 1 |  |
| 2 | 3 | -0.25 punt voor elke foute combinatie van bits |
| 3 | 3 | -0.25 punt voor elke foute combinatie van bits |
| 4 | 3 | -0.25 punt voor elke foute combinatie van bits |

Het uiteindelijke cijfer is: 1 + 9 \* (behaalde punten / 10)

# Bijlage B - Toets

**Voorbeeld van schriftelijke toetsopgaven:**

**Opgave 1**

Een fruitsap producent experimenteer met fruitsap combinaties.

Gegeven een dataset D met combinaties gesuggereerd door testpersonen:

|  |  |
| --- | --- |
| **tid** | **items** |
| 1  2  3  4  5  6  7  8  9  10 | {“Sinaasappel”,“Appel”,“Kers”}  {“Sinaasappel”, ”Ananas”, “Mango”}  {“Sinaasappel”,“Kers”}  {”Ananas”, “Mango”}  {“Appel”,“Kers”}  {“Sinaasappel”,“Kers”,”Ananas”, “Mango”}  {“Kers”,”Ananas”,“Mango”}  {“Appel”,“Kers”}  {“Sinaasappel”,“Kers”,”Ananas”, “Mango”}  {“Appel”} |

Gegeven de volgende gegevensverzamelingen:

IT1= {“Sinaasappel”,“Appel”,“Kers”}, IT4= {”Ananas”, “Mango”}, IT5= {“Appel”,“Kers”} en

IT6= {“Sinaasappel”,“Kers”,”Ananas”, “Mango”}

a. Schrijf de volgende verzamelingen op:

Gegeven de volgende items:

{“Appel”}, {“Kers”}, en {“Sinaasappel”, ”Ananas”},

b. Bereken support van deze items.

Gegeven de volgende associatieregels:

{“Appel”} ⟶{“Kers”}, {“Kers”} ⟶{“Ananas. Mango”} en {“Sinaasappel, Ananas”}⟶{“Mango”}

c. Bereken confidence voor deze regels.

Gegeven is een dataset waar elke transactie een itemset heeft (zie tabel hiernaast)

**Opgave 2**

Gegeven is een dataset waar elke transactie een itemset heeft (zie tabel hiernaast)

|  |  |
| --- | --- |
| Tid | Items |
| 1 | {“D”} |
| 2 | {“A”,“C”} |
| 3 | {“B”,“E”} |
| 4 | {“D”,“C”} |
| 5 | {“D”,“C”} |
| 6 | {“A”,“D”,“C”} |
| 7 | {“A”,“B”,“E”} |
| 8 | {“C”,“B”,“E”} |
| 9 | {“A”,“C”,“B”,“E”} |
| 10 | {“A”,“C”,“B”,“E”} |

Gegeven de volgende deelverzamelingen:

{“A”, “B”, “C” }⟶ {“D”} en {“A”, “B”}⟶ {“C” ,“D”}

a. Bereken confidence voor deze associatieregels

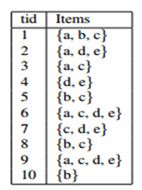
Gegeven de volgende twee associatieregels:

{“A”, “B”, “C” }⟶ {“D”} en {“A”, “B”}⟶ {“C” ,“D”}

b. Bereken confidence voor deze associatieregels.

**Opgave 3**

Gegeven is een dataset D waar elke transactie een itemset heeft (zie tabel hieronder)



a. Bereken confidence voor associatieregels:

{“d”}⟶{“e”} en {“b”} ⟶{“c”}

b. Illustreer het Apriori-algoritme door te tonen hoe dit algoritme alle 2-itemsets met een minimale support van 60% in deze dataset vindt.

**Opgave 4**

Gegeven 4 geneesmiddelen en elk heeft twee kenmerken (pH-Index en gewicht).

Het doel is deze objecten te clusteren in twee geneeskunde groepen)( zie tabel hieronder)

|  |  |  |
| --- | --- | --- |
| Geneesmiddel | Gewicht | pH-Index |
| A | 1 | 1 |
| B | 2 | 1 |
| C | 4 | 3 |
| D | 5 | 4 |

Hierbij wordt de normale (euclidische) afstand als afstandsmaat gebruikt.

a. Bereken de Euclidische afstand van van .

D={ A(1,1), B(2,1), C(4,3), D(5,4) }

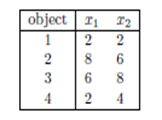
b. Bereken de centrum van de cluster D

c. Pas de k-means(k=2) methode op D tot een maximum van 3 stappen. Noteer voor elke iteratie welke cluster gevormd worden, wat de cluster centroide zijn en vul deze resultaten in de onderstaande tabel in:

|  |  |  |
| --- | --- | --- |
| **Start** |  | M1= (2.5, 2) M2= (3.5, 2,5) |
| **Iteratie 1** | C1={....................}  C2={………………….} | M1=…….  M2=…….. |
| **Iteratie 2** | C1={………………..}  C2={………………..} | M1= ……  M2= …… |
| **Iteratie 3** | C1={………………..}  C2={………………..} | M1= ……  M2= …… |

**Opgave 5**

Gegeven is een dataset met 4 objecten (zie tabel hieronder):



Cluster de waarden d.m.v. het k-means algoritme tot twee iteraties. Initialiseer random met

cluster-centra (2,2) en (6,8)

Noteer voor elke iteratie welke cluster gevormd worden en wat de cluster centroide zijn en vul

deze resultaten in de onderstaande tabel in:

|  |  |  |
| --- | --- | --- |
|  | **Clusters** | **Centrums** |
| **Start** |  | M1= (2,2) M2= (6,8) |
| **Iteratie 1** | C1={………………..}  C2={………………..} | M1=……  M2=…….. |
| **Iteratie 2** | C1={………………..}  C2={………………..} | M1= ……  M1= …… |

**Opgave 6**

Gegeven is dataset D met n = 8 datapunten, die elk 2-dimensionaal zijn (x, y):

D={(0,1),(1,0),(0,2),(1,3),(4,1),(5,2),(4,4),(5,3) }

We willen deze data clusteren, wat we kunnen doen door middel van de k-means methode.

Kies k = 2, en als initiële cluster centroids:

Pas de k-means methode toe op D tot een maximum van 3 stappen, Noteer voor elke iteratie

welke cluster gevormd worden, wat de cluster centroide zijn en vul deze resultaten in de

onderstaande tabel in:

|  |  |  |
| --- | --- | --- |
| **Start** |  | M1= (3, 1), M2= (3, 4) |
| **Iteratie 1** | C1={ ……………….}  C2={………………..} | M1=…….  M2=…….. |
| **Iteratie 2** | C1={………………..}  C2={………………..} | M1= ……  M2= …… |
| **Iteratie 3** | C1={………………..}  C2={………………..} | M1= ……  M2= …… |

# Bijlage C - Antwoorden practicum

**Opgave 1**

Gegeven de volgende dataset:  
S={{“A”,“B”},{“A”,“D”},{“B”,“C”,“D”},{“A”,“C”,“D”},{“A”,“C”,“D”,“F”},,{“A”,“B”,“E”}}

Gebruik het apriori algoritme om alle itemsets met een minimale support van 60% in deze

dataset te vinden.

**Opgave 2**

Een fruitsap producent experimenteer met fruitsap combinaties.

Gegeven een dataset D met combinaties gesuggereerd door testpersonen:

|  |  |
| --- | --- |
| **tid** | **items** |
| 1  2  3  4  5  6  7  8  9  10 | {“Sinaasappel”,“Appel”,“Kers”}  {“Sinaasappel”, ”Ananas”, “Mango”}  {“Sinaasappel”,“Kers”}  {”Ananas”, “Mango”}  {“Appel”,“Kers”}  {“Sinaasappel”,“Kers”,”Ananas”, “Mango”}  {“Kers”,”Ananas”,“Mango”}  {“Appel”,“Kers”}  {“Sinaasappel”,“Kers”,”Ananas”, “Mango”}  {“Appel”} |

1. Illustreer het Apriori-algoritme door te tonen hoe dit algoritme alle itemsets met een minimale support van 60% in deze dataset vindt.

**Opgave 3**

Bij de kassa van een supermarkt zijn de volgende boodschappenmandjes afgerekend:

|  |  |
| --- | --- |
| Tid | Items |
| 1 | {“I1”,“I2”,“I3”} |
| 2 | {“I1”,“I4”} |
| 3 | {“I2”,“I4”} |
| 4 | {“I1”,“I2”,“I3”} |
| 5 | {“I1”,“I4”} |
| 6 | {“I2”,“I4”} |
| 7 | {“I1”,“I2”,“I3”,“I4”} |
| 8 | {“I4”} |
| 9 | {“I1”,“I2”,“I3”} |

Bereken de associatieregels bij een minimale support van 50% en een minimale confidence van 60%.

**Uitwerkingen van schriftelijke opgaven**

|  |  |  |  |
| --- | --- | --- | --- |
| **1.a** | IT1∪ IT4={“Sinaasappel”,“Appel”,“Kers”, ”Ananas”, “Mango”}, } IT5 ∪ IT6={“Appel”,“Kers”, “Sinaasappel”,”Ananas”, “Mango” } IT6∩IT4={”Ananas”, “Mango”} IT6∩IT1={“Sinaasappel”,“Kers”} | 4p | 19p |
| **1.b** |  | 6p |
| **1.c** |  | 9p |

# 

|  |  |  |  |
| --- | --- | --- | --- |
| **2.a** |  | 6p | 14p |
| **2.b** |  | 8p |  |

# 

|  |  |  |  |
| --- | --- | --- | --- |
| **3.a** |  | 6p | 16p |
| **3.b** | **Iteration 1**:  C1={{ “a”}, { “b”}, { “c”}, { “d”}, { “e”}}    F1={{ “a”}, { “c”}, { “d”}, { “e”}}  **Iteration 2**:  C2={{ “a”,“c”}, {“a”,“d”}, {“a”,“e”}, {“c”,“d”}, {“c”,“e”}}      F2={{ }}  Geen 2-item gevonden | 5p  5p |

# 

|  |  |  |  |
| --- | --- | --- | --- |
| **4.a** |  | 2p | 18p |
| **4.b** |  | 2p |  |
| **4.c** | **Iteratie 1:**      **Iteratie 2:** | 6p  6p  2p |  |

# 

|  |  |  |  |
| --- | --- | --- | --- |
| **5.** | **Iteratie 1:**      **Iteratie 2:** | 6p  6p  2p | 14p |

# 

|  |  |  |  |
| --- | --- | --- | --- |
| **6.** | **Iteratie 1:**    **Iteratie 2:** | 6p  6p  2p | 14p |

**Uitwerkingen van praktische toetsopgaven(Geogebra)**

|  |  |  |  |
| --- | --- | --- | --- |
| **1.** | **Start: k=0**  Minimaal support = 60% | 2p | 12p |
| **Iteratie:** k=1  C1={{"A"},{"B"},{"C"},{"D"},{"F"},{"E"}}}    ⟶ F1={{"A"},{"B"},{"C"},{"D"}} | 4p |
| **Iteratie: k=2** C2={{"A","B"},{"A","C"},{"A","D"},{"B","C"},{"B","D"},{"C","D"}}    ⟶ F2={ {"A","D"},{"C","D"}} ⟶ C3={ {}}⟶Eind | 4p  2p |

|  |  |  |  |
| --- | --- | --- | --- |
| **2.** | **Start: k=0** Minimaal support = 50% | 2p | 14p |
|  | **Iteratie: k=1** C1={{"Appel"},{"Kers"},{"Sinaasappel"},{"Mango"},{"Ananas"}}    ⟶ F1={{{"Kers"},{"Sinaasappel"},{"Mango"},{"Ananas"}} | 4p |  |
|  | I**teratie: k=2** C2={{"Kers","Sinaasappel"},{"Kers","Mango"},{"Kers","Ananas"},{"Sinaasappel","Mango"},{"Sinaasappel","Ananas"}      ⟶ F2={ {"Ananas","Mango"}} ⟶ C3={ {}}⟶Eind | 4p  2p |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **3.** | **Start: k=0** Minimaal support = 50% minimale confidence = 60%. | 2p | 18p |
|  | **Iteratie: k=1** C1={{"I1"},{"I2"},{"I3"},{"I4"}}      ⟶ F1={{"I1"},{"I2"},{"I4"}} | 4p |  |
|  | **Iteratie: k=2** C2={{"I1","I2"},{"I1","I4"},{"I2", "I4"}}    ⟶ F2={{}} ⟶ Eind | 4p  2p |  |
|  | **Associatieregels:**  Confidence({"I1"}⇒{"I2"})=66.67%  Confidence({"I1"}⇒{"I4"})=50%  Confidence({"I2"}⇒{"I4"})=50%  Confidence({"I2"}⇒{"I1"})=66.67%  Confidence({"I4"}⇒{"I1"})=50%  Confidence({"I4"}⇒{"I2"})=50% | 6p |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **4.** | **Start: k=0**  Minimaal support = 50% minimale confidence = 60%. | 2p | 20p |
| I**teratie: k=1** C1={{"1"},{"2"},{"3"},{"4"},{"5"}}    ⟶ F1={{ {"2"},{"4"},{"5"}} | 4p |  |
| **Iteratie: k=2** C2={{"2","4"},{"2","5"},{"4", "5"}}    ⟶ F2={{"4","5"}} ⟶ F3={{}} Eind | 4p  2p |  |
| **Associatieregels:**  Confidence({"4"}⇒{"5"})=71,43% Confidence({"4","5"}⇒{"5"})=100%  Confidence({"5"}⇒{"4"})=71,43% Confidence({"4","5"}⇒{"4"})=100%  Confidence({"2" }⇒{"4"})=66,67% Confidence({"4","5"}⇒{"2"})=60% Confidence({"4"}⇒{"2"})=57,14% Confidence({{"2"}}⇒{"4","5"})=50% Confidence({"2"}⇒{"5"})=66,67% Confidence({{"4"}}⇒{"4","5"})=71,43% Confidence({"5"}⇒{"2"})=57.14% Confidence({{"5"}}⇒{"4","5"})=71,43% | 8p |  |

# Bijlage D - Uitwerking antwoorden 3.2

1.

a.

|  |  |
| --- | --- |
| **Uitwerking(Waarden)** | **Geogebra programma(Definities)** |
|  |  |

b.

|  |  |
| --- | --- |
| **Uitwerking(Waarden)** | **Geogebra programma(Definities)** |
|  |  |

c.

|  |  |
| --- | --- |
| **Uitwerking(Waarden)** | **Geogebra programma(Definities)** |
|  |  |

**2.**

|  |  |
| --- | --- |
| **Uitwerking(Waarden)** | **Geogebra programma(Definities)** |
|  |  |

a.

|  |  |
| --- | --- |
| **Uitwerking(Waarden)** | **Geogebra programma(Definities)** |
|  |  |

b.

|  |  |
| --- | --- |
| **Uitwerking(Waarden)** | **Geogebra programma(Definities)** |
|  |  |

c.

|  |  |
| --- | --- |
| **Uitwerking(Waarden)** | **Geogebra programma(Definities)** |
|  |  |

**3.**

a.

|  |  |
| --- | --- |
| **Uitwerking(Waarden)** | **Geogebra programma(Definities)** |
|  |  |

b.

|  |  |
| --- | --- |
| **Uitwerking(Waarden)** | **Geogebra programma(Definities)** |
|  |  |

c.

|  |  |
| --- | --- |
| **Uitwerking(Waarden)** | **Geogebra programma(Definities)** |
|  |  |

**4.**

**a**.

|  |  |
| --- | --- |
| **Uitwerking(Waarden)** | **Geogebra programma(Definities)** |
|  |  |

b.

|  |  |
| --- | --- |
| **Uitwerking(Waarden)** | **Geogebra programma(Definities)** |
|  |  |

c.

|  |  |
| --- | --- |
| **Uitwerking(Waarden)** | **Geogebra programma(Definities)** |
|  |  |

**5.**

|  |  |
| --- | --- |
| **Uitwerking(Waarden)** | **Geogebra programma(Definities)** |
|  |  |

a.

|  |  |
| --- | --- |
| **Uitwerking(Waarden)** | **Geogebra programma(Definities)** |
|  |  |

b.

|  |  |
| --- | --- |
| **Uitwerking(Waarden)** | **Geogebra programma(Definities)** |
|  |  |

|  |  |
| --- | --- |
| **Uitwerking(Waarden)** | **Geogebra programma(Definities)** |
|  |  |

**6.**

|  |  |
| --- | --- |
| **Start: k=0**  Minimaal support = 50% | |
|  | |

|  |  |
| --- | --- |
| **Iteratie: k=1**  C1={{"a"}, {"ba"}, {"bi"}, {"c"}, {"m"}, {"r"}, {"p"}, {"s"}} | |
|  | |
| De 1-items {"a"}, {"ba"}, {"bi"}, {"p"} en {"s"} zijn niet frequent ⟶F1={{"c"}, {"m"}, {"r"}} | |

|  |  |
| --- | --- |
| **Iteratie: k=2**  C2={{"c", "m"}, {"c", "r"}, {"m", "r"}} | |
|  | |
| De 2-items {"c", "m"} en {"m", "r"}} zijn niet frequent.  Dus: F2={{"c", "m"}} ⟶ C3={{}} } ⟶ Eind | |

**7.**

|  |  |
| --- | --- |
| **Start: k=0**  Minimaal support = 40% Minimaal confidence= 70% | |
|  | |

|  |  |
| --- | --- |
| **Iteratie: k=1**  C1={{"a"}, {"b"}, {"c"}, {"d"},{"e"}} | |

|  |  |
| --- | --- |
|  | |
| De 1-item {**e}** is niet frequent ⟶F1={{"a"}, {"b"}, {"c"}, {"d"}} | |

|  |  |
| --- | --- |
| **Iteratie: k=2**  C2={{"a", "b"}, {"a", "c"}, {"a", "d"}, {"b", "c"}, {"b", "d"}, {"c", "d"}} | |
|  | |
| De 2-item {**a,d}** is niet frequent ⟶F2={{"a","b"}, {"a","c"}, {"b","c"}, {"b","d"},{"c","d"}} | |

|  |  |
| --- | --- |
| **Iteratie: k=3**  C3={{"a", "b","c"}, {"b", "c", "d"}} | |
|  | |
| De 3-item {**b,c,d}** is niet frequent ⟶F3={{"a","b","c"}⟶C4={{}}}⟶Eind | |

|  |  |
| --- | --- |
| **Associtiesregels:** | **Geogebra applet:** |
| Confidence({a,c}⇒{b})=100%  Confidence({a}⇒{b})=75% Confidence({a}⇒{b})=100%  Confidence({c}⇒{b})=75% Confidence({b}⇒{a})=100% |  |

**8.**

|  |
| --- |
| **Start: k=0**  Minimaal support = 50% Minimaal confidence= 60% |
|  |

|  |
| --- |
| **Iteratie: k=1**  C1={{Brood},{Melk}, {Bier},{Eieren},{Luier},{Cola}} |
|  |
| **De 1-items {Eieren} en {Cola} zijn niet frequent⟶ F1={{Brood},{Melk}, {Bier},{Luier}}** |

|  |
| --- |
| **Iteratie: k=2**  **C2={{Brood,Melk}, {Brood,Bier},{Brood,Luier},{Melk,Luier},{Melk, Bier}**  **,{Bier,Luier}}** |
|  |
| De 2-items {**Bier,Brood},** {**Bier, Melk}** zijn niet frequent  F2= {{Brood, Melk}, {Brood,Luier},{Melk,Luier},{Bier,Luier}} |

|  |
| --- |
| **Iteratie: k=3**  C3= {{Brood,Melk,Luier}} |
|  |
| De 3-item {**{"Brood", "Melk", "Luier"}}** is niet frequent  **F3= {{}}** |

|  |  |
| --- | --- |
| **Associtiesregels:** | **Geogebra applet:** |
| Confidence({Bier}⇒{Luier,Melk})=66.7%  Confidence({Melk}⇒{Brood})=75% Confidence({Bier}⇒{Luier})=100% |  |

9.

|  |
| --- |
| **Start: k=0**  Minimaal support = 40% Minimaal confidence= 80% |
|  |

|  |
| --- |
| **Iteratie: k=1**  C1={{a},{b}, {c},{d},{e},{f},{g}} |
| **I** |
| **De 1-items {c}, {e} en {f} zijn niet frequent ⟶** F1={{**a},{b}, {c},**{**d},{e},**{**f},**{**g}}** |

|  |
| --- |
| **Iteratie: k=2**  **C2={{a,b}, {a,d},{a,g},{b,d},{b,g},{d,g}}** |
|  |
| De 2-items {**a,b},** {**a, g},** {**b, g} en** {**e, g}** zijn niet frequent**⟶ F2={{a,d},{b,d}} ⟶ C 3 is leeg⟶ Eind** |

|  |  |
| --- | --- |
| **Associtiesregels:** | **Geogebra applet:** |
| Confidence({a}⇒{d})=100%  Confidence({b}⇒{d})=80% |  |