



Voorbeeldexamen

Editie 202310

Copyright © EXIN Holding B.V. 2023. All rights reserved.
EXIN® is a registered trademark.

No part of this publication may be reproduced, stored, utilized or transmitted in any form or by any means, electronic, mechanical, or otherwise, without the prior written permission from EXIN.

Inhoud

Inleiding	4
Voorbeeldexamen	5
Antwoordsleutel	15
Evaluatie	34

Inleiding

Dit is het EXIN Data Analytics Foundation (DAF.NL) voorbeeldexamen. Op dit examen is het Reglement voor de Examens van EXIN van toepassing.

Dit examen bestaat uit 40 meerkeuzevragen. Elke vraag heeft een aantal antwoorden, waarvan er één correct is.

Het maximaal aantal te behalen punten is 40. Elke goed beantwoorde vraag levert u 1 punt op. U hebt minimaal 26 punten nodig om te slagen.

De beschikbare tijd is 60 minuten.

Veel succes!

Voorbeeldexamen

1 / 40

Data analytics is de wetenschap van het analyseren van data met als doel hier conclusies uit te trekken.

Waarop is data analytics **niet** gericht?

- A) Het omzetten van data in informatie
- B) Het aanmaken van data
- C) De presentatie van data
- D) De verwerking van data

2 / 40

Wat is een voorbeeld van een stap in data-opschoning?

- A) Een bedrijf verzamelt de omzetcijfers van alle filialen en wijzigt de monetaire eenheid in 'duizend euro' voordat de datasets wordt samengevoegd in het datawarehouse van het bedrijf.
- B) Een bedrijf verzamelt de omzetcijfers van alle filialen en bouwt een datamodel om per geografische locatie de omzet voor het komende jaar te voorspellen.
- C) Een bedrijf verzamelt de omzetcijfers van alle filialen en verwijdert duplicaten voordat de datasets worden samengevoegd in het datawarehouse van het bedrijf.
- D) Een bedrijf verzamelt de omzetcijfers van alle filialen, bepaalt de mediaanwaarde en stelt vast welk filiaal een omzet heeft die het verst van de mediaan is verwijderd.

3 / 40

Naleving van wetgeving is een belangrijk aandachtspunt voor organisaties om risico's te mijden.

Wat is bij data **geen** belangrijk aandachtspunt voor de naleving van relevante wetgeving?

- A) Data analytics
- B) Data-eigenaarschap
- C) Dataprivacy
- D) Intellectueel eigendom

4 / 40

Wat is **geen** kenmerk van interactieve dashboards?

- A) Ze stellen gebruikers in staat om in te zoomen op de details van data.
- B) Ze kunnen data uit meerdere bronnen combineren.
- C) Ze kunnen lastig te gebruiken zijn voor niet-technische gebruikers.
- D) Ze verstrekken wat-als-analyses aan het management.

5 / 40

Een organisatie wil tekst van veel verschillende webpagina's verzamelen.

Wat is de **beste** manier om openbare data te verkrijgen?

- A) Via alternatieve data
- B) Via interne verwervingssystemen
- C) Via interviews en experimenten
- D) Via web-schrapen

6 / 40

Wat is een kanaal om data te verzamelen?

- A) Continue data
- B) Intentie-abstractie
- C) Interne verwervingssystemen
- D) Sleutelwaarde-opslag
- E) Relational database management system (RDMS)
- F) Gestructureerde data

7 / 40

Wat is een voorbeeld van web-schrapen?

- A) Een CSV-bestand downloaden van een overheidswebsite
- B) Data downloaden door de inhoud van een website te parseren
- C) Data downloaden via een file transfer protocol verbinding (FTP-verbinding)
- D) Op internet gevonden Excel-bestanden downloaden

8 / 40

Wat is bij het verzamelen van data van essentieel belang om datawetten na te leven?

- A) Dat transacties worden voltooid met open data
- B) Dat transparantie bij het verzamelen van data wordt gewaarborgd
- C) Dat er minimale hoeveelheden openbare data worden gemined
- D) Dat er persoonlijke data wordt verworven

9 / 40

Welk type data wordt gekenmerkt door volume, snelheid en verscheidenheid?

- A) Alternatieve data
- B) Big data
- C) Gestructureerde data
- D) Ongestructureerde data

10 / 40

Welke dataoplossing heeft post-analytische besluitvorming als uitdrukkelijk doel?

- A) Enterprise data warehouse (EDW)
- B) Sleutelwaarde-opslag
- C) Relational database management system (RDMS)
- D) Ongestructureerde datasystemen

11 / 40

Wat is een verschil tussen de manier waarop een enterprise data warehouse (EDW) en een relational database management system (RDMS) worden gebruikt?

- A) Een EDW is een oplossing voor dagelijkse werkzaamheden en transacties, terwijl een RDMS wordt geraadpleegd voor vooraf gedefinieerde rapporten en dashboards.
- B) Een EDW maakt gebruik van online analytical processing (OLAP), terwijl een RDMS gebruikmaakt van online transaction processing (OLTP).
- C) Een RDMS is een tool die wordt toegepast op complexe rapporten en analyses, terwijl een EDW wordt toegepast op snelle transacties in real time.
- D) Een RDMS wordt gebaseerd op een model voor sleutelwaarde-opslag, terwijl een EDW wordt gebaseerd op een model voor objectopslag.

12 / 40

Wat is **geen** kenmerk van gedistribueerde bestandssystemen?

- A) Dat opslagruimte en servers tegen relatief lage kosten kunnen worden toegevoegd
- B) Dat data-chunks op meerdere servers tegelijk worden verwerkt
- C) Dat data makkelijk kan worden opgeslagen als blob zonder schemadefinitie
- D) Dat data zonder gerichte toepassing in de oorspronkelijke indeling worden opgeslagen

13 / 40

Een autoleasebedrijf gebruikt een cloudoplossing voor alle kernactiviteiten. Medewerkers kunnen hierin offertes opstellen, voertuigen aankopen, onderhoud inplannen en kredietcontroles uitvoeren.

Het bedrijf wil daarnaast een cloudoplossing van dezelfde cloudprovider gaan gebruiken voor data analytics-doeleinden.

Wat is **geen** voordeel van de beoogde data analytics-oplossing?

- A) De cloudoplossingen zijn goedkoper dan wanneer het bedrijf eigenaar zou zijn van de benodigde omgeving met hardware, software en IT-medewerkers.
- B) Het is niet nodig om enorme datasets naar verschillende servers te importeren of exporteren.
- C) De oplossing kan worden opgeschaald in hetzelfde tempo als de activiteiten van het autoleasebedrijf.
- D) De oplossing kan worden afgeschaald wanneer data analytics niet meer nodig is.

14 / 40

Wat is het verband tussen een onafhankelijke variabele en een afhankelijke variabele?

- A) Een afhankelijke variabele wordt veranderd door een datawetenschapper en een onafhankelijke variabele wordt beïnvloed door deze verandering.
- B) De waarden van een afhankelijke variabele weerspiegelen de effecten die bij wijziging van de onafhankelijke variabele worden waargenomen en vastgelegd.
- C) Een onafhankelijke variabele wordt uitgedrukt als een kleine letter 'y' en een afhankelijke variabele als een hoofdletter 'X'.
- D) Een onafhankelijke variabele moet altijd worden beschreven als een categorische variabele en een afhankelijke variabele kan elke vorm aannemen.

15 / 40

Welk type variabele is een categorische variabele?

- A) Booleaanse variabelen
- B) Discrete variabelen
- C) Nominale variabelen
- D) Numerieke variabelen

16 / 40

Welk type variabele is de kilometerstand van een voertuig aan het begin en het einde van een jaar?

- A) Booleaans
- B) Categorisch
- C) Numeriek
- D) Datum/tijd

17 / 40

Waarom is het belangrijk om onderscheid te maken tussen discrete en continue variabelen?

- A) Omdat de geschikte algoritmen voor de variabelen dan beter kunnen worden geselecteerd
- B) Omdat dit belangrijk is om het conceptuele datamodel te definiëren
- C) Omdat alleen discrete variabelen met andere variabelen kunnen worden samengevoegd
- D) Omdat in relationele databases alleen continue variabelen kunnen worden opgeslagen

18 / 40

Welke techniek voor het scrubben van data houdt in dat variabelen worden getransformeerd naar een integer formaat?

- A) One-hot codering
- B) Samenvoeging van variabelen
- C) Selectie van variabelen
- D) Web-schrapen

19 / 40

Een autoleasebedrijf baseert zijn data analytics op een dataset met kolommen als contractnummer, kenteken, toelatingsdatum, merk, model, kleur en einddatum contract.

De kolommen 'toelatingsdatum' en 'einddatum contract' worden teruggebracht tot een integer met zes posities, waarbij de eerste vier posities overeenkomen met het jaar en de laatste twee met de maand. Zo kunnen de contracten worden onderverdeeld in categorieën met alle contracten die in een bepaalde maand aflopen.

Waarvan is dit een voorbeeld?

- A) Binning
- B) Samenvoeging van variabelen
- C) One-hot codering
- D) Selectie van variabelen

20 / 40

Bij data-scrubbing is het belangrijk om de gebruikte oorspronkelijke brondata te behouden, zodat eventuele revisies mogelijk blijven.

Wat moet er worden gebruikt om dit te waarborgen?

- A) Een beleid voor dataretentie
- B) Een datawarehouse
- C) Een sleutelwaarde-opslag
- D) Kunstmatige intelligentie (KI) algoritmen

21 / 40

Wat is een kenmerk van inferentiële (toetsende) methoden?

- A) Ze hebben de voorkeur voor situaties waarin data goed is gedocumenteerd en tot één informatiepool is gestandaardiseerd.
- B) Ze helpen complexe informatie te comprimeren tot een handige, makkelijk leesbare indeling door duidelijke trends samen te vatten.
- C) Ze draaien om de isolatie en analyse van een deel van de data, waarna de resultaten worden getest ten opzichte van een andere subset van de data.
- D) Ze bieden een handige en beknopte samenvatting van historische data die is gegenereerd uit een potentieel enorm aantal op zichzelf staande gebeurtenissen.

22 / 40

In welk geval moet beschrijvende analyse worden gebruikt?

- A) Wanneer een dataset niet makkelijk kan worden geanalyseerd
- B) Wanneer er beperkte informatie beschikbaar is
- C) Wanneer de opgehaalde data erg gedetailleerd is
- D) Wanneer er sprake is van hiaten in de administratie

23 / 40

Wat heeft een menselijke operator nodig om aan de slag te gaan met datamining?

- A) Een gedefinieerde hypothese zoals 'waarschijnlijk' en 'onwaarschijnlijk'
- B) Een groot aantal mogelijke inputcombinaties
- C) Een model dat aangeeft welke input een bepaalde output oplevert
- D) Een probleem waarvan wordt aangenomen dat het de moeite waard is om het op te lossen

24 / 40

Bij welke machine learning-methode worden trainingsdata en testdata gebruikt?

- A) Datamining-methode
- B) Beschrijvende methode
- C) Inferentiële (toetsende) methode
- D) Opgesplitste validatie-methode

25 / 40

Het verkoopproces van een autoleasebedrijf heeft een grote dataset opgeleverd. Het bedrijf gebruikt machine learning om de puntjes op de i te zetten bij het opstellen van offertes. Alleen de data uit het verkoopproces wordt ingevoerd in de software, zonder aanvullende regels. Vervolgens zoekt de software naar patronen en retourneert een model om zonder al te veel korting de offerteprijs te berekenen.

Welk type machine learning is dit?

- A) Reinforcement learning
- B) Supervised learning
- C) Unsupervised learning

26 / 40

Een belangrijk aspect van natuurlijke taalverwerking (NTV) is syntactische analyse.

Wat wordt er bij syntactische analyse geanalyseerd?

- A) De betekenis van een zin
- B) De benoemde entiteiten
- C) De tekst van zoekopdrachten
- D) De structuur van een zin

27 / 40

Een van de taken bij natuurlijke taalverwerking (NTV) omvat het herkennen van belangrijke entiteiten in tekst, zoals een naam, locatie of activiteit.

Welke NTV-taak is dit?

- A) Klasse-identificatie
- B) Herkenning van benoemde entiteiten
- C) Sentimentanalyse
- D) Stemming

28 / 40

Wat is het verband tussen data en algoritmen?

- A) Algoritmen worden ontwikkeld om data te verwerken en analyseren.
- B) Data wordt ontwikkeld om te worden omgezet naar algoritmische kennis.
- C) Verschillende mensen gebruiken hetzelfde algoritme voor verwerking van dezelfde data.
- D) Inputdata verandert nooit voor algoritmen, wat resulteert in uniciteit.

29 / 40

Uit data van een makelaar blijkt dat de gemiddelde prijs van verkochte huizen afneemt in lijn met een stijgende hypotheekrente.

Welke soort regressieanalyse moet worden gebruikt om dit patroon in de data te beschrijven?

- A) Exponentiële regressieanalyse
- B) Lineaire regressieanalyse
- C) Logistische regressieanalyse
- D) Niet-lineaire regressieanalyse

30 / 40

Wat is het doel van regressieanalyse?

- A) De prestaties van een organisatie analyseren als basis voor besluitvorming
- B) Datapunten categoriseren in groeperingen wanneer er geen voorgedefinieerde klassen bestaan
- C) Een lijn of curve vinden die patronen in de data zo goed mogelijk beschrijft
- D) Begrijpen hoe data wordt verzameld, geordend en geïnterpreteerd

31 / 40

Welk type model genereert categorische voorspellingen en clustering?

- A) Algoritmisch model
- B) Classificatiemodel
- C) Regressiemodel

32 / 40

Een bedrijf beschikt over een grote dataset voor de verkoop van overhemden. Het personeel wil de cijfers beter begrijpen en de data analist vraagt zich af of clusteranalyse wellicht de gewenste inzichten oplevert.

Hoe zorgt k-means clustering voor een beter begrip van de verkoopcijfers?

- A) Door data te clusteren in voorgedefinieerde groeperingen
- B) Door onverwachte relaties tussen clusters te beschrijven
- C) Door de rol van zwaartepunten te verklaren
- D) Door een nieuwe manier te bieden om klanten te groeperen

33 / 40

Wat is een **belangrijk** verschil tussen associatie analyse en sequentiële mining?

- A) Bij associatie analyse wordt de gegeneraliseerde sequentiepatronen (GSP) techniek toegepast, terwijl sequentiële mining a priori wordt toegepast.
- B) Bij associatie analyse wordt niet gekeken naar de volgorde waarin items voorkomen, terwijl hier bij sequentiële mining rekening mee wordt gehouden.
- C) Bij associatie analyse wordt een betrouwbaarheidsniveau gebruikt ter ondersteuning van de besluitvorming, terwijl sequentiële mining een vorm van supervised learning is.
- D) Associatie analyse werkt beter bij grote datasets, terwijl sequentiële mining het meest geschikt is voor kleine datasets.

34 / 40

Sequentiële mining gebruikt een betrouwbaarheidsniveau om beslissingen te onderbouwen. Het gebruikt ook een soort criterium om te focussen op frequente patronen om deze besluitvorming te onderbouwen.

Welk criterium is dit?

- A) A priori-algoritme
- B) Frequentie itemset
- C) Minimum support
- D) Recursieve eliminatie

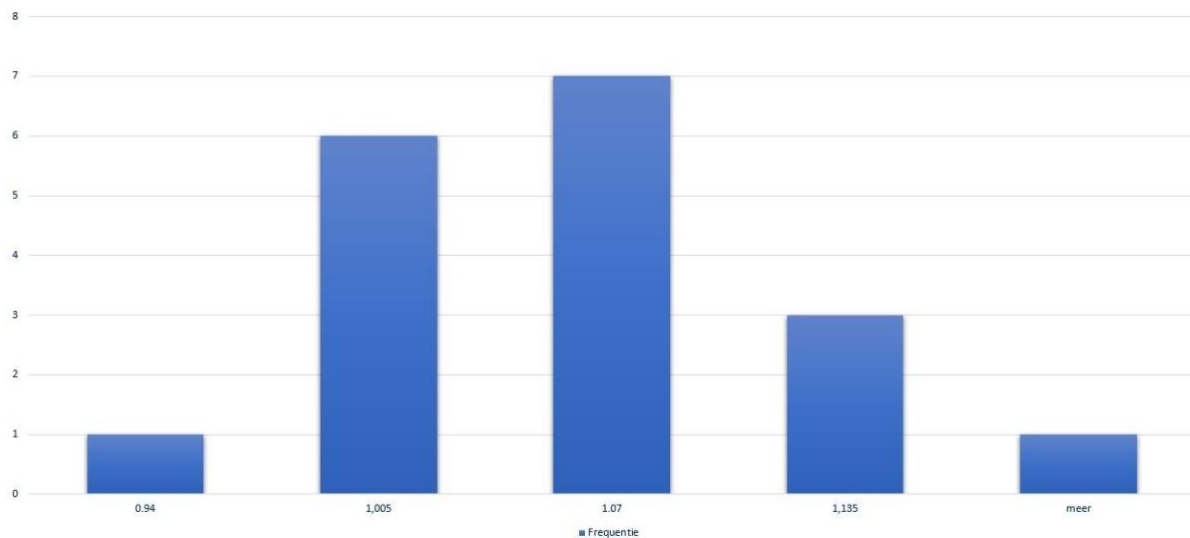
35 / 40

Welke grafische techniek wordt doorgaans gebruikt om data te presenteren aan een extern publiek?

- A) Expansief
- B) Verklarend
- C) Uiteenzettend
- D) Verkennend

36 / 40

Bekijk de onderstaande afbeelding:



Waar is dit een voorbeeld van?

- A) Staafdiagram
- B) Box plot
- C) Heatmap
- D) Histogram

37 / 40

Welk type plot wordt gebruikt om de spreiding en scheefheid van een dataset weer te geven?

- A) Box plot
- B) Rug plot
- C) Scatter plot
- D) Violin plot

38 / 40

Om meer inzicht te krijgen in de prijzen van verkochte huizen in verschillende wijken van een stad wil een makelaar een heatmap gebruiken.

Wat toont deze heatmap?

- A) Een stadsplattegrond waarop de niveaus van de huizenprijzen met kleuren worden aangeduid
- B) Een stadsplattegrond met verkochte huizen en de verkoopprijs per huis
- C) Een tabel met per kwartaal de gemiddelde huisprijzen voor het afgelopen jaar, gesorteerd op de gemiddelde prijs
- D) Een tabel met de prijzen van verkochte huizen het afgelopen jaar, geordend op prijs

39 / 40

Waarom is een esthetisch ontwerp belangrijk?

- A) Omdat daarmee de technieken voor machine learning worden uitgebreid
- B) Omdat daarmee natuurlijke taalverwerking (NTV) mogelijk wordt
- C) Omdat het is gericht op modellen voor regressieanalyse
- D) Omdat daarmee de bruikbaarheid van datavisualisaties toeneemt

40 / 40

Welke tool voor datavisualisatie heeft het uiterlijk van Microsoft Excel?

- A) Data Wrapper
- B) Google Charts
- C) Power BI
- D) Tableau

Antwoordsleutel

1 / 40

Data analytics is de wetenschap van het analyseren van data met als doel hier conclusies uit te trekken.

Waarop is data analytics **niet** gericht?

- A) Het omzetten van data in informatie
 - B) Het aanmaken van data
 - C) De presentatie van data
 - D) De verwerking van data
- A) Incorrect. Hierop zijn de kernactiviteiten van data analytics gericht: de analyse en verwerking van data om deze te doorgronden.
- B) Correct. Het aanmaken van data gaat vooraf aan de behoefte om deze te doorgronden en is daarom geen onderdeel van data analytics. (Literatuur: B, Hoofdstuk 2 en 3)
- C) Incorrect. Data analytics is gericht op de presentatie van data om te zorgen dat de resultaten van de analyse geschikt zijn voor een beoogde doelgroep.
- D) Incorrect. De manier waarop data wordt verwerkt en geanalyseerd om deze te kunnen doorgronden, is een centraal thema binnen data analytics.

2 / 40

Wat is een voorbeeld van een stap in data-opschoning?

- A) Een bedrijf verzamelt de omzetcijfers van alle filialen en wijzigt de monetaire eenheid in 'duizend euro' voordat de datasets wordt samengevoegd in het datawarehouse van het bedrijf.
 - B) Een bedrijf verzamelt de omzetcijfers van alle filialen en bouwt een datamodel om per geografische locatie de omzet voor het komende jaar te voorspellen.
 - C) Een bedrijf verzamelt de omzetcijfers van alle filialen en verwijdert duplicaten voordat de datasets worden samengevoegd in het datawarehouse van het bedrijf.
 - D) Een bedrijf verzamelt de omzetcijfers van alle filialen, bepaalt de mediaanwaarde en stelt vast welk filiaal een omzet heeft die het verst van de mediaan is verwijderd.
- A) Incorrect. Dit is geen data-opschoning, maar datanormalisatie/-standaardisatie.
- B) Incorrect. Dit is geen data-opschoning, maar voorspellende data-analyse.
- C) Correct. Het verwijderen van duplicaten is een typische stap in data-opschoning om de uniciteit van de data te waarborgen. (Literatuur: B, Hoofdstuk 4)
- D) Incorrect. Dit is geen data-opschoning, maar dataprofilering. Dataprofilering heeft betrekking op de beoordeling van data met behulp van statistische functies voordat er technieken voor machine learning op toegepast worden.

3 / 40

Naleving van wetgeving is een belangrijk aandachtspunt voor organisaties om risico's te mijden.

Wat is bij data **geen** belangrijk aandachtspunt voor de naleving van relevante wetgeving?

- A) Data analytics
 - B) Data-eigenaarschap
 - C) Dataprivacy
 - D) Intellectueel eigendom
- A) Correct. In wetgeving wordt niet beschreven hoe data wordt geanalyseerd, maar juist welke data mag worden verzameld en geanalyseerd en voor welk doeleinde. (Literatuur: B, Hoofdstuk 7)
- B) Incorrect. Steeds meer landen voeren momenteel regelgeving in om vast te leggen wie de eigenaar van verzamelde data is en wat de rechten van die eigenaar zijn.
- C) Incorrect. Er bestaan momenteel internationale verordeningen voor dataprivacy en die vormen een aandachtspunt voor organisaties.
- D) Incorrect. Wetgeving met betrekking tot intellectueel eigendom is een van de eerste wetgevingen die van invloed is op data analytics.

4 / 40

Wat is **geen** kenmerk van interactieve dashboards?

- A) Ze stellen gebruikers in staat om in te zoomen op de details van data.
 - B) Ze kunnen data uit meerdere bronnen combineren.
 - C) Ze kunnen lastig te gebruiken zijn voor niet-technische gebruikers.
 - D) Ze verstrekken wat-als-analyses aan het management.
- A) Incorrect. Interactieve dashboards stellen gebruikers in staat om in te zoomen op meer gedetailleerde data.
- B) Incorrect. Interactieve dashboards combineren data uit meerdere bronnen om een uitgebreid inzicht te bieden.
- C) Correct. Interactieve dashboards worden zodanig ontworpen dat ze makkelijk te gebruiken zijn door niet-technische gebruikers. (Literatuur: B, Hoofdstuk 5)
- D) Incorrect. Interactieve dashboards bieden wat-als-inzichten ter ondersteuning van de besluitvorming.

5 / 40

Een organisatie wil tekst van veel verschillende webpagina's verzamelen.

Wat is de **beste** manier om openbare data te verkrijgen?

- A) Via alternatieve data
 - B) Via interne verwervingssystemen
 - C) Via interviews en experimenten
 - D) Via web-schrapen
- A) Incorrect. Met alternatieve data wordt informatie van niet-traditionele bronnen verzameld, meestal in verband met bankzaken en financiën. Dit is daarom niet de beste manier.
- B) Incorrect. Het verkrijgen van data uit interne verwervingssystemen is de meest gangbare manier om interne data te verkrijgen.
- C) Incorrect. Interviews en experimenten zijn niet geschikt voor automatisering en code. Bovendien is uit interviews verkregen informatie geen openbare data.
- D) Correct. Met web-schrapen wordt informatie van het web verzameld met behulp van code en automatisering. (Literatuur: A, Hoofdstuk 1)

6 / 40

Wat is een kanaal om data te verzamelen?

- A) Continue data
 - B) Intentie-abstractie
 - C) Interne verwervingssystemen
 - D) Sleutelwaarde-opslag
 - E) Relational database management system (RDMS)
 - F) Gestructureerde data
- A) Incorrect. Continue data is een type variabele, geen kanaal om data te verzamelen.
- B) Incorrect. Intentie-abstractie is een onderdeel van natuurlijke taalverwerking (NTV), geen kanaal om data te verzamelen.
- C) Correct. Interne verwervingssystemen zijn een kanaal om data te verzamelen. (Literatuur: A, Hoofdstuk 1)
- D) Incorrect. Sleutelwaarde-opslag is een opslagsysteem voor data, geen kanaal om data te verzamelen.
- E) Incorrect. RDMS is een softwareprogramma dat wordt gebruikt om gestructureerde data op te slaan. Het is geen kanaal om data te verzamelen.
- F) Incorrect. Gestructureerde data is geen kanaal om data te verzamelen, maar een indeling om data op te slaan.

7 / 40

Wat is een voorbeeld van web-schrapen?

- A) Een CSV-bestand downloaden van een overheidswebsite
 - B) Data downloaden door de inhoud van een website te parseren
 - C) Data downloaden via een file transfer protocol verbinding (FTP-verbinding)
 - D) Op internet gevonden Excel-bestanden downloaden
- A) Incorrect. Bij web-schrapen wordt waardevolle data geëxtraheerd door data te combineren. Dat is iets anders dan het downloaden van een enkel bestand.
- B) Correct. Bij web-schrapen worden websites doorzocht en wordt waardevolle data geëxtraheerd. (Literatuur: A, Hoofdstuk 1)
- C) Incorrect. Dit heeft geen betrekking op het web-gedeelte van web-schrapen.
- D) Incorrect. Hoewel de bestanden zijn gevonden op internet, is hier geen sprake van schrapen.

8 / 40

Wat is bij het verzamelen van data van essentieel belang om datawetten na te leven?

- A) Dat transacties worden voltooid met open data
 - B) Dat transparantie bij het verzamelen van data wordt gewaarborgd
 - C) Dat er minimale hoeveelheden openbare data worden gemined
 - D) Dat er persoonlijke data wordt verworven
- A) Incorrect. Open data kan persoonsgegevens bevatten. Als dat het geval is, voldoet de data niet noodzakelijkerwijs aan datawetten.
- B) Correct. Om te voldoen aan actuele datawetten is transparantie over de manier waarop data wordt verzameld van essentieel belang. (Literatuur: A, Hoofdstuk 1)
- C) Incorrect. In datawetten staat niet dat mining van openbare gegevens van essentieel belang is voor de naleving.
- D) Incorrect. Persoonlijke data mogen alleen worden verworven conform de wet. Verwerving is dus niet van essentieel belang voor de naleving.

9 / 40

Welk type data wordt gekenmerkt door volume, snelheid en verscheidenheid?

- A) Alternatieve data
 - B) Big data
 - C) Gestructureerde data
 - D) Ongestructureerde data
- A) Incorrect. Alternatieve data is een kanaal voor dataverzameling.
- B) Correct. Big data wordt gekenmerkt door volume, snelheid en verscheidenheid. (Literatuur: A, Hoofdstuk 2)
- C) Incorrect. Gestructureerde data beschrijft informatie met een goed-gedefinieerde indeling waaruit algoritmen makkelijk cruciale informatie kunnen ophalen en verwerken.
- D) Incorrect. Ongestructureerde data bestaat uit informatie zonder herkenbare indeling die niet in een standaardstructuur past, bijvoorbeeld een dataset in tabelvorm.

10 / 40

Welke dataoplossing heeft post-analytische besluitvorming als uitdrukkelijk doel?

- A) Enterprise data warehouse (EDW)
 - B) Sleutelwaarde-opslag
 - C) Relational database management system (RDMS)
 - D) Ongestructureerde datasystemen
- A) Correct. Een EDW heeft post-analytische besluitvorming in de organisatie als uitdrukkelijk doel. (Literatuur: A, Hoofdstuk 2)
- B) Incorrect. In een sleutelwaarde-opslag kan data zonder schemadefinitie worden opgeslagen, als blob. Hierdoor is de waarde van de data vaag en kan het resultaat van een verzoek niet worden beheerst.
- C) Incorrect. Een RDMS is gericht op dataopslag en wordt niet gebruikt voor post-analytische besluitvorming.
- D) Incorrect. Ongestructureerde datasystemen worden voornamelijk gebruikt om ongestructureerde data op te slaan en te verwerken. Ze worden niet gebruikt voor post-analytische besluitvorming.

11 / 40

Wat is een verschil tussen de manier waarop een enterprise data warehouse (EDW) en een relational database management system (RDMS) worden gebruikt?

- A) Een EDW is een oplossing voor dagelijkse werkzaamheden en transacties, terwijl een RDMS wordt geraadpleegd voor vooraf gedefinieerde rapporten en dashboards.
 - B) Een EDW maakt gebruik van online analytical processing (OLAP), terwijl een RDMS gebruikmaakt van online transaction processing (OLTP).
 - C) Een RDMS is een tool die wordt toegepast op complexe rapporten en analyses, terwijl een EDW wordt toegepast op snelle transacties in real time.
 - D) Een RDMS wordt gebaseerd op een model voor sleutelwaarde-opslag, terwijl een EDW wordt gebaseerd op een model voor objectopslag.
- A) Incorrect. Een EDW is geen oplossing voor dagelijkse werkzaamheden, omdat het data van verschillende bronnen moet consolideren en integreren. Een EDW is gericht op analytische toepassingen. Een RDMS is een tool voor het beheer van transactiedata.
- B) Correct. Een EDW wordt gebaseerd op OLAP en een RDMS op OLTP. (Literatuur: A, Hoofdstuk 2)
- C) Incorrect. Een RDMS wordt gebruikt voor snelle transacties in real time en een EDW voor complexe rapporten en analyses.
- D) Incorrect. Een RDMS wordt niet gebaseerd op een model voor sleutelwaarde-opslag, maar op een relationeel model.

12 / 40

Wat is **geen** kenmerk van gedistribueerde bestandssystemen?

- A) Dat opslagruimte en servers tegen relatief lage kosten kunnen worden toegevoegd
 - B) Dat data-chunks op meerdere servers tegelijk worden verwerkt
 - C) Dat data makkelijk kan worden opgeslagen als blob zonder schemadefinitie
 - D) Dat data zonder gerichte toepassing in de oorspronkelijke indeling worden opgeslagen
-
- A) Incorrect. De kosten van incrementele opslag en servers zijn relatief lager bij gedistribueerde bestandssystemen.
 - B) Incorrect. In gedistribueerde bestandssystemen kunnen data-chunks op verschillende servers tegelijk worden verwerkt door taken op te splitsen binnen het netwerk.
 - C) Correct. Data kan in een sleutelwaarde-opslag worden opgeslagen als blob zonder schemadefinitie. Dit is geen kenmerk van gedistribueerde bestandssystemen. (Literatuur: A, Hoofdstuk 2)
 - D) Incorrect. Gedistribueerde bestandssystemen maken het mogelijk om data op te slaan die in het oorspronkelijke formaat geen expliciete functie hebben.

13 / 40

Een autoleasebedrijf gebruikt een cloudoplossing voor alle kernactiviteiten. Medewerkers kunnen hierin offertes opstellen, voertuigen aankopen, onderhoud inplannen en kredietcontroles uitvoeren.

Het bedrijf wil daarnaast een cloudoplossing van dezelfde cloudprovider gaan gebruiken voor data analytics-doeleinden.

Wat is **geen** voordeel van de beoogde data analytics-oplossing?

- A) De cloudoplossingen zijn goedkoper dan wanneer het bedrijf eigenaar zou zijn van de benodigde omgeving met hardware, software en IT-medewerkers.
 - B) Het is niet nodig om enorme datasets naar verschillende servers te importeren of exporteren.
 - C) De oplossing kan worden opgeschaald in hetzelfde tempo als de activiteiten van het autoleasebedrijf.
 - D) De oplossing kan worden afgeschaald wanneer data analytics niet meer nodig is.
-
- A) Correct. Dit kan niet als algemene regel worden gesteld. De cloudoplossing zou zelfs duurder kunnen zijn. (Literatuur: A, Hoofdstuk 2)
 - B) Incorrect. Alle data staat al in de cloud, klaar voor gebruik.
 - C) Incorrect. Cloudoplossingen kunnen doorgaans worden opgeschaald zonder enige inspanning of investering van de klant.
 - D) Incorrect. Op- en afschalen zijn meestal voordelen van een cloudoplossing.

14 / 40

Wat is het verband tussen een onafhankelijke variabele en een afhankelijke variabele?

- A) Een afhankelijke variabele wordt veranderd door een datawetenschapper en een onafhankelijke variabele wordt beïnvloed door deze verandering.
 - B) De waarden van een afhankelijke variabele weerspiegelen de effecten die bij wijziging van de onafhankelijke variabele worden waargenomen en vastgelegd.
 - C) Een onafhankelijke variabele wordt uitgedrukt als een kleine letter 'y' en een afhankelijke variabele als een hoofdletter 'X'.
 - D) Een onafhankelijke variabele moet altijd worden beschreven als een categorische variabele en een afhankelijke variabele kan elke vorm aannemen.
-
- A) Incorrect. De onafhankelijke variabele is het onderdeel dat wordt gewijzigd. Deze wijziging is van invloed op de afhankelijke variabele.
 - B) Correct. Wanneer de onafhankelijke variabele wordt gewijzigd, heeft dit effect op de afhankelijke variabele. (Literatuur: A, Hoofdstuk 2)
 - C) Incorrect. De onafhankelijke variabele wordt in een vergelijking uitgedrukt als 'X'.
 - D) Incorrect. Een onafhankelijke variabele kan ook worden beschreven als een numerieke, Booleaanse of datum/tijd-waarde, en niet alleen als een categorische variabele.

15 / 40

Welk type variabele is een categorische variabele?

- A) Booleaanse variabelen
 - B) Discrete variabelen
 - C) Nominale variabelen
 - D) Numerieke variabelen
-
- A) Incorrect. Booleaanse variabelen kunnen waar of onwaar zijn en worden opgeslagen als 16-bits (2-bytes) waarden. Dit is een ander type variabele dan een categorische variabele.
 - B) Incorrect. Een discrete variabele is een variabele die afzonderlijke, telbare waarden aanneemt. Categorische variabelen zijn niet telbaar.
 - C) Correct. Een nominale variabele is een type categorische variabele, oftewel een variabele die één waarde uit een beperkt en doorgaans vastliggend aantal mogelijke waarden kan aannemen. Deze waarden zijn kwalitatief van aard en kunnen niet worden opgeteld. (Literatuur: A, Hoofdstuk 3)
 - D) Incorrect. Numerieke variabelen hebben een waarde waarmee een meetbare hoeveelheid wordt beschreven als een getal, bijvoorbeeld 'hoeveel'. Categorische variabelen zijn niet telbaar.

16 / 40

Welk type variabele is de kilometerstand van een voertuig aan het begin en het einde van een jaar?

- A) Booleaans
 - B) Categorical
 - C) Numeriek
 - D) Datum/tijd
- A) Incorrect. De kilometers kunnen onder of boven een jaarlimiet vallen, maar de waarde op zich is niet Booleaans. De waarde kan elk getal boven 0 zijn.
- B) Incorrect. De kilometerstand kan worden vergeleken met een categorische variabele (de opgelegde grenswaarden onder een bepaald type leaseovereenkomst), maar is zelf geen categorische variabele.
- C) Correct. De variabele kan wiskundig uitgedrukt worden als een integer. (Literatuur: A, Hoofdstuk 3)
- D) Incorrect. De kilometers zijn gekoppeld aan een tijdsinterval, maar het getal zelf is geen datum of tijd.

17 / 40

Waarom is het belangrijk om onderscheid te maken tussen discrete en continue variabelen?

- A) Omdat de geschikte algoritmen voor de variabelen dan beter kunnen worden geselecteerd
 - B) Omdat dit belangrijk is om het conceptuele datamodel te definiëren
 - C) Omdat alleen discrete variabelen met andere variabelen kunnen worden samengevoegd
 - D) Omdat in relationele databases alleen continue variabelen kunnen worden opgeslagen
- A) Correct. Bij het analyseren van data is het belangrijk om te herkennen of een variabele discreet of continu is, aangezien dit bepalend is voor de compatibiliteit met het geselecteerde algoritme. (Literatuur: A, Hoofdstuk 2)
- B) Incorrect. Herkennen of variabelen continu of discreet zijn, is niet van invloed op het conceptuele datamodel waarin inhoud met een hogere abstractie wordt vastgelegd (entiteiten en relaties).
- C) Incorrect. Discrete variabelen kunnen niet met andere variabelen worden samengevoegd of wiskundig worden bewerkt.
- D) Incorrect. In relationele databases kunnen zowel discrete als continue variabelen worden opgeslagen.

18 / 40

Welke techniek voor het scrubben van data houdt in dat variabelen worden getransformeerd naar een integer formaat?

- A) One-hot codering
 - B) Samenvoeging van variabelen
 - C) Selectie van variabelen
 - D) Web-schrapen
- A) Correct. Bij de one-hot codering worden variabelen getransformeerd tot een integer, zodat de algoritmen succesvol kunnen worden uitgevoerd. (Literatuur: A, Hoofdstuk 4)
- B) Incorrect. Bij het samenvoegen van variabelen worden verwante variabelen gecombineerd om zo veel mogelijk informatie te behouden.
- C) Incorrect. Bij het selecteren van variabelen worden binnen een dataset variabelen geselecteerd die waardevoller zijn dan de overige variabelen.
- D) Incorrect. Web-schrapen is geen techniek voor data-scrubbing, maar een methode voor dataverzameling.

19 / 40

Een autoleasebedrijf baseert zijn data analytics op een dataset met kolommen als contractnummer, kenteken, toelatingsdatum, merk, model, kleur en einddatum contract.

De kolommen 'toelatingsdatum' en 'einddatum contract' worden teruggebracht tot een integer met zes posities, waarbij de eerste vier posities overeenkomen met het jaar en de laatste twee met de maand. Zo kunnen de contracten worden onderverdeeld in categorieën met alle contracten die in een bepaalde maand aflopen.

Waarvan is dit een voorbeeld?

- A) Binning
 - B) Samenvoeging van variabelen
 - C) One-hot codering
 - D) Selectie van variabelen
- A) Correct. Bij binning worden numerieke waarden of tijdstempels geconverteerd tot een categorie aan de hand van een discrete integer. (Literatuur: A, Hoofdstuk 4)
- B) Incorrect. De twee variabelen 'toelatingsdatum' en 'einddatum contract' worden op geen enkele wijze samengevoegd.
- C) Incorrect. One-hot codering zet discrete variabelen om in een binair formaat. De resulterende integers kunnen hier meer dan twee waarden bevatten, die geen toestanden zijn.
- D) Incorrect. Bij de selectie van variabelen worden variabelen die geen inzichten opleveren weggelaten. In dit voorbeeld wordt niets weggelaten.

20 / 40

Bij data-scrubbing is het belangrijk om de gebruikte oorspronkelijke brondata te behouden, zodat eventuele revisies mogelijk blijven.

Wat moet er worden gebruikt om dit te waarborgen?

- A) Een beleid voor dataretentie
 - B) Een datawarehouse
 - C) Een sleutelwaarde-opslag
 - D) Kunstmatige intelligentie (KI) algoritmen
- A) Correct. Om de brondata te behouden, is het belangrijk dat er een beleid voor dataretentie wordt opgesteld en gebruikt. (Literatuur: A, Hoofdstuk 4)
- B) Incorrect. Een datawarehouse is een relationele database waarin de data worden opgeslagen. Hiermee wordt niet gewaarborgd dat de brondata behouden blijft. Dit kan alleen worden gewaarborgd door middel van data-governance en regels voor datamanagement.
- C) Incorrect. Een sleutelwaarde-opslag is een database waarin data wordt opgeslagen als een sleutelwaarde-paar.
- D) Incorrect. KI-algoritmen waarborgen niet dat de oorspronkelijke brondata behouden blijft. KI wordt toegepast om inzichten, voorspellingen en conclusies te verkrijgen.

21 / 40

Wat is een kenmerk van inferentiële (toetsende) methoden?

- A) Ze hebben de voorkeur voor situaties waarin data goed is gedocumenteerd en tot één informatiepool is gestandaardiseerd.
 - B) Ze helpen complexe informatie te comprimeren tot een handige, makkelijk leesbare indeling door duidelijke trends samen te vatten.
 - C) Ze draaien om de isolatie en analyse van een deel van de data, waarna de resultaten worden getest ten opzichte van een andere subset van de data.
 - D) Ze bieden een handige en beknopte samenvatting van historische data die is gegenereerd uit een potentieel enorm aantal op zichzelf staande gebeurtenissen.
- A) Incorrect. Dit heeft betrekking op beschrijvende analyse, niet op inferentiële methoden.
- B) Incorrect. Dit is een kenmerk van beschrijvende analyse. Inferentiële methoden worden niet gebruikt om informatie te comprimeren.
- C) Correct. Dit is wat er gebeurt bij opgesplitste validatie, een techniek die wordt gebruikt bij inferentiële methoden. (Literatuur: A, Hoofdstuk 5)
- D) Incorrect. Dit heeft betrekking op beschrijvende analyse. Inferentiële methoden kijken naar de natuurlijke variantie van relatief minder beknopte data.

22 / 40

In welk geval moet beschrijvende analyse worden gebruikt?

- A) Wanneer een dataset niet makkelijk kan worden geanalyseerd
 - B) Wanneer er beperkte informatie beschikbaar is
 - C) Wanneer de opgehaalde data erg gedetailleerd is
 - D) Wanneer er sprake is van hiaten in de administratie
-
- A) Incorrect. In dit geval moet er inferentiële (toetsende) statistische analyse worden gebruikt.
 - B) Incorrect. In dit geval moet er inferentiële (toetsende) statistische analyse worden gebruikt.
 - C) Correct. Beschrijvende analyse moet worden gebruikt wanneer er gedetailleerde data beschikbaar is. (Literatuur: A, Hoofdstuk 5)
 - D) Incorrect. In dit geval moet er inferentiële (toetsende) statistische analyse worden gebruikt.

23 / 40

Wat heeft een menselijke operator nodig om aan de slag te gaan met datamining?

- A) Een gedefinieerde hypothese zoals 'waarschijnlijk' en 'onwaarschijnlijk'
 - B) Een groot aantal mogelijke inputcombinaties
 - C) Een model dat aangeeft welke input een bepaalde output oplevert
 - D) Een probleem waarvan wordt aangenomen dat het de moeite waard is om het op te lossen
-
- A) Incorrect. Dit heeft een operator nodig voor een standaard statistisch model.
 - B) Incorrect. Dit heeft een operator nodig voor reinforcement learning.
 - C) Incorrect. Dit heeft een operator nodig voor supervised learning.
 - D) Correct. Een operator heeft een doel nodig, zoals een probleem dat de moeite waard is om op te lossen. (Literatuur: A, Hoofdstuk 5)

24 / 40

Bij welke machine learning-methode worden trainingsdata en testdata gebruikt?

- A) Datamining-methode
 - B) Beschrijvende methode
 - C) Inferentiële (toetsende) methode
 - D) Opgesplitste validatie-methode
-
- A) Incorrect. Bij datamining worden vooral nieuwe dingen aangeleerd en wordt niet zozeer geprobeerd te begrijpen hoe aangeleerde zaken toepasbaar zijn in een specifieke context, zoals bij machine learning het geval is.
 - B) Incorrect. Beschrijvende analyse is vooral geschikt voor situaties waarin data goed is gedocumenteerd en tot één informatiepool is gestandaardiseerd.
 - C) Correct. Bij inferentiële methoden wordt 70 tot 80% van een dataset gebruikt om een analytische agent te trainen in het herkennen van patronen en die vervolgens te testen op de resterende 20 tot 30%. (Literatuur: A, Hoofdstuk 5)
 - D) Incorrect. Bij inferentiële methoden wordt vaak een techniek gebruikt die opgesplitste validatie wordt genoemd. Opgesplitste validatie is echter niet de volledige methode.

25 / 40

Het verkoopproces van een autoleasebedrijf heeft een grote dataset opgeleverd. Het bedrijf gebruikt machine learning om de puntjes op de i te zetten bij het opstellen van offertes. Alleen de data uit het verkoopproces wordt ingevoerd in de software, zonder aanvullende regels. Vervolgens zoekt de software naar patronen en retourneert een model om zonder al te veel korting de offerteprijs te berekenen.

Welk type machine learning is dit?

- A) Reinforcement learning
 - B) Supervised learning
 - C) Unsupervised learning
-
- A) Incorrect. Er is sprake van een groot aantal mogelijke inputcombinaties, maar die worden niet willekeurig beoordeeld op hun prestaties.
 - B) Incorrect. Er is geen specifieke of precieze regel om de juiste prijs te berekenen op basis van de onafhankelijke inputvariabelen.
 - C) Correct. Er zijn vooraf gelabelde input- en outputcombinaties, maar geen bekende outputs die fungeren als referentiepunt. Bij unsupervised learning worden inputwaarden geanalyseerd om patronen te ontdekken en de prijs voor de best mogelijke prestaties te bepalen. (Literatuur: A, Hoofdstuk 5)

26 / 40

Een belangrijk aspect van natuurlijke taalverwerking (NTV) is syntactische analyse.

Wat wordt er bij syntactische analyse geanalyseerd?

- A) De betekenis van een zin
 - B) De benoemde entiteiten
 - C) De tekst van zoekopdrachten
 - D) De structuur van een zin
-
- A) Incorrect. Dit wordt semantische analyse genoemd.
 - B) Incorrect. Dit wordt classificatie genoemd.
 - C) Incorrect. Dit wordt tekstparsering genoemd.
 - D) Correct. Analyse van de structuur van een zin wordt syntactische analyse genoemd. (Literatuur: A, Hoofdstuk 10)

27 / 40

Een van de taken bij natuurlijke taalverwerking (NTV) omvat het herkennen van belangrijke entiteiten in tekst, zoals een naam, locatie of activiteit.

Welke NTV-taak is dit?

- A) Klasse-identificatie
- B) Herkenning van benoemde entiteiten
- C) Sentimentanalyse
- D) Stemming

- A) Incorrect. Bij klasse-identificatie wordt een document gelabeld met binaire vectoren van opgegeven tokens.
- B) Correct. Bij herkenning van benoemde entiteiten worden in het oog springende tekstonderdelen opgepikt, zoals wat, waar en hoe. (Literatuur: A, Hoofdstuk 10)
- C) Incorrect. Sentimentanalyse wordt gebruikt om de emotionele intentie van een document te detecteren.
- D) Incorrect. Stemming wordt gebruikt om de trefwoorden van een document te parseren in kernbetekenissen.

28 / 40

Wat is het verband tussen data en algoritmen?

- A) Algoritmen worden ontwikkeld om data te verwerken en analyseren.
 - B) Data wordt ontwikkeld om te worden omgezet naar algoritmische kennis.
 - C) Verschillende mensen gebruiken hetzelfde algoritme voor verwerking van dezelfde data.
 - D) Inputdata verandert nooit voor algoritmen, wat resulteert in uniciteit.
-
- A) Correct. Een algoritme is een reeks stappen die reageren op signalen en veranderende patronen uit data. (Literatuur: A, Hoofdstuk 6)
 - B) Incorrect. De ontwikkeling van data houdt geen verband met algoritmische kennis.
 - C) Incorrect. Mensen gebruiken hun eigen interne algoritme voor de evaluatie van data.
 - D) Incorrect. Data verandert altijd en is niet uniek voor algoritmen.

29 / 40

Uit data van een makelaar blijkt dat de gemiddelde prijs van verkochte huizen afneemt in lijn met een stijgende hypotheekrente.

Welke soort regressieanalyse moet worden gebruikt om dit patroon in de data te beschrijven?

- A) Exponentiële regressieanalyse
- B) Lineaire regressieanalyse
- C) Logistische regressieanalyse
- D) Niet-lineaire regressieanalyse

- A) Incorrect. Exponentiële regressieanalyse is niet geschikt voor lineaire relaties tussen variabelen.
- B) Correct. De regressie is lineair en negatief. (Literatuur: A, Hoofdstuk 7)
- C) Incorrect. Logistische regressie levert discrete variabelen op als output.
- D) Incorrect. De regressie is hier lineair, dus is een lineaire regressieanalyse het meest geschikt.

30 / 40

Wat is het doel van regressieanalyse?

- A) De prestaties van een organisatie analyseren als basis voor besluitvorming
- B) Datapunten categoriseren in groeperingen wanneer er geen voorgedefinieerde klassen bestaan
- C) Een lijn of curve vinden die patronen in de data zo goed mogelijk beschrijft
- D) Begrijpen hoe data wordt verzameld, geordend en geïnterpreteerd

- A) Incorrect. Dit heeft betrekking op business intelligence (BI), wat kan worden gedefinieerd als een toolset om informatie over de prestaties van een organisatie te verzamelen, analyseren en rapporteren aan besluitvormers.
- B) Incorrect. Hoewel logistische regressie kan worden gebruikt om records met vergelijkbare of dicht bij elkaar liggende waarden te herkennen, geldt dit niet voor alle regressiemethoden.
- C) Correct. Dit is het doel van regressieanalyse. Hoewel data mogelijk al te eenvoudig worden voorgesteld met slechts één lijn of curve, wordt zo wel een nuttig referentiepunt geboden om algemene voorspellingen over toekomstige data te doen. (Literatuur: A, Hoofdstuk 7)
- D) Incorrect. Dit heeft te maken met statistiek, dat als primair doel heeft de betekenis van data en variaties hierin te bepalen.

31 / 40

Welk type model genereert categorische voorspellingen en clustering?

- A) Algoritmisch model
- B) Classificatiemodel
- C) Regressiemodel

- A) Incorrect. Algoritmen worden gebruikt voor alle datamodellerings technieken en zijn geen model op zich.
- B) Correct. Classificatie is de overkoepelende term voor algoritmen die categorische voorspellingen en clusteranalyses genereren voor onderwerpen als marktonderzoek en klantprofilering. (Literatuur: A, Hoofdstuk 8)
- C) Incorrect. Regressieanalyse is een populaire statistische techniek die wordt gebruikt om de relatie tussen een of meer onafhankelijke variabelen en de afhankelijke variabele te modelleren.

32 / 40

Een bedrijf beschikt over een grote dataset voor de verkoop van overhemden. Het personeel wil de cijfers beter begrijpen en de data analyst vraagt zich af of clusteranalyse wellicht de gewenste inzichten oplevert.

Hoe zorgt k-means clustering voor een beter begrip van de verkoopcijfers?

- A) Door data te clusteren in voorgedefinieerde groeperingen
 - B) Door onverwachte relaties tussen clusters te beschrijven
 - C) Door de rol van zwaartepunten te verklaren
 - D) Door een nieuwe manier te bieden om klanten te groeperen
-
- A) Incorrect. Met k-means clustering is het mogelijk om niet-geïdentificeerde groeperingen te vinden.
 - B) Incorrect. Met k-means clustering is het mogelijk om groepen te vinden die zo ver mogelijk van elkaar gescheiden zijn. Het is echter niet mogelijk om de relatie tussen die groepen te analyseren.
 - C) Incorrect. Zwaartepunten zijn een technische oplossing om groepen te maken, maar spelen verder geen rol van betekenis.
 - D) Correct. K-means clustering is nuttig wanneer er geen kennis over een bestaande categorie beschikbaar is, maar het wel wenselijk is om nieuwe niet-geïdentificeerde groeperingen te vinden. Op basis van deze groeperingen kan een nieuw inzicht ontstaan. (Literatuur: A, Hoofdstuk 8)

33 / 40

Wat is een **belangrijk** verschil tussen associatie analyse en sequentiële mining?

- A) Bij associatie analyse wordt de gegeneraliseerde sequentiepatronen (GSP) techniek toegepast, terwijl sequentiële mining a priori wordt toegepast.
 - B) Bij associatie analyse wordt niet gekeken naar de volgorde waarin items voorkomen, terwijl hier bij sequentiële mining rekening mee wordt gehouden.
 - C) Bij associatie analyse wordt een betrouwbaarheidsniveau gebruikt ter ondersteuning van de besluitvorming, terwijl sequentiële mining een vorm van supervised learning is.
 - D) Associatie analyse werkt beter bij grote datasets, terwijl sequentiële mining het meest geschikt is voor kleine datasets.
-
- A) Incorrect. Bij sequentiële mining wordt GSP gebruikt en associatie analyse gebruikt a priori.
 - B) Correct. Bij associatie analyse wordt geen rekening gehouden met de volgorde van gebeurtenissen, die bij sequentiële mining juist belangrijk is. (Literatuur: A, Hoofdstuk 9)
 - C) Incorrect. Sequentiële mining is geen vorm van supervised learning, maar een type associatie analyse.
 - D) Incorrect. De grootte van datasets voor associatie analyse en sequentiële mining verschilt niet.

34 / 40

Sequentiële mining gebruikt een betrouwbaarheidsniveau om beslissingen te onderbouwen. Het gebruikt ook een soort criterium om te focussen op frequente patronen om deze besluitvorming te onderbouwen.

Welk criterium is dit?

- A) A priori-algoritme
- B) Frequente itemset
- C) Minimum support
- D) Recursieve eliminatie

- A) Incorrect. A priori is een algoritme voor de mining van frequente itemsets en het leren van associatieregels via relationele databases.
- B) Incorrect. Een frequente itemset komt voor in transacties uit de transactiedatabase met minimum support, waarbij minimum support een door de gebruiker ingestelde drempel is.
- C) Correct. Net als bij associatie analyse worden ook bij sequentiële mining criteria voor minimum support gebruikt om te focussen op frequente patronen, evenals een betrouwbaarheidsniveau om geïnformeerde beslissingen te nemen en prioriteiten toe te kennen tijdens de implementatie. (Literatuur: A, hoofdstuk 10)
- D) Incorrect. Recursieve eliminatie is een algoritme om frequente itemsets in een transactiedatabase te ontdekken.

35 / 40

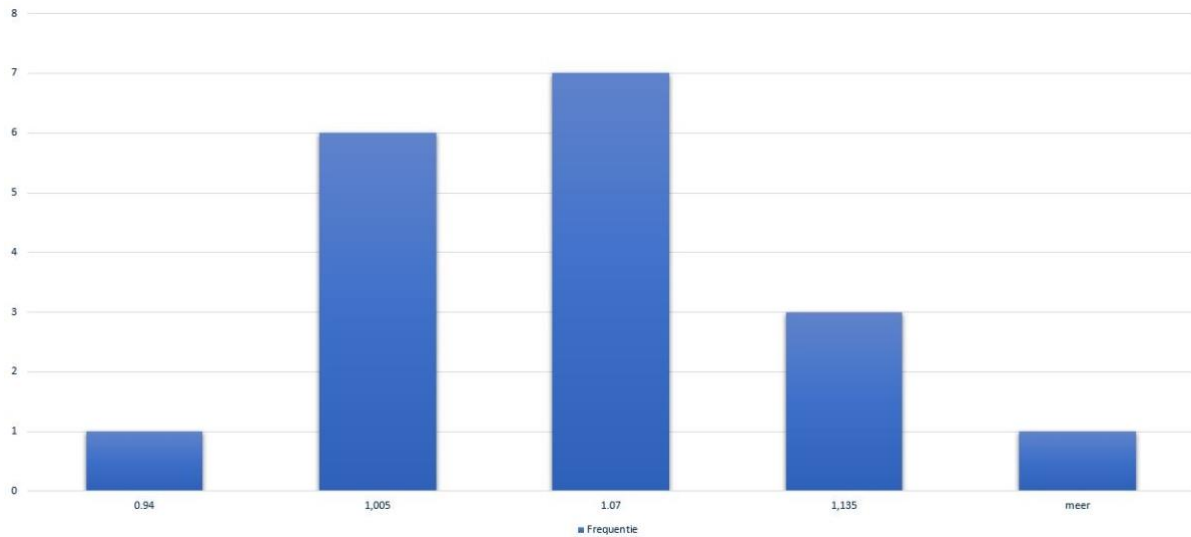
Welke grafische techniek wordt doorgaans gebruikt om data te presenteren aan een extern publiek?

- A) Expansief
- B) Verklarend
- C) Uiteenzettend
- D) Verkennend

- A) Incorrect. Expansief is een term die doorgaans wordt gebruikt in combinatie met het woord 'beleid'. Het is een economisch concept dat verband houdt met technieken om recessie en werkloosheid te voorkomen.
- B) Correct. Verklarende grafische voorstellingen worden gebruikt om data en bevindingen eenvoudiger uit te leggen en het publiek te helpen complexe data beter te begrijpen. (Literatuur: A, Hoofdstuk 11)
- C) Incorrect. Uiteenzetting is een filosofische en literaire techniek die niet wordt gebruikt in data analytics.
- D) Incorrect. Verkennende grafische voorstellingen worden gegenereerd terwijl de analyse wordt uitgevoerd en het model nog in productie is, om zodoende het interne begrip te vergroten.

36 / 40

Bekijk de onderstaande afbeelding:



Waar is dit een voorbeeld van?

- A) Staafdiagram
 - B) Box plot
 - C) Heatmap
 - D) Histogram
- A) Correct. Een staafdiagram is een grafische voorstelling van categorische data met behulp van rechthoekige staven, waarbij de lengte van elke staaf evenredig is aan de voorgestelde waarde. (Literatuur: A, Hoofdstuk 12)
- B) Incorrect. Een box plot toont de vijf-getallensamenvatting van een dataset.
- C) Incorrect. Een heatmap is een tweedimensionale voorstelling van data waarin waarden worden voorgesteld door middel van kleuren.
- D) Incorrect. Een histogram is een grafische voorstelling van data gegroepeerd in continue numerieke bereiken die overeenkomen met verticale staven. Een histogram heeft geen witruimte tussen de staven.

37 / 40

Welk type plot wordt gebruikt om de spreiding en scheefheid van een dataset weer te geven?

- A) Box plot
 - B) Rug plot
 - C) Scatter plot
 - D) Violin plot
-
- A) Correct. Een box plot beschrijft de symmetrie van data en wordt gebruikt om de verdeling van een set continue data samen te vatten en weer te geven. (Literatuur: A, Hoofdstuk 11)
 - B) Incorrect. Een rug plot visualiseert de verdeling van een losse variabele, niet de scheefheid van een dataset.
 - C) Incorrect. Een scatter plot visualiseert informatie over de relatie tussen continue variabelen, niet over de verdeling en scheefheid van een dataset.
 - D) Incorrect. Een violin plot beschrijft de verdeling en dichtheid van data, maar niet de scheefheid ervan.

38 / 40

Om meer inzicht te krijgen in de prijzen van verkochte huizen in verschillende wijken van een stad wil een makelaar een heatmap gebruiken.

Wat toont deze heatmap?

- A) Een stadsplattegrond waarop de niveaus van de huizenprijzen met kleuren worden aangeduid
 - B) Een stadsplattegrond met verkochte huizen en de verkoopprijs per huis
 - C) Een tabel met per kwartaal de gemiddelde huisprijzen voor het afgelopen jaar, gesorteerd op de gemiddelde prijs
 - D) Een tabel met de prijzen van verkochte huizen het afgelopen jaar, geordend op prijs
-
- A) Correct. In een heatmap worden data met kleuren aangeduid. (Literatuur: A, Hoofdstuk 12)
 - B) Incorrect. In een heatmap worden data met kleuren aangeduid. Uit niets blijkt dat er voor deze plattegrond kleuren zijn gebruikt.
 - C) Incorrect. In een heatmap worden data met kleuren aangeduid. Uit niets blijkt dat er in deze tabel, die sowieso geen plattegrond is, kleuren zijn gebruikt.
 - D) Incorrect. In een heatmap worden data met kleuren aangeduid. Uit niets blijkt dat er in deze tabel, die sowieso geen plattegrond is, kleuren zijn gebruikt. Bovendien is het voor een heatmap niet nodig om prijzen te ordenen.

39 / 40

Waarom is een esthetisch ontwerp belangrijk?

- A) Omdat daarmee de technieken voor machine learning worden uitgebreid
 - B) Omdat daarmee natuurlijke taalverwerking (NTV) mogelijk wordt
 - C) Omdat het is gericht op modellen voor regressieanalyse
 - D) Omdat daarmee de bruikbaarheid van datavisualisaties toeneemt
-
- A) Incorrect. Machine learning is niet afhankelijk van datavisualisatie. Het wordt gebaseerd op datasets.
 - B) Incorrect. NTV heeft te maken met taal, syntaxis en semantiek. Het is niet afhankelijk van datavisualisatie.
 - C) Incorrect. Met een esthetisch ontwerp wordt elke vorm van datavisualisatie verbeterd. Het ontwerp is niet specifiek gericht op regressieanalyses.
 - D) Correct. Principes voor een esthetisch ontwerp omvatten zorgvuldige ontwerp- en kleurcombinaties die de bruikbaarheid van datavisualisaties ten goede komt. (Literatuur: A, Hoofdstuk 11)

40 / 40

Welke tool voor datavisualisatie heeft het uiterlijk van Microsoft Excel?

- A) Data Wrapper
 - B) Google Charts
 - C) Power BI
 - D) Tableau
-
- A) Incorrect. Data Wrapper heeft een eigen interface.
 - B) Incorrect. Google Charts is compatibel met Google-producten.
 - C) Correct. Power BI is compatibel met andere Microsoft-producten. (Literatuur: A, Hoofdstuk 12)
 - D) Incorrect. Tableau heeft een eigen interface.

Evaluatie

De juiste antwoorden op de vragen in dit voorbeeldexamen staan in onderstaande tabel.

Vraag	Antwoord	Vraag	Antwoord
1	B	21	C
2	C	22	C
3	A	23	D
4	C	24	C
5	D	25	C
6	C	26	D
7	B	27	B
8	B	28	A
9	B	29	B
10	A	30	C
11	B	31	B
12	C	32	D
13	A	33	B
14	B	34	C
15	C	35	B
16	C	36	A
17	A	37	A
18	A	38	A
19	A	39	D
20	A	40	C



Driving Professional Growth

Contact EXIN

www.exin.com