

Provkonstruktion

Årskurs: Gymnasiet

Ämne eller kurs: Artificiell Intelligens 2

Tema: Datasets och datarengöring

Syfte

Syftet med detta prov är att bedöma elevernas förmåga att arbeta med och förbereda datasets för maskininlärning, inklusive deras kunskaper i datarengöring, hantering av saknade värden och normalisering av data.

Koppling till styrdokument

Centralt innehåll

Kursens centrala innehåll omfattar att arbeta med och förbereda datasets för maskininlärning. Detta inkluderar datarengöring, hantering av saknade värden och normalisering av data för att säkerställa att modellen får högkvalitativ input.

Kunskapskrav

Eleven ska kunna analysera och förbereda data för maskininlärning och beskriva olika steg i datarengöring och normalisering.

Prov

Faktafrågor

1. Vad innebär datarengöring?

- A. Att skapa nya dataset
- B. Processen att förbereda och korrigera data för analys
- C. Att visualisera data
- D. Att lagra data i en databas**

2. Varför är normalisering viktigt?

- A. För att öka datamängden
- B. För att säkerställa att data står i rätt förhållande till varandra och inte snedvrider resultaten**

- C. För att minska antalet datapunkter
 - D. För att spara tid i analysen
3. Nämn ett exempel på ett problem som uppstår i datasets.
- A. För mycket data
 - B. För lite data
 - C. Saknade värden**
 - D. Förändrade datatyper
4. Vad är en vanlig metod för att hantera saknade värden?
- A. Ta bort hela datasetet
 - B. Ignorera saknade värden
 - C. Att fylla i med medelvärde eller median**
 - D. Använda slumpmässiga värden
5. Vilken av följande metoder är en typ av normalisering?
- A. Korrelationsanalys
 - B. Min-max normalisering**
 - C. Klasstillhörighet
 - D. Dataaugmentation
6. Vad är en konsekvens av dålig datakvalitet i maskininlärning?
- A. Högre noggrannhet
 - B. Fler dataresultat
 - C. Felaktiga förutsägelser och bristande resultat**
 - D. Förbättrad modellprestanda
7. Vad innebär normalisering av data?
- A. Att rensa bort onödiga data
 - B. Att skapa standardformat för datalagring
 - C. Att skala data så att olika datamängder kan jämföras**
 - D. Att duplicera data för säkerhetskopiering
8. Vilken typ av dataset används oftast i maskininlärning?
- A. Textdata
 - B. Image data
 - C. Tabulär data**
 - D. Videodata
9. Vad kan anomalier i ett dataset leda till?

- A. Ökad datakvalitet
- B. Förbättrade modeller
- C. Missledande slutsatser**
- D. Inga effekter

10. Vilket av följande är inte en teknik för datarengöring?

- A. Imputation av saknade värden
- B. Borttagning av outliers
- C. Skapa nya datalager**
- D. Korrigering av felaktiga värden

11. Vad gör en datarapport användbar?

- A. Den är kort och koncis
- B. Den innehåller väl presenterad och korrekt data**
- C. Den fokuserar endast på stora datamängder
- D. Den innehåller onödig information

12. Vad innebär det att fylla saknade värden med median?

- A. Att använda det mest förekommande värdet
- B. Att använda det mittersta värdet av datasetet**
- C. Att använda ett slumpmässigt värde
- D. Att ta bort hela raden

13. Vilket program används ofta för datarengöring?

- A. Python**
- B. Word
- C. Photoshop
- D. HTML

14. Vad står "GIGO" för?

- A. Garanti in, garanti ut
- B. Garbage in, garbage out**
- C. God information, god analys
- D. Ge informativ, ge om

15. Hur kan datakvalitet mätas?

- A. Genom att räkna antalet datapunkter
- B. Genom att analysera modellernas hastighet
- C. Genom att granska data för konsistens och noggrannhet**
- D. Genom att observera användarbeteende

Resonerande frågor

1. Diskutera vikten av datarengöring i maskininlärning och ge exempel på hur dåliga data kan påverka en modell. Syftet med denna fråga är att ge eleverna möjlighet att visa sin förståelse för datarengöringens roll i maskininlärning.
2. Beskriv hur normalisering kan påverka resultaten från en maskininlärningsmodell. Genom att besvara denna fråga kan eleverna visa sin kännedom om normaliseringens betydelse.
3. Analysera ett dataset och identifiera minst tre olika typer av problem som kan förekomma. Denna fråga låter eleverna demonstrera sina analytiska färdigheter och förståelse för databasproblem.
4. Resonera kring skillnaden mellan olika metoder för att hantera saknade värden och deras inverkan på dataanalys. Genom att diskutera metoderna får eleverna möjlighet att jämföra och kontrastera lösningar.
5. Reflektera över hur kvaliteten på data kan påverka affärsbeslut i en organisation som använder maskininlärning. Syftet är att ge eleverna möjlighet att utforska praktiska tillämpningar av sina kunskaper.
6. Vad kan konsekvenserna av att ignorera datarengöring vara för en organisation som implementerar AI-lösningar? Genom att svara på denna fråga kan eleverna tänka kritiskt om realvärlden.
7. Diskutera hur metoder för datarengöring utvecklas med ny teknik och nya dataformat. Här kan eleverna visa sin insikt i trendanalys och teknologiska framsteg.
8. Analysera hur datarengöring och normalisering är relaterade till varandra och deras kollektiva inverkan på modellernas precision. Denna fråga utmanar eleverna att tänka på samband och detaljer i processen.

Bedömning

Provet bedöms med totalt 30 poäng möjligt, där faktafrågorna ger totalt 15 poäng (1 poäng vardera) och resonerande frågorna ger totalt 15 poäng (upp till 3 poäng vardera beroende på djup och kvalitet av svaret).

För betyg E krävs minst 8 poäng, för C 12 poäng (varav minst 3 poäng från resonerande frågor) och för A 18 poäng (varav minst 5 poäng från resonerande frågor).

Tags: [Prov](#)