

Hemläxa

Årskurs: Gymnasiet

Ämne eller kurs: Artificiell Intelligens 2

Tema: Datasets och datarengöring

Ordkollen

Här listas fem ämnesord på läxans tema som är bra att känna till betydelsen av.

- **Dataset:** En samling av data som används för analys eller träning av en maskininlärningsmodell.
- **Datarengöring:** Processen att förbereda och korrigera data för att säkerställa dess kvalitet innan analys.
- **Saknade värden:** Datapunkter i ett dataset som inte har något registrerat värde.
- **Normalisering:** En metod för att skala data så att den hamnar inom ett specifikt intervall, vilket förenklar jämförelser mellan olika datatyper.
- **Anomalier:** Avvikande värden i datasetet som inte passar in med resten av datan och som kan påverka analysen negativt.

Instuderingsfrågor

- Vad är syftet med att arbeta med datasets inom maskininläring?
- Beskriv några vanliga problem som kan förekomma i datasets.
- Varför är datarengöring viktigt innan analys av data?
- Ge exempel på metoder för att hantera saknade värden.
- Vad innebär begreppet normalisering av data?
- Nämn två olika metoder för normalisering och förklara kort hur de fungerar.
- Hur kan anomalier påverka resultaten av en analys?
- Varför är det viktigt att diskutera mål och åtgärder med datarengöring efter en gruppaktivitet?
- Vad kan hända om dålig datakvalitet inte åtgärdas innan en maskininlärningsmodell implementeras?
- Reflektera över vikten av samarbete inom grupper när datarengöring utförs. Vad kan fördelarna vara?

Skrivuppgift

Uppgift 1: Analysera ett dataset

Välj ett dataset som du har tillgång till. Analysera det och identifiera eventuella problem med datakvaliteten. Skriv en sammanställning av dina iakttagelser.

Svarslängd: ca. 300 ord (En halv sida)

Uppgift 2: Datarengöringsplan

Skapa en plan för hur du skulle hantera de problem du har identifierat i ditt valda dataset, inklusive hur du skulle angripa saknade värden och anomalier.

Svarslängd: ca. 500 ord (En sida)

Uppgift 3: Normaliseringsexempel

Välj en metod för normalisering (exempelvis min-max eller z-score) och tillämpa den på en del av datan i ditt valda dataset. Redovisa processen och diskutera eventuella effekter av normaliseringen.

Svarslängd: ca. 400 ord (En sida)

Kom ihåg att dokumentera alla steg noggrant så att du kan referera till dem senare och dela med andra.

Tags: [Läxa](#)