

KURS 2 · ØKT 2

Hva lærte treet egentlig?

8. trinn

40 av 60 min

Uten KI

Mål for økta: elevene erfarer at et system kan lære et mønster som passer treningsdataene, men som ikke er det vi mente, og kan forklare forskjellen på treningsdata og testdata.

Kort fortalt: gruppene lager et enkelt beslutningstre som skal skille ulv fra hund. Halvparten får et skjevt treningssett der alle ulvene står i snø. Treet deres lærer å kjenne igjen snø, ikke ulv, og bommer på testen. Elevene skal gå ut av timen med at et system kan være perfekt på treningen og likevel ha lært feil ting, og at bare en test med nye eksempler avslører det.

Før timen

rundt 25 min

- Skriv ut dyrekortene: [dyrekort.pdf](#) . Side 1 er treningssett 1, side 2 er treningssett 2 og side 3 er testsettet. Halvparten av gruppene får treningssett 1, resten treningssett 2. Alle får testsettet. Klipp ut kortene.
- Legg treningssettet og testsettet i hver sin konvolutt. Merk hvilke grupper som får sett 1 og sett 2.
- Skriv ut beslutningstre-arket, ett per gruppe: [beslutningstre.pdf](#) .
- Elevene trenger KI-loggen sin ([ki-logg.pdf](#) hvis noen trenger nytt ark).
- Les «Det du trenger å vite» under, og fasiten for testsettet i fasiten for kurset (egen PDF på øktsiden). Du leser opp fasiten i trinn 2.
- Åpne presentasjonen: [okt-2.html](#) . Trykk N for å se lærernotatene under lysbildet. PDF-en er reserve hvis nettleseren ikke virker.

Ikke si til elevene at treningssettene er forskjellige. Det skal de oppdage selv på tavla.

Det du trenger å vite

Kortene har ingen bilder, bare egenskaper: pels, ører, hale, størrelse og hvor bildet er tatt (snø, skog, gress eller by). Hvert treningskort har merkelapp: ulv eller hund. Dyrene er oppdiktet og forenklet.

Et beslutningstre er en rekke ja-nei-spørsmål. For eksempel: «Er bildet tatt i snø? Ja: ulv. Nei: hund.» Gruppene lager et tre med høyst tre spørsmål.

Treningssett 1 er skjevt. Alle ulvene står i snø, ingen av hundene gjør det. Bakgrunnen er det eneste som skiller alle tolv kortene. Gruppene vil derfor lage et tre som spør om snø.

Treningssett 2 er balansert. Både ulver og hunder står i snø. Her er det halen som skiller kortene: ulvenes hale henger ned, hundenes krøller seg opp.

Testsettet har dyr der bakgrunnen er snudd: hunder i snø og ulver andre steder. Fasit: T1 hund, T2 ulv, T3 ulv, T4 hund, T5 hund, T6 ulv, T7 ulv, T8 hund.

Hva som trolig skjer: et tre som spør om snø, får 1 av 8 riktige. Et tre som spør om halen, får 7 av 8. T5 er en hund med hale som henger ned, og blir feil også med halen. Kontrasten på tavla er poenget med økta.

Begreper: treningsdata er eksemplene systemet lærer av. Testdata er nye eksempler systemet ikke har sett.

Snarveilæring er når et system scorer høyt ved å bruke noe tilfeldig i dataene, som snøen, i stedet for det det skulle lære.

I kurs 1 til 3 prøver elevene først og får forklaringen etterpå. Derfor kommer det praktiske før faktadelen.

Slik går timen

Del	Lysbilde	Tid
Start: hva tror du?	1 og 2	3 til 5 min
Praktisk: tren, test, sammenlign	3 til 5	14 til 18 min
Fakta: treningsdata og testdata	6 til 9	6 til 10 min

Del	Lysbilde	Tid
Diskusjon	10	6 til 10 min
KI-logg	11	3 til 5 min

Lysbilde 12 er kreditering, og trenger ikke vises i timen.

Start: hva tror du?

3 til 5 min

KJERNE

Lysbilde 1 og 2. Ble KI-loggen fra økt 1 flyttet, gjør elevene den nå.

Ellers deler du ut en lapp til hver elev. **Si:** «Hvis en maskin får se mange bilder av ulver og hunder, hva tror du den legger merke til? Skriv det på lappen.» Lappen legges i konvolutten.

1. Praktisk

14 til 18 min

KJERNE

Grupper på tre eller fire. Halvparten av gruppene har treningssett 1, resten treningssett 2.

Lysbilde 3. Trinn 1: tren et tre (6 til 8 min). Del ut treningssettet og beslutningstre-arket.

Si: «Dere har tolv dyrekort med merkelapp: ulv eller hund. Lag et tre med høyst tre ja-nei-spørsmål som sorterer alle tolv riktig. Spørsmålene må kunne besvares ut fra det som står på kortet. Skriv hvor mange kort treet sorterer riktig.»

Se etter, og spør:

- Grupper med treningssett 1 som spør om bakgrunnen. Ikke stopp dem. Det er det eneste som skiller alle kortene.
- Grupper med treningssett 1 som unngår bakgrunnen fordi den «ikke handler om dyret». De får høyst 11 av 12. Spør: «Hva er det enkleste som skiller kortene?»
- Grupper med treningssett 2. Halen skiller alle kortene der.

Lysbilde 4. Trinn 2: test treet (4 til 5 min). Del ut testsettet.

Si: «Her er åtte nye kort uten merkelapp. Sorter dem bare med treet, og skriv treets svar på arket.»

Les opp fasiten: T1 hund, T2 ulv, T3 ulv, T4 hund, T5 hund, T6 ulv, T7 ulv, T8 hund. Gruppene teller hvor mange som ble riktige.

Lysbilde 5. Trinn 3: sammenlign (3 til 4 min). Skriv to kolonner på tavla: treningssett 1 og treningssett 2. Hver gruppe leser opp hvor mange de fikk riktige og hva spørsmål 1 i treet var.

Si, når tallene står på tavla: «Alle trærne var perfekte på treningen. Hvorfor gikk det så forskjellig på testen?»

2. Fakta og info

6 til 10 min

KJERNE

Lysbilde 6. Si: «Treningsdata er eksemplene systemet lærer av. Testdata er nye eksempler det ikke har sett. Bare testen viser om systemet har lært noe som virker utenfor treningen.»

Lysbilde 7. Si: «Systemet finner det enkleste som skiller eksemplene. I treningssett 1 var det snøen, ikke dyret. Treet var perfekt på treningen og dårlig på testen.»

Lysbilde 8. Si: «Dette har skjedd på ordentlig. I 2016 trente forskere med vilje en maskin på 20 bilder der alle ulvene sto i snø. Den kalte bilder ulv når det var snø i bakgrunnen, uansett hvilket dyr det var. Forskerne laget den slik for å teste om folk oppdaget feilen.»

Lysbilde 9. Si: «Forskere kaller dette snarveilæring. Systemet bruker noe tilfeldig i dataene i stedet for det det skulle lære. Det skjer fortsatt, og snarveiene er ofte vanskelige å oppdage.»

Knytt til økt 1: Nim-maskinen ble bare god mot den motstanden den møtte.

3. Diskusjon

6 til 10 min

Lysbilde 10. Still ett spørsmål om gangen. La elevene tenke alene i et halvt minutt, snakke med den som sitter ved siden av i ett minutt, og si det til hele klassen etterpå.

- **Kjernespørsmål:** «Hva lærte treet egentlig?» (Treet fra sett 1 lærte å kjenne igjen snø, ikke ulv.)
- **Utdyping:** «Hvordan kunne dere oppdaget feilen før testen?» (Ved å se om noe annet enn dyret skilte kortene, eller ved å sjekke om dataene hadde ulver uten snø.)
- **Utdyping:** «Var treningssett 2 perfekt? Hva gikk galt med kort T5?» (T5 er en hund med hale som henger ned. Ingen regel virker på alle.)
- **Utdyping:** «Hvem har ansvaret når et system lærer feil mønster: den som samlet dataene, den som laget systemet, eller den som bruker det?» (Ikke ett riktig svar.)

4. Skrivning og refleksjon

3 til 5 min

KAN FLYTTES

Lysbilde 11. Elevene ser på lappen fra starten. Så skriver de i KI-loggen:

- Dagens spørsmål: «Hva ville du gjort med treningssett 1 for å rette det?» (Godt svar: ta med ulver som ikke står i snø, og hunder som gjør det.)
- «Nå tenker jeg...»
- «Jeg lurar på...»

Har dere dårlig tid, flytter du denne delen til starten av økt 3.

Misforståelser å følge med på

- At et system som er perfekt på treningsdataene, er et godt system.
- At mer data alltid hjelper. Tolv kort til med ulver i snø hadde ikke rettet noe.
- At systemet «ser» dyret. Det ser bare egenskapene det får, og bruker dem som passer best.

Tilpasning

- **Kort tid:** alle grupper får treningssett 1. Du viser treet for treningssett 2 på tavla etterpå.
- **Mer tid:** gruppene bytter treningssett og lager et nytt tre. Eller de lager et eget testkort som lurar treet til en annen gruppe.

Kilder og kreditering

- Ribeiro, Singh og Guestrin, «"Why Should I Trust You?": Explaining the Predictions of Any Classifier» (2016), om klassifisereren for ulv og husky som ble trent med vilje på bilder med snø.
- Pathak m.fl., oversiktsartikkel om snarveilæring og «Clever Hans»-effekten, *Frontiers in Artificial Intelligence* (2025).
- Kortene er oppdiktet og kontrollert med eget skript.