

KURS 4 · ØKT 1

Hvem er med i dataene?

9. trinn

40 av 60 min

Uten KI

Mål for økta: elevene kan forklare hvorfor et system kan fungere dårligere for grupper som er lite med i treningsdataene, og viser at et gjennomsnitt kan skjule store forskjeller.

Kort fortalt: elevene får høre om forskning som viser at ansiktsgjenkjenning feiler oftere for noen grupper. Så teller de 60 datakort fra en oppdiktet ansiktslås til elevskap, tegner stolper og regner ut hvor ofte systemet feiler for hver gruppe. Elevene skal gå ut av timen med at grupper som er lite med i dataene, rammes oftest, og at et samlet tall på 82 prosent riktige skjuler det.

Før timen

rundt 25 min

- Skriv ut datakortene, ett sett per gruppe på tre eller fire: [datakort.pdf](#) (tre sider, 60 kort). Klipp ut kortene og legg hvert sett i en konvolutt.
- Skriv ut stolpearket, ett per gruppe: [stolpeark.pdf](#).
- Skriv ut testresultatene, ett per gruppe: [testresultater.pdf](#). Arket deles ut først i trinn 3.
- Elevene trenger KI-loggen sin ([ki-logg.pdf](#) hvis noen trenger nytt ark).
- Les «Det du trenger å vite» under, og fasiten i fasiten for kurset (egen PDF på øktsiden).
- Åpne presentasjonen: [økt-1.html](#). Trykk N for å se lærernotatene under lysbildet. PDF-en er reserve hvis nettleseren ikke virker.

Det du trenger å vite

Hovedideen: et system som lærer av data, blir best på det det har sett mest av. Er noen grupper lite med i treningsdataene, får systemet lite øvelse på dem, og feiler oftere for dem.

Forskningen elevene får høre om:

- **Gender Shades (2018):** to forskere testet tre kommersielle systemer som skulle avgjøre om et ansikt tilhørte en kvinne eller en mann. For menn med lys hud var feilprosenten høyst 0,8. For kvinner med mørk hud var den opptil 34,7. To mye brukte bildesamlinger hadde 79,6 og 86,2 prosent personer med lys hud.
- **NIST (2019 og 2025):** det amerikanske standardinstituttet testet 189 algoritmer fra 99 utviklere. De fleste feilet ulikt for ulike grupper, for noen typer feil mellom 10 og over 100 ganger så ofte. De mest nøyaktige algoritmene hadde minst forskjell. Underrepresentasjon i treningsdata kan være én årsak. Dårlig lys og bildekvalitet er en annen.

Øvelsen: 60 datakort, hvert merket med hudtone og kjønn. Kortene har ingen bilder. Tallene er oppdiktet, men mønsteret er det forskerne har funnet.

Gruppe	Hudtone, kjønn	Antall kort	Feil av 12 forsøk	Feilprosent
A	lys, gutt	22	0	0
B	lys, jente	20	1	8
C	middels, gutt	6	1	8
D	middels, jente	6	3	25
E	mørk, gutt	4	3	25
F	mørk, jente	2	5	42

Til sammen 59 av 72 riktige, altså 82 prosent. Det høres bra ut, men for gruppe F feiler systemet nesten annenhver gang.

Viktig: elevene skal ikke telle eller plassere seg selv eller hverandre i gruppene. Si det høyt før trinn 1.

I kurs 4 kommer faktadelen før det praktiske, fordi kurset er på 9. trinn.

Slik går timen

Del	Lysbilde	Tid
Start: hva tror du?	1 og 2	3 til 5 min
Fakta: forskningen og gjennomsnittet	3 til 6	6 til 10 min
Praktisk: sorter, tegn, les testresultater	7	14 til 18 min
Diskusjon	8	6 til 10 min
KI-logg	9	3 til 5 min

Lysbilde 10 er kilder, og trenger ikke vises i timen.

Start: hva tror du?

3 til 5 min

KJERNE

Lysbilde 1 og 2. Del ut en lapp til hver elev.

Si: «Hvis et system skal kjenne igjen ansikter, hva må være med i bildene det lærer av, for at det skal virke for alle? Skriv det på lappen.» Lappen legges i konvolutten.

1. Fakta og info

6 til 10 min

KJERNE

Lysbilde 3. Si: «Et system som lærer av data, vet bare det som finnes i eksemplene. Er noen lite med i dataene, får systemet lite øvelse på dem.»

Lysbilde 4. Si: «I 2018 testet to forskere tre systemer som skulle avgjøre om et ansikt tilhørte en kvinne eller en mann. For menn med lys hud tok systemene feil i under én prosent av tilfellene. For kvinner med mørk hud tok de feil i opptil 35 prosent. Bildesamlingene mange brukte, hadde rundt 80 prosent personer med lys hud.»

Lysbilde 5. Si: «Det amerikanske standardinstituttet NIST testet 189 algoritmer. De fleste feilet oftere for noen grupper enn for andre. De beste algoritmene hadde minst forskjell. En mulig årsak er at noen grupper er lite med i treningsdataene. Dårlig lys kan også gi feil.»

Lysbilde 6. Si: «Et system kan ha 82 av 100 riktige til sammen, og likevel feile nesten annenhver gang for én gruppe. Det skal dere se selv nå.»

2. Praktisk

14 til 18 min

KJERNE

Lysbilde 7. Grupper på tre eller fire. Del ut datakortene og stolpearket.

Si først: «Dere får 60 datakort fra en oppdiktet ansiktslås til elevskap. Tallene er oppdiktet, men mønsteret er det forskerne har funnet. Kortene har ingen bilder. Ingen skal telle eller plassere seg selv eller andre i gruppene.»

Trinn 1: sorter og tell (5 til 6 min). **Si:** «Sorter kortene i seks bunker etter det som står på kortet. Skriv antallet i hver bunke øverst på stolpearket.»

Trinn 2: tegn stolpene (3 til 4 min). **Si:** «Fargelegg én rute per kort, så dere får seks stolper. Svar på spørsmålet under: hvilken stolpe er høyest, hvilken er lavest, og hva tror dere det betyr for systemet?»

Trinn 3: les testresultatene (6 til 8 min). Del ut testresultatene.

Si: «Hver rad er 12 forsøk der systemet skulle åpne et skap. Et kryss betyr at systemet ikke kjente igjen eleven. Tell kryssene i hver rad, regn ut prosent ved å dele på 12, og skriv det under stolpen for samme gruppe. Til slutt: hvor mange av alle 72 forsøkene gikk riktig?»

Se etter, og spør:

- Grupper som regner prosent ut fra 60 i stedet for 12. Pek på at hver rad har 12 forsøk.
- Grupper som sier at systemet «bare er dårlig på mørk hud». Be dem se på gruppene A til F hver for seg. (D, middels hud og jente, har like mange feil som E.)
- Grupper som ser sammenhengen mellom lav stolpe og høy feilprosent. Spør: «Hvorfor tror dere det er slik?»

3. Diskusjon

6 til 10 min

Lysbilde 8. 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:** «Hvem merker feilen først, og hvem får vite om den?» (Eleven som ikke får åpnet skapet merker det med en gang. De som laget systemet, får kanskje aldri vite det.)
- **Utdyping:** «82 prosent riktige til sammen. Hvorfor kan det være en dårlig målestokk alene?» (Det skjuler at gruppe F feiler 42 prosent av gangene.)
- **Utdyping:** «Hvordan kunne de som laget systemet oppdaget feilen før det ble brukt?» (Telle hvem som var med i dataene, og teste hver gruppe for seg.)
- **Utdyping:** «Hvem har ansvaret hvis eleven i gruppe F ikke får åpnet skapet: de som samlet bildene, de som laget systemet, eller skolen som tok det i bruk?» (Ikke ett riktig svar. Kurs 9 handler om dette.)

4. Skrivning og refleksjon

3 til 5 min

KAN FLYTTES

Lysbilde 9. Elevene ser på lappen fra starten. Så skriver de i KI-loggen:

- Dagens spørsmål: «Hva måtte vært annerledes i dataene for at systemet skulle virke for alle?» (Godt svar: flere kort fra gruppene som var lite med, og test for hver gruppe.)
- «Nå tenker jeg ...»
- «Jeg lurte på ...»

Har dere dårlig tid, flytter du denne delen til starten av økt 2.

Misforståelser å følge med på

- At skjeve resultater alltid skyldes at noen har laget systemet med vilje. Her skyldes det hvem som var med i dataene.
- At mer data løser det. Det hjelper bare hvis dataene inneholder de som mangler.
- At et høyt gjennomsnitt betyr at systemet virker for alle.
- At forskjellene bare gjelder ansikter. Det samme kan skje med stemmer, dialekter og språk.

Tilpasning

- **Kort tid:** hopp over trinn 2. Gruppene skriver antallene på stolpearket og går rett til testresultatene.
- **Mer tid:** gruppene foreslår hvor mange kort av hver gruppe de ville hatt i et nytt datasett på 60, og begrunner.
- **Knytt til dialekter:** kortet «Tale til tekst» i kurs 1 kan hentes fram. Spør hvilke dialekter som er lite med i et norsk datasett.

Kilder og kreditering

- Buolamwini og Gebru, «Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification», Proceedings of Machine Learning Research 81 (2018).
- Grother, Ngan og Hanaoka, NISTIR 8280, «Face Recognition Vendor Test Part 3: Demographic Effects» (2019), og NISTs «FRTE Demographics» (tabeller oppdatert 5. mars 2025).
- Tallene i øvelsen er oppdiktet.