

## Tekst-annotation og Inter-Rater Reliability øvelse (25 min)

Når vi forsøger at forstå hvorfor nogle grupper klarer sig bedre end andre, trækker vi (ofte ubevidst) på fundamentalt forskellige forklaringsmodeller:

- **Strukturelle forklaringer** peger på systemiske faktorer: økonomiske strukturer, institutionel diskrimination, nedarvede privilegier, adgang til ressourcer. Her ligger årsagen uden for individet.
- **Individuelle forklaringer** peger på personlige faktorer: motivation, valg, indsats, kultur, mindset. Her placeres ansvaret hos individet selv.

I den virkelige verden møder vi ikke rene “strukturelle” eller “individuelle” forklaringer. I stedet møder vi dem i:

- Politiske taler og debatter
- Mediernes dækning af sociale problemer
- Sociale medier og kommentarspor
- Interviews og surveys

For at kunne *systematisk analysere* hvordan ulighed forklares/tales om i offentligheden, skal vi kunne:

1. **Identificere** hvilken forklaringstype der anvendes
2. **Kategorisere** store mængder tekst konsistent
3. **Måle** hvor enige vi er om kategoriseringen

> *Kategoriser hver udtalelse som enten STRUKTUREL eller INDIVIDUEL forklaring på ulighed. Sammenlign derefter jeres resultater og diskutér: Hvor var I uenige, og hvorfor? Hvad fortæller det om teksterne og den “menneskelige” vinkel i denne form for kvantitativt tekstanalyse?*

1. **Individuel annotation:**
    - a. Kategoriser hver tekst som enten STRUKTUREL eller INDIVIDUEL
  2. **Gruppesammenligning:**
    - a. Sammenlign jeres labels
    - b. Udregn procentvis enighed:  $(\text{antal enige} / \text{total antal tekster}) \times 100$
    - c. Udregn Cohen's Kappa (justerer for tilfældig enighed, se nedenfor)
    - d. Diskutér uenigheder: Hvorfor var I uenige?
  3. **Refleksion:**
    - a. Manglede i guidelines? Hvilke?
    - b. Hvordan kunne annotation forbedres?
- 

### Cohen's Kappa

$$\kappa = \frac{(P_o - P_e)}{(1 - P_e)}$$

hvor:

$P_o$  er observeret enighed (antal enige / total)

$P_e$  er forventet tilfældig enighed

```
from sklearn.metrics import cohen_kappa_score
annotator1 = [1, 0, 0, 0, 1, 1, 0, 0, 1, 0, 0, 1, 1, 0, 0, 1, 1, 0, 0, 1]
annotator2 = [1, 0, 1, 0, 0, 1, 0, 0, 1, 1, 0, 0, 1, 0, 1, 0, 1, 0, 0, 1]
kappa = cohen_kappa_score(annotator1, annotator2)
print(f"Cohen's Kappa: {kappa:.2f}")
```

**Tekster: forestil jer de kommer fra en "fri-tekst" besvarelse i en survey**

1. "Folk er fattige fordi de træffer dårlige valg, men når systemet gør det dyrt at være fattig, er det svært at træffe gode valg."
2. "Uddannelsessystemet favoriserer børn fra ressourcestærke hjem, det er bare fakta."
3. "Selvom der er strukturelle barrierer, kan man stadig komme langt med den rette indstilling og vilje."
4. "Det handler om at tage ansvar for sit eget liv, uanset hvilke kort man er blevet delt."
5. "Når praktikpladser og elevpladser primært går til 'dem man kender', er der ikke lige konkurrence."
6. "Man kan godt anerkende at forholdene er svære, men folk har også et personligt ansvar for at forbedre deres situation."
7. "Det offentlige transportsystem er dårligt i udkantsområderne, så mange kan ikke komme til jobsamtaler i byen."
8. "Nogle mennesker fra svære kår klarer sig fremragende, det beviser at det kan lade sig gøre hvis man vil det nok."
9. "Lav mindsteløn betyder at folk kan arbejde fuld tid og stadig leve i fattigdom."
10. "Kultur og normer i nogle miljøer nedprioriterer uddannelse, og det skal vi tale åbent om."
11. "Selvfølgelig er der ulige startbetingelser, men det må ikke blive en undskyldning for ikke at kæmpe."
12. "Når vuggestuepladser koster 3000 kr om måneden, kan enlige forsørgere på kontanthjælp ikke komme i arbejde."
13. "Der er mange strukturelle udfordringer, men samtidig ser vi også at nogle ikke gør nok for at hjælpe sig selv."
14. "Systemet straffer folk der forsøger at komme ud af kontanthjælp; man mister boligstøtte og friplads samtidig."
15. "Man skal have den rigtige mentalitet og ikke lade sig definere af sin baggrund."
16. "Forskellen i ressourcer mellem folkeskoler i rige og fattige kvarterer er dokumenteret og problematisk."
17. "Problemet er manglende motivation og for meget fokus på at være offer for systemet."
18. "Folk fra akademikerfamilier har et netværk der åbner døre, det handler ikke kun om kvalifikationer."
19. "Ja, udgangspunktet er ulige, men Danmark har SU og gratis uddannelse - mulighederne er der hvis man griber dem."
20. "Når børn vokser op i fattigdom, påvirker det deres hjerneudvikling og fremtidige læring - det er ikke et valg."