

# Etikk-rapport

Videregående elevers avsluttende vitnemål består som regel av standpunktkarakterer og eksamenskarakterer. I denne oppgaven drøfter vi noen etiske problemstillinger man bør ta hensyn til dersom man i stedet for å gjennomføre eksamen, predikerer elevenes eksamenskarakterer på bakgrunn av deres og andre elevers tidligere skolearbeid, og eventuelt annen persondata om elevene.

Dataene i det aktuelle datasettet inneholder mye informasjon om elevene utover deres tidligere skoleprestasjoner. Eksempler på slike variabler er: kjønn, helse, sivilstatus, alkoholforbruk, foreldrenes utdanning, foreldrenes jobb og andre sosioøkonomiske variabler. All inndata vi gir algoritmene om elevene vil kunne bli brukt til å predikere deres eksamenskarakterer.

Det er flere argumenter for å ta med sosioøkonomiske variabler i predikeringen. Dersom det finnes signifikante korrelasjoner mellom for eksempel foreldrenes utdanning og elevenes prestasjon på eksamen, så vil en algoritme kunne benytte seg av denne informasjonen og predikere resultatet mer nøyaktig. Man kan tenke at det er en svakhet at elevers karakterer påvirkes av faktorer utenfor elevens kontroll, som for eksempel foreldrenes mulighet til å hjelpe dem med skolearbeidet. Men det er samtidig naivt å tenke at elevene stiller med like forutsetninger og at faktorer som foreldrenes nivå av utdanning ikke har en reell effekt på elevers prestasjoner. Det kan likevel være etiske utfordringer ved å benytte seg av denne informasjonen, som for eksempel personvern.

Fra modelleringene vi har gjennomført ser vi for eksempel at både helse og kjønn har en signifikant effekt på prediksjon av karakterene. Kjønn og helse har tydeligvis hatt (en minimal) påvirkning på de tidligere elevenes karakterer. Vi ønsker imidlertid ikke at eksamenskarakterer skal påvirkes av elevenes kjønn eller helse, vi ønsker at karakterene skal gjenspeile elevenes faktiske kompetanse og ikke fordommer knyttet til hverken kjønn eller helse. Det er ingen overraskelse at det finnes fordommer i samfunnet, og det at algoritmene synliggjør dem for oss, kan være en god ting.

Vi kan også bruke algoritmer til å utligne fordommer og/eller sosioøkonomiske variabler i prediksjonen av karakterer. Men det er viktig å huske på at vår modell ikke har som mål i seg selv å utligne sosioøkonomiske forskjeller, og at det å ta med sosioøkonomiske variabler med i modellen, i verste fall kan være med på å forsterke mekanismen som man ønsker å utligne ved å inkludere. Det kommer an på hvilken (om noen) bias og eventuelle fordommer som ligger i datasettet vi bruker. Dessuten er ikke nødvendigvis eksamenskarakterer de rette instrumentene å benytte dersom man ønsker å utligne sosiale forskjeller.

Om et overordnet ideal for karaktersetting er at den skal være rettferdig, så er det klart at maskinlæringsmodellen er mer kvalifisert til denne oppgaven. Alternativet til bruk av maskinlæringsmodellen er at det er lærere som avgjør eksamenskarakterene, dagens modell. Målt opp mot mennesker, er maskinlæringsmodellen mer presis, objektiv, konsistent og transparent i sin karaktersetting. Med maskinlæringsmodellen har vi også mye mer kontroll og tilgang på hva slags faktorer som blir vektet enn om et menneske skulle satt karakteren. Den er derfor mer kvalifisert enn lærere til å gi en rettferdig vurdering basert på elevens kunnskaper fordi den i større grad utelukker menneskelige feil og eksterne, tilfeldige faktorer som stemningsleie og dagsform hos både elev og sensor på vurderingsdagen, i tillegg til sensorens holdninger og verdier.

Det ser tilsynelatende ut som om maskinlæringsmodellen kan ta mer rettferdige avgjørelser, men hva har det å si om elevene ikke har tillit til systemet og føler seg fremmedgjort? Noen elever føler allerede på en slags fremmedgjøring og reduksjonisme fra før; at ens arbeid blir redusert til et tall. Ved å overlate karaktersettingen til maskinlæringsalgoritmer, kan dette føre til at denne følelsen forsterkes; man ender opp uten noen form for menneskelig samhandling i vurderingssituasjonen av mennesket.

Det er heller ikke nødvendigvis slik at man er bedre egnet bare fordi man er mer presis; det å være egnet er mer enn presisjon. Faktorer som tillit, vil også ha noe å si; det er viktig at elevene har tillit til karaktersetteren. Vi er nok mer villige til å godta noe som er vurdert av en lærer enn en maskin til dels på grunn av artstilørighet og arts-forskjeller mellom oss. Vi har mindre problemer med å godta menneskelige feil fordi vi forventer ikke at menneske er feilfrie. Vi måler derimot maskiners prestasjoner opp mot et ideal fordi vi forventer at de skal innfri feilfritt, men dette er kanskje noe vi burde endre på om vi vektlegger rettferdighet mer enn tillit.

Hvorfor vi burde vektlegge tillit er fordi vi vil kunne føle at vi er i kontroll over våre egne karakterer. Det kan se ut som om at man blir maktesløs i hendene på maskinen som skal sette eksamenskarakteren din, men du er egentlig i mye mer kontroll enn du selv tror (mest relevant for modell 1). Bare eleven blir fortalt om prosessen bak modellen, tror vi at denne innsikten vil bidra til å gi eleven forståelse for avgjørelsene maskinlæringsmodellen tar. Ved å la alle involverte parter gjøre seg kjent med denne vurderingstypen, kan skolen også være rettfærdiggjort i å la disse velge hvilke elever som skal bli tilbudt ekstra støtteordninger gjennom skoleløpet og imøtekomme deres behov. De involverte vil forstå hvorfor akkurat vedkommende trenger en slik støtteordning, og elevene kommer til å forstå at man i større grad er i kontroll over hva utfallet blir, siden algoritmen baserer karakteren på faktorer som man tidligere har kunnet påvirke, slik som skolevalg og faglige prestasjoner. Man blir i større grad vurdert ut ifra en helhetsvurdering heller enn én enkelt vurderingssituasjon slik man blir på en vanlig skoleeksamen der tilfeldigheter har mer å si for hvilken karakter du får på prestasjonen den dagen.

Basert på disse refleksjonene har vi valgt å inkludere elevenes tidligere eksamenskarakterer (G1 og G2), antall ganger elevene har strøket og hvilken skole de går på, som variablene i modell 3.