

# Oblig 1

Classification:

< Iris />

Jeg brukte «iris» datasettet for å få en grunnleggende forståelse for hvordan man utfører train, validation og split via scikitlearn.

Til å begynne med virka det uklart igjennom alle oppgavene hva det var jeg egentlig var på utkikk etter i koden og resultatene jeg fikk. Startet rolig med å følge eksemplene til faglærer, for å skjønne fremgangsmåte og se hva det var som fokuset skulle rettes mot.

Igjennom første datasettet «iris», virka det uforståelig hva tallene sa og betydde.

Jeg satt inn variablene train, val, og test, og skikket på tallene, og klødde meg en del i hodet av tallene. Etter at dette virka litt uforståelig søkte jeg nemlig opp hva det egentlig var jeg gjorde «her».

Det viser seg jo det at dette kun er ren prosessering av data, for å få en underbyggende forståelse av hvordan datasettene skal se ut for å klargjøres til å kjøres på en modell. Eller en «maskin».

Jeg fikk fint til å bruke split metodene gitt av faglærer, og jeg kunne se at «test» settet til slutt var fordelt på de 3 typene av iris-blomsten, var på omlag 0.33% gjennomsnitt ca. De var alle ca fordelt på settet, og det var ikke noe særlig overfitting eller underfitting av instansene.

Regression:

< insurance />

Her hadde jeg et stort datasett med forsikring for kunder. Hva prisen på forsikringen var, avhengig av kundens alder og hverdagskost i form av om de var røykere, hadde barn etc..

Jeg splittet datasettet opp og gjorde det klart for «modellen».

Clustering:

< Spotify dataset />

Dette datasettet bestod i om lag 6500 rader, og det var en god del. Var opprinnelig 22 kolonner også, som jeg etter hvert kuttet ned på til om lag 15-16 stk.