

# Intervjuoppgave: Maskinlæring

Sammen med dette dokumentet finner du to filer:

- TrainAndValid.csv (ca. 119MB)
- Data Dictionary.xlsx

Ved å bruke dataene i disse filene, samt noen fritt valgte metoder og teknologier, ønsker vi at du predikere forventet salgspris (SalePrice) på anleggsmaskiner. Vi ønsker videre at oppgaven løses med et fokus på forklarbarhet.

## Løs følgende oppgaver:

1. Lag en enkel statistikk for noen utvalgte variabler, både visuelt og numerisk. (Snitt, Max, Min).
2. Gjør en vurdering av mengden manglende verdier, konsekvens av dette, og nevner noen metoder for å håndtere det.
3. Gjør en vurdering av datatyper, kontinuerlige/kategoriske variabler, og hvordan dette håndteres.
4. Si litt om hvilke eksterne datakilder som kunne hjulpet å modellere problemet. (Bonuspoeng dersom du henter inn noe)
5. Gjør et modellvalg, og utform en kort begrunnelse som belyser fordeler og ulemper.
6. Velg og beskriv en metrikk.
7. Beskriv resultatene (hvilke "features" er viktigst?)
8. Lag en beskrivelse av hvilke data du valgte som testsett og hvilke data du valgte som treningssett.
9. Redegjør for hvordan du har tatt høyde for "overfitting" og hvor god overførbarhet du mener modellen har til nye data.
10. Lag et REST API som eksponerer prediksjonen fra modellen din der du kan få prediksjoner for ny input.

## Bonusoppgaver du kan gjøre dersom du ønsker og har tid:

11. Bygg en Docker-container som eksponerer API-et ditt.
12. Sett modellen din (Docker-containeren) i drift som en skytjeneste. Det skal være mulig å få gratis-credits / free tier hos f.eks. Azure/AWS/GCP for dette formål. Heroku er også en mulighet.

Struktur, kommentarer, kodekvalitet og testing er noe vi vil vektlegge for kandidater som har noen års erfaring.

Lag en presentasjon på ca. 8-16 slides med besvarelser på disse oppgavene.

Ved andregangs/teknisk intervju legger du kildekoden på GitHub i forkant og sender link til kontaktpersonen din senest dagen før avtalt presentasjon. Presentasjonen fremvises under intervjuet.

Lykke til!