

Intervjuoppgave: Maskinlæring

Sammen med dette dokumentet finner du to filer:

- TrainAndValid.csv (ca. 117MB)
- Data Dictionary.xlsx

Ved å bruke dataene i disse filene samt noen fritt valgte metoder og teknologier skal du predikere forventet salgspris (SalePrice) på anleggsmaskiner.

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 om datatyper, kontinuerlige/kategoriske variable, og hvordan dette håndteres.
4. Gjør et modellvalg, og utform en kort begrunnelse som belyser fordeler og ulemper.
5. Velg og beskriv en metric.
6. Beskriv resultatene, samt hvilke features er viktigst.
7. Lag en beskrivelse av hvilke data du valgte som testsett og hvilke data du valgte som treningssett.
8. Redegjør for hvordan du har tatt høyde for overfitting og hvor god overførbarhet modellen har til nye data.

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. Presentasjonen fremvises under intervjuet.

Lykke til!