

Bruktbiler kan være vanskelig å prise skikkelig. Forhandlere og byttebiler kan være vanskelig å prise skikkelig, og kan ofte ta mye lengre tid enn forventet.

Det finnes lignende AI og applikasjoner for å sjekke dette på nettet, men om det faktisk brukes av store og små forhandlere er ikke sikkert. Det brukes nok mange timer og ressurser til å regne ut, noe som koster ekstra penger.

For at prosjektet skal være en suksess, må vi kunne lage, oppdatere og vedlikeholde både webside og server. Vi må nå en høy nok accuracy til at gjennomsnittet holder seg bra, og vi får forventet verdi.

Jeg skal bruke et dataset fra kaggle, det inkluderer diverse informasjon som: bilmerke, modell, drivstoff, girkasse og kjørelengde. Siden pris er target, og den er kontinuerlig, blir dette et regresjonsproblem. Denne dataen kan også hentes inn fra diverse bruktbil databaser eller nettsider, f.eks finn.no.

Dataen bør gjennomgås for å sikre at det ikke er åpenbar feil, for eksempel en bil til en milliard inne i datasettet.

Personvern er også viktig ved slik henting av data, på finn.no er det ofte fullt navn, registrasjon osv.

Fjerning av nullverdier og endring av urealistisk data trengs for å fikse opp feil.

Jeg bruker en simpel Random Forest Regressor i applikasjonen, dette er helst for å ta forhold til ikke-lineære data. For senere implementasjon kan vi bruke mer avanserte og robuste modeller.

For å estimere baseline ytelse og oppførsel, har vi lineær regresjon mellom variabler og bilpris.

Enkle beregninger som gjennomsnittspris på bilmerker vil også bli brukt som baseline.

Jeg har lagt til feature_importances i koden for å se hvilke 10 funksjoner som er viktige for pris.

For å forbedre modellen kan vi utforske interaksjonen mellom for eksempel modellår og kjøreavstand. Vi kan og analysere feature importance til å se hvilke funksjoner som er viktig og lite viktig, og eventuelt fjerne dem for å redusere kompleksiteten.

Forbedring:

Lagre bruker inputs, for å kunne trene modellen videre. Det er viktig for å holde seg oppdatert.

Bygge for skalerbarhet, spesielt ved mange requests på en gang.

Oppdatere modellen til en med bedre ytelse, eller kanskje ensemble mange modeller i lag.

Referanser:

<https://www.kaggle.com/datasets/sidharth178/car-prices-dataset/data?select=test.csv>

<https://gdpr.eu>