

Machine learning - DTU

Rapport

Anna Louise Hansen

Fra Udviklings- og Forenklingsstyrrelsen

Chapter 1

Part I

1.1 Beskrivelse af data

Alle boligejere i Danmark betaler en skat, ejendomsværdiskat, som er baseret på værdien af deres ejendom. Dette vil sige hele ejendommen inkl. grunden som boligen ligger på. For at kunne gøre dette laver den danske stat offentlige ejendomsvurderinger som disse skatter bliver baseret på. Det er derfor vigtigt at disse vurderinger er retvisende og ikke mindst forklarbare, således at en borger kan forstå hvilke parametre der ligger til grund for ejendomsvurderingen. Til dette project har jeg valgt at arbejde med anonymiseret data fra mit arbejde i udviklings- og forenklingsstyrrelsen, hvor jeg til dagligt arbejder med netop dette. Datasættet består af ejendomssalg fra en 6 årig periode. Ud over selve huspriserne består data også af en lang række attributter som beskriver karakteristika ved selve boligen. Det kan f.eks. være tagmateriale, boligens opførelsesår, information om størrelsen af huset og grunden eller bbr koder som dækker over boligens anvendelse. Der ud over består data også af en lang række attributter som fortæller noget om hvor boligens beliggenhed. Det kan f.eks. være boligens koordinater eller information om afstanden til kyst og skov eller afstand til motorvej og jernbane. Data kommer fra en række forskellige registre og offentlige styrelser som eks. BBR og Styrrelsen for Dataforsyning og Effektivisering.