FINANSTILSYNET

NOTAT

Finanstilsynet

10. juli 2019

God praksis ved brug af superviseret machine learning

1. Indledning

Den finansielle sektor bruger i stigende grad nye teknologier som f.eks. superviseret machine learning. Der ligger et stort potentiale i dette, og Finanstilsynet forventer, at teknologien vil finde bred anvendelse i den finansielle sektor. Finanstilsynet ønsker at bidrage til en positiv udvikling på området. Virksomhederne bør dog ikke være ukritiske i deres brug af de "nye" teknologier.

Finanstilsynet og e-nettet har i perioden fra august 2018 til marts 2019 specifikt drøftet, hvordan en virksomhed kan bruge superviseret machine learning til at værdiansætte danske ejerboliger. Det er sket i Finanstilsynets regulatoriske sandkasse, FT Lab.

Hvad er machine learning?
Machine learning er en undergruppe af kunstig intelligens. Machine learning kan kort beskrives som algoritmer, der behandler data, lærer fra dette data og derefter bruger det lærte til at træffe velinformerede beslutninger.

På baggrund af de erfaringer, Finanstilsynet bl.a. har gjort i FT Lab, er det formålet med dette papir at gøre opmærksom på nogle af de risici, som superviseret machine learning indebærer. Det er desuden formålet at vejlede virksomhederne og inspirere dem til at træffe de rigtige beslutninger i forhold til håndteringen af disse risici.

Papiret oplister en række emner, som virksomheder, der er omfattet af den finansielle lovgivning, bør overveje i forbindelse med brug af superviseret machine learning. Papiret er et foreløbigt og ikkeudtømmende katalog med overvejelser, som virksomheder i den finansielle sektor løbende bør gøre sig, hvis de benytter superviseret machine learning.

Papiret skal dermed ses som et supplement til de krav, som virksomheder i øvrigt skal overholde, hvis de er omfattet af den finansielle regulering, som

f.eks. specifikke krav til virksomhedernes brug af modeller. Det gælder eksempelvis de krav, der stilles til institutters brug af modeller til udregning af kapitalgrundlag.

Hvad er superviseret machine learning?
Superviseret machine learning er en delmængde af machine learning, hvor man både kender variable for input og output. På baggrund af de kendte variable udledes den optimale kobling og vægtning mellem alle inputvariable og outputvariablen. Denne sammenhæng kan dernæst beskrive nye eksempler.

Machine learning vil i flere tilfælde indgå i finansielle virksomheders modelarbejde. Hvilken øvrig lovgivning, der finder anvendelse, vil afhænge af de konkrete aktiviteter, som den enkelte virksomhed udfører ved hjælp af machine learning. Virksomheder skal derfor holde sig for øje, at regler for almindelig risikostyring ved modelanvendelse, som bl.a. fremgår af CRR[1], fortsat gælder. Dette papir vil ikke komme nærmere ind på disse aspekter.

Finanstilsynet tager ikke stilling til, hvilke konkrete værktøjer eller modeller, en virksomhed bør bruge. Papiret foreskriver ikke specifikke standarder, men beskriver en god praksis ved brug af superviseret machine learning. Kravene vil være højere, hvis der er tale om aktiviteter, der er væsentlige, enten for virksomhedens forretningsmodel, risikostyring eller for forbrugere, end hvis aktiviteterne ikke er væsentlige.

Virksomheden skal ikke inddrage Finanstilsynet i enhver brug af superviseret machine learning. Når en virksomhed ønsker at benytte machine learning til at udføre en reguleret aktivitet, kan det dog i konkrete tilfælde kræve tilladelse eller dispensation fra Finanstilsynet. I sådanne tilfælde bør virksomheden som minimum have gjort sig de overvejelser, der er beskrevet i dette papir, før den beder om tilladelse eller dispensation.

Grundlæggende er Finanstilsynets betragtninger udtryk for, at finansielle virksomheders brug af superviseret machine learning ikke adskiller sig markant fra anden statistisk analyse eller aktivitet generelt. I alle tilfælde skal virksomheden sikre, at interne processer understøtter en betryggende drift med tilstrækkelig risikostyring, rapportering, regnskabsaflæggelse og kundebetjening, der er tilpasset de udførte eller udbudte aktiviteter.

2. Baggrund

Finanstilsynet har i regi af den regulatoriske sandkasse, FT Lab, sammen med e-nettet undersøgt, hvordan en virksomhed kan bruge superviseret machine learning ved hjælp af et neuralt netværk til at værdiansætte ejerboliger.

[1] Europa-Parlamentets og Rådets forordning (EU) nr. 575/2013 af 26. juni 2013

Finanstilsynet og e-nettet undersøgte, hvordan processer, udvikling og resultater kan beskrives, dokumenteres og forklares, når der er tale om komplekse modeller baseret på superviseret machine learning.

Værdiansættelse af ejerboliger med superviseret machine learning indebærer, at modellen trænes på historiske data, hvor man kender både input, såsom antal kvadratmeter, afstand til skole, placering på støjkort, og output, altså den faktiske salgspris (den pris en ejendom reelt er solgt til). Modellen afsøger forbindelser og vægtning mellem inputvariablene for at finde en optimal sammenhæng til bedst muligt at forklare outputvariablen, den faktiske salgspris. Når modellen er blevet tilstrækkeligt god til at ramme historiske data, dvs. når forudsigelserne ligger tæt nok på salgspriserne, kan man bruge modellen til at estimere en eventuel salgspris på en vilkårlig ejerbolig. Modellen kan med andre ord give et bud på, hvad en specifik bolig med specifikke karakteristika (input) kan handles for (output).

Brug af andre varianter af machine learning skal givetvis tage højde for andre specifikke forhold, som ikke bliver behandlet her. Finanstilsynet forventer dog, at de overordnede principper i papiret kan benyttes generelt på al brug af machine learning.

Hvad er neurale netværk?
Neurale netværk betegner her en række algoritmer, som forsøger at efterligne de processor, der udføres af biologiske neurale netværk som f.eks. menneskets hjerne. Neurale netværk "lærer" at udføre opgaver ved at betragte konkrete eksempler uden på forhånd at være programmeret med kriterierne for løsning af opgaven. På den måde kan neurale netværk f.eks. bruges til billedgenkendelse, idet de kan ”lære” at identificere billeder. Det klassiske eksempel er et billede af en kat. Netværket lærer at identificere disse billeder ved at analysere træningsdata, der manuelt er mærket som enten "kat" eller "ikke kat". Ved hjælp af disse resultater kan netværket identificere katte på nye billeder. De neurale netværk gør netop dette uden forudgående viden om katte, f. eks. viden om, at de har pels, knurhår og haler. I stedet genererer netværket automatisk identifikationsegenskaber fra det læringsmateriale, det behandler.

Papiret afspejler ikke kun de konkrete drøftelser, Finanstilsynet har haft med e-nettet i regi af FT Lab, men er en sammenfatning af overvejelser, Finanstilsynet indtil videre har gjort sig, bl.a. i forskellige internationale sammenhænge.

3. Formål med brug af superviseret machine learning og beskrivelse af modellen

Machine learning-modeller kan bruges til at løse en lang række opgaver i den finansielle sektor. Det kan eksempelvis være at monitorere mistænkelige transaktioner, værdiansætte aktiver og automatisere kunderådgivning. Afhængigt af, hvad den konkrete model mere specifikt skal bruges til, vil det

være nødvendigt at gøre sig forskellige overvejelser. Det er altså helt centralt, at en finansiel virksomhed, der ønsker at benytte en machine learning-model, først gør sig klart, hvad formålet med modellen er.

Formålsvurderingen er central, fordi alle efterfølgende modelvalg bør afgøres i lyset af denne. Virksomheden skal træffe mange valg i løbet af en models livscyklus, og alle disse valg bør træffes på baggrund af formålet. Formålet med en konkret model bør derfor være klart og tydeligt beskrevet. I beskrivelsen bør det som minimum fremgå, hvilken konkret opgave en model skal udføre, hvorfor løsningen bedst bliver opnået med machine learning, og hvilke interne eller eksterne interessenter modellen skal gavne.

Træning, brug og opdatering af en model skal understøtte modellens formål. Det er derfor væsentligt, at formålet er formuleret, så alle interessenter forstår det. Det forudsætter en konkret og klar formulering, som i stor udstrækning er renset for fagtekniske termer. En sådan formulering vil også sikre, at alle involverede, herunder beslutningstagere, forstår, hvorfor det konkrete formål bedst bliver opnået ved hjælp af den konkrete model.

Det bør samtidig, og før modellen bliver sat i drift, stå klart for organisationen, hvorfor en konkret opgave bedst kan løses ved hjælp af machine learning. Dette er særligt vigtigt, hvis den udførte aktivitet er særskilt reguleret, eksempelvis af den finansielle regulering. Denne kan fastsætte særlige krav, virksomheden skal tage hensyn til.

En virksomhed kan udvikle machine learning-modeller med udgangspunkt i et konkret formål eller mere undersøgende, hvor virksomheden afsøger tilgængelige interne data for at finde uopdagede sammenhænge. Selvom der kan være værdi i en ren undersøgende tilgang af en virksomheds interne data for at finde sådanne sammenhænge, bør den finansielle virksomhed nøje overveje og beskrive formålet med en model, inden modellen bliver taget i drift. Det skal sikre, at virksomheden først tager en model i brug, når det står klart, hvad den kan bidrage med, og på hvilket grundlag dette er afgjort.

God praksis for brug af superviseret machine learning indebærer, at:

- virksomheden har beskrevet formålet med modellen klart, så den senere kan træffe beslutninger på et fyldestgørende grundlag
- beskrivelsen af modellen understøtter en tilstrækkelig forståelse på alle relevante niveauer i organisationen.

4. Governance (modeludvikling, –anvendelse og -opdatering), politikker og forretningsgange

Finansielle virksomheder skal allerede overholde krav til governance. Virksomhederne skal f.eks. udarbejde politikker og forretningsgange på en række

områder i medfør af bl.a. ledelsesbekendtgørelserne. Disse krav skal sikre, at virksomhederne bliver drevet forsvarligt og håndterer alle relevante risici. Machine learning har potentiale til at skabe nye typer risici, som virksomhedernes governancestruktur derfor bør håndtere.

Udvikling og brug af machine learning-modeller bør som udgangspunkt følge samme procedurer og metoder som virksomhedens øvrige IT-udvikling. Selvom der kan være behov for at tilpasse virksomhedens procedurer og metoder, så de tager højde for machine learning-modellers særegne karakteristika, bør virksomheden ikke udvikle disse modeller uden at inddrage den generelle viden om IT-udvikling, som den i forvejen besidder.

Der findes adskillige modeltyper og måder at optimere dem på. Udviklingen af modeller bør derfor løbende dokumenteres og valideres. Logning af ændringer til modeller bør også foregå systematisk. Virksomheden bør altså kunne dokumentere valg og fravalg under udvikling, drift og opdatering af en konkret model.

I valget mellem forskellige modeltyper og dokumentationen heraf bør virksomheden opsætte kriterier for udvælgelse. Kriterierne bør indeholde mere end blot performance. De bør også forholde sig til robusthed, jf. afsnit 7, og forklarlighed, jf. afsnit 9. Modellens begrænsninger, også i kvalitet af inputdata, bør desuden være beskrevet, så det står klart for modellens interessenter, under hvilke omstændigheder den bør bruges.

Disse forhold afhænger af modellens formål og det forhold, som modellen beskriver eller forklarer. Virksomheden bør overveje, hvilke faktorer der kan påvirke modellen, og hvordan disse bedst bliver håndteret. Der vil eksempelvis være forskel på, om en model skal forklare et marked, hvor de samme relativt få faktorer er væsentlige over en lang periode, såsom boligmarkedet, eller et marked, hvor faktorernes væsentlighed er mere varierende, såsom de finansielle markeder.

Virksomheden bør også tænke sin brug af machine learning ind i den løbende risikostyring og implementere klare retningslinjer for ledelsesmæssig opfølgning. Det indebærer, at virksomheden bør sikre en passende governancestruktur til at håndtere brugen af teknologien. Virksomheden bør have eller implementere tilstrækkelige foranstaltninger for IT-sikkerhed og for at modstå eksempelvis cyberangreb, som forsøger at påvirke modellen eller inputdata.

Machine learning bør med andre ord indgå i de sædvanlige governancestrukturer, som allerede findes på de områder, hvor teknologien kan benyttes. I det omfang, der opstår nye processer, når virksomheden bruger machine learning, som den ikke har taget stilling til i sin sædvanlige governancestruktur, vil det være relevant at genbesøge og opdatere det eksisterende set-up. Dette

skal sikre, at virksomheden også kan håndtere de nye processer. Virksomheden bør derfor overveje, i hvilket omfang machine learning giver anledning til justeringer i politikker, instrukser og forretningsgange samt rapporterings- og kontrolprocedurer.

God praksis for brug af superviseret machine learning indebærer, at:

- brugen indgår med de nødvendige tilpasninger i virksomhedens sædvanlige governance
- virksomheden dokumenterer og logger valg og fravalg igennem modellens livscyklus, så modellens historik er sikret
- modellens begrænsninger er grundigt beskrevet, også i forhold til datakvalitet.

5. Datahåndtering

Machine learning er ikke et nyt fænomen. Teknologien har været kendt i mange årtier. Den er dog i de senere år blevet betydeligt mere udbredt, i takt med at virksomhederne i stigende omfang har fået adgang til den nødvendige computerkraft og datamængde. Særligt tilgængeligheden af større og mere komplekse datamængder stiller en række krav til finansielle virksomheder, der benytter machine learning.

Virksomheden bør tage aktivt stilling til den data, som modellen bruger, i tilknytning til formålet med machine learning. Virksomheden bør beskrive, hvordan den sikrer datakvalitet og stabilitet i indhentning af data. Særligt, hvis data hentes fra tredjemand, er det vigtigt, at virksomheden forholder sig til, hvordan den kan sikre kvaliteten af data, og at den har stabil adgang hertil. Tilgangen til dette arbejde bør være risikobaseret, sådan at de mest kritiske datakilder sikres bedst. Disse overvejelser forudsætter, at virksomheden afdækker sit databehov i forhold til den konkrete model.

Det kan også være en fordel at have en liste over alternative datakilder i tilfælde af, at en ekstern leverandør ikke til stadighed kan levere det nødvendige data. Jo vigtigere data er for modellen, jo mere nødvendige bliver alternative datakilder.

Virksomheden bør dokumentere processen for håndtering af alt data, der indgår i modellen. Det inkluderer, hvordan data er behandlet forud for, at det bruges. Det gælder eksempelvis, om outliers er fjernet, variable er normaliseret eller ustruktureret data er opdelt i intervaller. Virksomheden bør som en del af dette arbejde også beskrive, hvorfor den har bearbejdet data, og hvordan data eventuelt er annoteret.

God praksis for brug af superviseret machine learning indebærer, at:

- virksomheden afdækker sine databehov
- virksomheden sikrer, at datakvalitet og stabilitet i levering kan opretholdes på et tilfredsstillende niveau
- virksomheden dokumenterer sin proces for håndtering af data.

6. Træning af modellen

Når virksomheden udformer en model, der er baseret på machine learning, bør den vælge, hvordan den vil optimere modellen. Denne optimering bør direkte understøtte det opsatte formål. Hvis formålet med modellen eksempelvis er at estimere output på enkelte observationer, bør modellens træning minimere afvigelser på enkeltniveau. Hvis formålet derimod er at estimere output, der er grupperet i en større portefølje, bør træningen minimere afvigelser på gruppe- eller porteføljeniveau.

En virksomhed bør derfor kunne forklare de valg, den har truffet i forbindelse med optimering af modellen. Det er en forudsætning for at bruge machine learning, at virksomheden kan beskrive sammenhængen mellem modellens formål og den tilgang, der er benyttet til træning og optimering.

Virksomheden bør beskrive eventuelle effekter fra datahåndtering i forbindelse med træning af modellen. Er der eksempelvis frasorteret data, som kan betyde, at modellen performer bedre eller værre i visse datasegmenter, eller betyder den valgte optimering, at specifikke datapunkter tillægges større vægt end andre? Virksomheden bør i alle tilfælde kunne beskrive, hvordan modellens træning understøtter, at formålet med at bruge machine learning bedst muligt opnås.

En superviseret machine learning-model kræver ofte meget store mængder data for at kunne opnå resultater, som er væsentligt bedre, end hvad man kan opnå med mere klassiske statistiske modeller.

Den store mængde data bør opdeles i flere delmængder for at sikre, at en model er konsistent på tværs af data. En model bør i første omgang trænes på et datasæt, her kaldet *træningssættet*. Hver gang modellen er blevet optimeret, bør den valideres på et datasæt, som ikke er indgået i træningen, her kaldet *valideringssættet*. Når modellen opnår en tilstrækkelig præcision på valideringssættet, bør den testes på et yderligere datasæt, som ikke tidligere har indgået i træningen, her kaldet *testsættet*. Denne tilgang minimerer risikoen for, at modellen bliver overparametriseret eller overfitted[2], hvilket kan

[2] Overparametrisering eller overfitting betyder, at en model beskriver et datasæt så nøjagtigt, at modellen ikke længere er egnet til at forklare yderligere eller fremtidig data. En model mister derved sin funktionalitet.

medføre, at modellen ikke er retvisende. Denne type risiko eksisterer allerede ved klassiske statistiske modeller, men ved machine learning er risikoen forøget.

Virksomheden bør derfor forholde sig til, hvordan de data, den bruger til henholdsvis træning, validering og test, opdeles og benyttes for bedst muligt at understøtte modellens træning og dermed formål. Opdeling af data skal understøtte, at modellen beskriver den virkelighed, virksomheden ønsker beskrevet, og ikke blot det træningssæt, modellen er udviklet på baggrund af.

God praksis for brug af superviseret machine learning indebærer, at:

- virksomheden tilrettelægger træning af modellen, så modellens formål bedst muligt understøttes, herunder valg af optimering
- virksomheden kan beskrive, om datahåndtering kan have påvirket træning og optimering af modellen
- data bliver opdelt til træning, validering og test
- opdelingen af data sker med udgangspunkt i modellens formål.

7. Performance og robusthed

Machine learning-modeller kan i visse tilfælde forbedre resultater og øge præcisionen i forhold til mere traditionelle statistiske modeller. Forbedringen kan indebære mere præcise resultater, hurtigere resultater eller belysning af sammenhænge, der ikke tidligere var kendt. Med et fælles begreb kan disse forbedringer i forhold til eksisterende modeller kaldes *performance*. Begrebet performance dækker over modellens evne til korrekt at estimere en ønsket parameter. Det er centralt for god brug af superviseret machine learning, at en model performer godt, men det er i lige så høj grad centralt, at modellen er robust.

En model bør, udover at estimere korrekt, være robust overfor ændringer i data og andre udefrakommende påvirkninger. Modellen bør bl.a. kunne modstå ondsindede aktørers forsøg på at påvirke modellens udfald. Det kunne eksempelvis være en bruger, som forsøger at snyde en model – direkte eller ved at ændre detaljer i inputdata. Finansielle virksomheder bør overveje risikoen for misbrug, når de benytter machine learning.

Robusthed omfatter også modellens evne til at håndtere ændringer i den virkelighed, som modellen forsøger at beskrive, dvs. modellens evne til at håndtere opdaterede data. Denne egenskab skal sikre, at modellen er tilstrækkeligt konsistent over tid. Output fra en model, der er tilstrækkelig robust, bør derved som udgangspunkt ikke ændre sig væsentligt fra én modelversion til den næste, eller hvis der testes på data fra forskellige tidsperioder. Virksomheden bør derfor sikre, at versionering af konkrete modeller bliver dokumenteret og gemt. Dette skal gøre det muligt at genskabe tidligere resultater for

derved at undersøge, hvorfor en eventuel markant ændring i modellen er indtruffet.

Der kan opstå situationer, hvor det er nødvendigt at afveje forholdet mellem performance og robusthed. I en sådan situation bør virksomheden i særlig grad overveje, hvordan og på hvilken baggrund afvejningen mellem performance og robusthed sker. Virksomheden bør også kunne dokumentere, hvorfor den har truffet beslutningen.

I forhold til ovenstående afvejning er det vigtigt, at virksomheden bruger det domænekendskab, dvs. dybdegående kendskab til virksomhedens produkter, tjenester og målgrupper, der allerede er opbygget i virksomheden. Rammer en model i en periode korrekt (høj performance), uden at det kan forklares hvorfor, kan det være et tegn på, at modellen har en ringe robusthed.

God praksis for brug af superviseret machine learning indebærer, at:

- virksomheden har forholdt sig aktivt til modellens robusthed
- virksomheden har en dokumenteret tilgang til versionering af sine modeller
- virksomheden overvejer risikoen for misbrug
- virksomheden kan dokumentere eventuelle afvejninger mellem performance og robusthed.

8. Ansvarlighed (accountability)

Virksomheden bør træffe beslutninger på baggrund af resultater fra en machine learning-model i samme kontekst, som den træffer andre beslutninger. En sådan beslutning vil dermed som udgangspunkt skulle leve op til samme krav som en beslutning, virksomheden har truffet på traditionel vis. For en finansiel virksomhed betyder det bl.a., at den bør træffe beslutning på samme ledelsesniveau, hvad enten en machine learning-model har været involveret i processen eller ej.

Ansvarlighed indebærer derfor, at den ledelsesmæssige godkendelse af en model og brugen af den ligger på det niveau i virksomheden, der i forvejen har ansvaret for den konkrete aktivitet, som modellen vedrører. Det formelle ansvar for machine learning-modeller i en virksomhed bør altid være forankret på et passende højt ledelsesmæssigt niveau.

God praksis for brug af superviseret machine learning indebærer, at:

- virksomheden sikrer, at modeller er godkendt på et tilstrækkeligt højt ledelsesniveau
- ansvaret for brugen af machine learning er veldefineret.

9. Forklarlighed (explainability)

Machine learning indebærer på nuværende tidspunkt processer og resultater, som kan være sværere at forklare end klassiske statistiske metoder. Det medfører, at virksomheden bør overveje, hvilke metoder den kan bruge for at sikre, at den benytter teknologien på en betryggende måde.

Forklarlighed indebærer, at virksomheden kan forklare og forstå, hvorfor en machine learning-model har produceret et givent resultat. Det kan eksempelvis ske ved hjælp af forskellige statistiske eller matematiske værktøjer. Forklarlighed er en væsentlig forudsætning for, at brug af en machine learning-model kan kontrolleres, og er afgørende for, at en model kan bruges på en betryggende måde.

I forbindelse med modellens vedligehold bør virksomheden løbende vurdere risikoen for uhensigtsmæssige udfald og effekter. Virksomheden bør derfor i sine procedurer for modellens udvikling, vedligehold og brug eksempelvis inkludere følsomhedsanalyser af modellens komponenter og vurdere, om vægtningen af de væsentligste elementer giver mening. En model bør som hovedregel give resultater, som er intuitive og konsistente med økonomisk teori.

Virksomheden bør overveje, hvordan den løbende forholder sig til modellens resultater, så resultaterne kan bidrage til virksomhedens beslutninger på en betryggende måde. Da modeller kan have mange forskellige udformninger og anvendelsesformer, er det ikke muligt at fastsætte en udtømmende liste over, hvilke værktøjer, metoder eller lignende der vil forbedre en models forklarlighed. Som minimum bør virksomheden dog bruge de industristandarder for værktøjer til forklarlighed af machine learning-modeller, som til enhver tid er fremherskende. Desuden bør virksomheden bruge velkendte statistiske metoder, f.eks. back-tests af resultater og følsomhedsanalyser, til at vurdere parametres modelvægte.

I den forbindelse er det desuden væsentligt at forholde sig til modellens formål. De værktøjer, virksomheden kan bruge til at forklare en models resultater, kan være forskellige, alt efter hvilket formål modellen har. Er det eksempelvis ikke muligt at vurdere, om en model lægger vægt på de mest logiske input, bør virksomheden kunne forklare, hvorfor modellen alligevel kan bruges til at opfylde formålet. En høj grad af forklarlighed vil især være aktuel, hvis machine learning bliver brugt på aktiviteter, som er underlagt den finansielle regulering, eller til aktiviteter, der har direkte konsekvens for forbrugere.

En model, som derudover ligger til grund for beslutninger, der påvirker enkeltpersoner, skal gøre det muligt for de berørte personer at få forklaret grundlaget for denne beslutning.

God praksis for brug af superviseret machine learning indebærer, at:

- virksomheden kan forklare, hvordan en model fungerer, og hvad der særligt ligger til grund for dens resultater
- forklaringer bør være tilgængelige for relevante interessenter, inklusive virksomhedens egen ledelse
- værktøjer til forklarlighed bliver benyttet i relevant omfang, og at virksomheden f.eks. kan vise, hvilke komponenter der tillægges størst vægt for konkrete udfald af modellen
- enkeltpersoner kan få forklaret grundlaget for en beslutning, der påvirker dem hver især.

10. Dataetik, skævhed i data (bias) og rimelighed (fairness)

Virksomheder må nødvendigvis forholde sig til dataetik som supplement til den teknologiske udformning og tilgang til machine learning[3].

Dataetik er ansvarlig brug af data. Det bygger på principperne i bl.a. persondatalovgivningen, som omfatter både persondataforordningen og databeskyttelsesloven. Den finansielle lovgivning indeholder endnu ikke udtrykkelige krav til dataetik, der rækker videre end de generelle god skik-regler. En virksomhed bør ikke desto mindre gøre sig en række etiske overvejelser både inden og undervejs i forløbet. Forkerte beslutninger på det etiske område kan påføre virksomheden omkostninger og indebære væsentlige operationelle risici, særligt såsom omdømmerisici. Desuden er der stigende fokus på dataetik både nationalt og internationalt.

I det omfang, virksomheden behandler personoplysninger, skal den ansvarlige for personoplysningerne (den dataansvarlige) være opmærksom på at efterleve reglerne i persondataforordningen. Dette indbefatter også principperne for behandling af personoplysninger oplistet i artikel 5. Den dataansvarlige er ansvarlig for og skal kunne dokumentere, at principperne er overholdt.

Et væsentligt område indenfor dataetik er problemer med skævhed i data (herefter bias). Bias kan have mange kilder og kan give anledning til uhensigtsmæssige udfald af modellen. Det er derfor vigtigt, at en virksomhed, der benytter machine learning, forholder sig aktivt til, hvordan den kan mindske effekten af bias i modellens tilblivelse og brug.

Bias kan eksempelvis opstå fra data, som indeholder variable, der anses for diskriminerende, såsom køn eller etnicitet. Bias kan også opstå indirekte ved interaktioner mellem flere variable, som ikke i sig selv er diskriminerende.

[3] VLAK-regeringens faktaark om dataetiske tiltag for erhvervslivet, 29. januar 2019: https://em.dk/media/12932/faktaark_dataetiske-initiativer.pdf, og anbefalinger fra Ekspertgruppen om dataetik, november 2018: https://em.dk/media/12191/ekspertgruppens-afrapportering-inkl-anbefalinger.pdf

Sidstnævnte situation kan være svær at teste statistisk, og udvikling af modellen og evaluering af resultater bør derfor i væsentlig grad involvere eksperter med domænekendskab om det givne emne. Bias bør identificeres og fjernes i det omfang, det er muligt.

Udviklingsprocessen bør inddrage eksperter både på teknologisiden og på forretningssiden af virksomheden for at sikre forankring i virksomheden generelt og for at sikre, at flere forskellige typer af eksperter har gennemgået modellen med henblik på at identificere eventuelle uhensigtsmæssige udfald.

Udover bias er det væsentligt for virksomheden at forholde sig til rimelighed. Rimelighed dækker her over den til enhver tid gældende forståelse i samfundet af, hvad der er rigtigt og forkert. Rimelighed er altså et kulturelt og flydende begreb. En model kan dermed være fri for bias fra både datakilder og modeludvikling, men alligevel lede til udfald, som vurderes ikke at være fair overfor eksempelvis bestemte kundesegmenter. Virksomheden bør kunne dokumentere, hvordan den har forholdt sig til rimelighed.

God praksis for brug af superviseret machine learning indebærer, at:

- virksomheden har forholdt sig aktivt til bias og til, hvordan risikoen for uhensigtsmæssige udfald kan minimeres, og kan dokumentere dette
- virksomheden har forholdt sig aktivt til rimelighed og kan dokumentere dette.

11. Gennemsigtighed (transparency)

Brug af machine learning kan indebære en risiko for, at virksomheden træffer beslutninger på et grundlag, der er svært at efterprøve. Det uklare grundlag kan medføre utryghed hos dem, som er genstand for virksomhedens beslutning. Uklarhed kan desuden skjule fejl i modellen, der har potentiale til negativ effekt på en større gruppe modtagere, end enkelte sagsbehandlerfejl kan have.

En virksomhed, som arbejder med machine learning, bør derfor forholde sig til, hvordan den informerer sine interessenter om dette. Det bliver i stigende grad nødvendigt, når en model har effekt på beslutninger, som direkte eller indirekte påvirker enkeltpersoner, særligt forbrugere. Virksomheden skal være opmærksom på reglerne i artikel 22 i persondataforordningen om automatiske individuelle afgørelser, herunder profilering[4].

[4] Artikel 22 i Europa-Parlamentets og Rådets forordning (EU) 2016/679 af 27. april 2016 om beskyttelse af fysiske personer i forbindelse med behandling af personoplysninger og om fri udveksling af sådanne oplysninger og om ophævelse af direktiv 95/46/EF (generel forordning om databeskyttelse)

Det skal være muligt for den enkelte forbruger at orientere sig om, hvilke processer der kan forventes at ligge til grund for behandling af vedkommendes sager med pågældende virksomhed.

God praksis for brug af superviseret machine learning indebærer, at:

- virksomheden stiller tilstrækkelig information om sin brug af machine learning til rådighed for de berørte parter
- enkeltpersoner kan få indblik i, hvilke processer der ligger til grund for behandling af deres sager.

giver den registrerede ret til ikke at være genstand for en afgørelse, der alene er baseret på automatisk behandling, herunder profilering, som har retsvirkning eller på tilsvarende vis betydeligt påvirker den pågældende. Dette gælder f.eks. ikke, hvis det følger af en kontrakt, eller den registrerede giver samtykke.

Bilag 1

Samlede anbefalinger

Formål

God praksis for brug af superviseret machine learning indebærer, at:

- virksomheden har beskrevet formålet med modellen klart, så den senere kan træffe beslutninger på et fyldestgørende grundlag
- beskrivelsen af modellen understøtter en tilstrækkelig forståelse på alle relevante niveauer i organisationen.

Governance

God praksis for brug af superviseret machine learning indebærer, at:

- brugen indgår med de nødvendige tilpasninger i virksomhedens sædvanlige governance
- virksomheden dokumenterer og logger valg og fravalg igennem modellens livscyklus, så modellens historik er sikret
- modellens begrænsninger er grundigt beskrevet, også i forhold til datakvalitet.

Datahåndtering

God praksis for brug af superviseret machine learning indebærer, at:

- virksomheden afdækker sine databehov
- virksomheden sikrer, at datakvalitet og stabilitet i levering kan opretholdes på et tilfredsstillende niveau
- virksomheden dokumenterer sin proces for håndtering af data.

Træning af modellen

God praksis for brug af superviseret machine learning indebærer, at:

- virksomheden tilrettelægger træning af modellen, så modellens formål bedst muligt understøttes, herunder valg af optimering
- virksomheden kan beskrive, om datahåndtering kan have påvirket træning og optimering af modellen
- data bliver opdelt til træning, validering og test
- opdelingen af data sker med udgangspunkt i modellens formål.

Performance og robusthed

God praksis for brug af superviseret machine learning indebærer, at:

- virksomheden har forholdt sig aktivt til modellens robusthed
- virksomheden har en dokumenteret tilgang til versionering af sine modeller
- virksomheden overvejer risikoen for misbrug
- virksomheden kan dokumentere eventuelle afvejninger mellem performance og robusthed.

Ansvarlighed

God praksis for brug af superviseret machine learning indebærer, at:

- virksomheden sikrer, at modeller er godkendt på et tilstrækkeligt højt ledelsesniveau
- ansvaret for brugen af machine learning er veldefineret.

Forklarlighed

God praksis for brug af superviseret machine learning indebærer, at:

- virksomheden kan forklare, hvordan en model fungerer, og hvad der særligt ligger til grund for dens resultater
- forklaringer bør være tilgængelige for relevante interessenter, inklusive virksomhedens egen ledelse
- værktøjer til forklarlighed bliver benyttet i relevant omfang, og at virksomheden f.eks. kan vise, hvilke komponenter der tillægges størst vægt for konkrete udfald af modellen
- enkeltpersoner kan få forklaret grundlaget for en beslutning, der påvirker dem hver især.

Dataetik

God praksis for brug af superviseret machine learning indebærer, at:

- virksomheden har forholdt sig aktivt til bias og til, hvordan risikoen for uhensigtsmæssige udfald kan minimeres, og kan dokumentere dette
- virksomheden har forholdt sig aktivt til rimelighed og kan dokumentere dette.

Gennemsigtighed

God praksis for brug af superviseret machine learning indebærer, at:

- virksomheden stiller tilstrækkelig information om sin brug af machine learning til rådighed for de berørte parter
- enkeltpersoner kan få indblik i, hvilke processer der ligger til grund for behandling af deres sager.