
Refleksioner over portfolio 1 -3 i form af læringsudbytte, fremtidig anvendelse i jeres studier eller i praksis mv

Overordnet har kurset været meget praksisorienteret, hvor der undervejs har været mulighed for at arbejde med konkrete øvelser. Gennem kurset har øvelserne dannet grundlag for løsningen af de i alt tre portfolio opgaver. Dermed har kurset været præget af en "learn by doing"-tilgang, og har i sidste ende givet viden om og kompetencer i, hvordan data kan bearbejdes og analyseres ved hjælp af Python.

I besvarelsen af portfolio 1 har læringsudbyttet hovedsageligt bestået af, hvordan man med Python kan udtrække deskriptiv statistik ud af et afgrænset datasæt. For at kunne trække statistikkerne frem, har en række af Pythons regular expressions været benyttet som fremgangsmetode. Formålet med den deskriptive statistik har været at give et statistisk billede af aspekter fra Titanic som passagerfordeling, antal overlevende, antal omkomne og inddeling på de forskellige passagerklasser. I besvarelsen af portfolio 2 var der et større fokus på text mining, hvordan man arbejder med et API, samt hvordan APIet kan afgrænses, og hentes ned på egen computer. I forlængelse af afgrænsningen var et af det største læringsudbytte, hvordan man i arbejdet med data kan være nødsaget til at rense den pågældende data før en text mining kan påbegyndes. Ønsker man eksempelvis at anviser de 10 mest benyttede ord i en række tekster, men har ikke rensset ens data, får man intetsigende resultater som; og, at, der, som etc. Har man derimod rensset dataen, får man, som i tilfældet med en af Guardian-artiklerne, resultater som; world, cup, players, England og team. I besvarelsen af portfolio 3 blev der igen arbejdet med et API – her bestod læringsudbyttet blandt andet af at hente data ned ud fra en selvsammensat URL ved hjælp af et API. Her var samtidig et fokus på metadata, og opgaven har givet anledning til videre overvejelse af, hvilken metadata, der er relevant at have med i eksempelvis et datasæt. Samtidig gav opgaven mulighed for at undersøge, hvorfor en institution som SMK ønsker at digitalisere sin samling, samt hvilke fordele og udfordringer et sådant ønske medfører.

Etiske overvejelser

I arbejdet med data vil der være en større sandsynlighed for, at man kommer i berøring med data, som kan kategoriseres som værende personfølsomme. Derfor har man et etisk ansvar for, at den data bliver behandlet forsvarligt. I arbejdet med Titanic datasættet vil dataen, af Datatilsynet, betegnes som værende "almindelig persondata", som dækker over; navne, adresser og e-mails

(”Hvad er personoplysninger?” u.å.). Datasættet indeholder blandt andet navne på passagerer ombord, men denne data er altså ikke personfølsom jævnfør Datatilsynet. Kigger man i stedet mod, hvem der høj grad beskæftiger sig med personfølsomme data, vil man eksempelvis få øje på forskere inden for det sundhedsvidenskabelige område. Disse forskere vil i forbindelse med deres empiriske forsøg komme i besiddelse af indsamlet data som; cpr-nr., sygdomshistorik og anden genetisk data, som af Datatilsynet kategoriseres som personfølsom. Denne data vil som selvfølge blive skrevet ned og dermed gjort tekstbaseret, og herefter har forskerne et etiske ansvar for, at disse oplysninger bliver opbevaret forsvarligt og sikkert. Københavns Universitet har eksempelvis en række regler for god videnskabelig praksis, hvor universitet har vedtaget, at deres forskere blandt andet skal vise hensyn over for; ”*de personer eller det objekt, som forskningen vedrører, eller som den får betydning for,*” (”Københavns Universitets regler om god videnskabelig praksis”, 2007, s. 1). Forskernes etisk ansvar følger med, når de endvidere publicerer deres resultater. Ansvarer består i at resultaterne præsenteres, som de nu engang foreligger, og at alle involverede parter bliver deklareret. I Danmark har vi for nyligt set et tilfælde, hvor en forskningsrapport inden for fødevareområdet, ikke er blevet deklareret på ordentlig vis. Problematikken bestod i, at en interesseorganisation havde påvirket rapporten og dens resultater ved at forfatte afsnit i den, hvor egen agenda blev fremmet. Dette er et eksempel, hvor god videnskabelig praksis, samt det etiske ansvar, ikke er blevet overholdt.

I forhold til opbevaringen af den personfølsomme data – tager man eksempelvis et offentligt sundhedssystem indeholdende data om borgernes sygdomshistorik - har de, der udvikler systemet, ligesom forskerne, et etisk ansvar for, at systemet ikke kan hackes og at dataen ikke kan tilgås af uvedkommende. Association for Computing Machinery (ACM) har udviklet en række etiske retningslinjer – Code of Ethics - som man i arbejdet med computere/systemer bør rette sig efter. I disse retningslinjer finder man blandt andet et afsnit, der skriver, at; ”*Computing professionals should perform due diligence to ensure the system functions as intended, and take appropriate action to secure resources against accidental and intentional misuse, modification, and denial of service.*” (”ACM Code of Ethics”, 2018). Hvilket netop pointerer et ansvar for sikker dataopbevaring, og at systemet ikke misbruges – bevidst som ubevidst.

Litteraturliste

Hvad er personoplysninger? (u.å.). Lokaliseret den 3. januar 2020. Fra:

<https://www.datatilsynet.dk/generelt-om-databeskyttelse/hvad-er-personoplysninger/>

Københavns Universitets regler om god videnskabelig praksis. (1. september 2007). Lokaliseret den 3. januar 2020. Fra:

https://praksisudvalget.ku.dk/regler_vejl/portal/pu_regler_1_.pdf

ACM Code of Ethics and Professional Conduct. (u.å.). Lokaliseret den 3. januar 2020. Fra:

<https://www.acm.org/code-of-ethics>