

Refleksjonsnotat

Refleksjoner over læring og ferdigheter

Gjennom dette prosjektet har vi lært masse nytt og tilegnet oss ferdigheter gjennom prøving og feiling. Vi har fått erfaring med hele prosessen fra datainnsamling til visualisering. Vi har lært å hente inn miljødata fra åpne API-er, og hvordan man må tolke strukturen til disse og eventuelle feilmeldinger som oppstår. Vi har også lært hvordan man jobber med CSV- og JSON-filer, og at koden må tilpasses etter hvordan dataene er organisert.

For å rense og strukturere dataene laget vi funksjoner som håndterer feil vi mente kunne finnes i datasettene. Vi brukte Pandas, NumPy og pandasql til å analysere dataene, og valgte statistiske mål som gjennomsnitt og median for å kunne sammenligne utvikling og sammenhenger i ulike datasett.

For å visualisere resultatene brukte vi med Matplotlib, Seaborn og Scikit-learn, og testet også interaktive løsninger med Plotly og Bokeh. Her fikk vi noen problemer, men dette ble løst ved å bruke alternative verktøy. Vi har laget en «data-app» med ipywidgets, som har gitt oss masse nyttig erfaring med brukerinteraksjon og logisk styring av visninger basert på valg.

Vi har jobbet strukturert med Git og GitHub, og fått mye erfaring ved å bruke tags, branches, merge og skrive tydelige commit-meldinger. Dette var noe ingen av oss hadde gjort før, så det var veldig nyttig å lære seg. Vi har lært hvordan man skal organisere et prosjekt i ulike mapper, og hvorfor det er viktig å bruke .gitignore og requirements.txt. Vi har også fått erfaring med hvordan man dokumenterer prosjektet i README.md, og lært å skrive tester ved hjelp av unittest.

Identifisering av spesifikke utfordringer

Johanne:

En av de første store utfordringene jeg fikk i prosjektet var å hente historisk temperaturdata fra Frost API. I starten var målet å hente ut temperaturdata hvert minutt i hele 2024. Vi prøvde å hente ut alt i en forespørsel, men fikk feilmelding (statuskode 400). Dette kom av at det er begrenset hvor mye data API-et returnerer per spørring. Jeg løste problemet ved å lage en for-løkke som hentet data måned for måned. Da vi utvidet prosjektet til å hente ut temperaturdata for de siste 50 årene, fungerte løsningen fortsatt som ønsket.

En annen utfordring var utviklingen av «data-appen». Målet var å gjøre analysene mer tilgjengelige for en bruker ved å bruke widgets. Dette var litt mer komplisert enn forventet. Her var en av de største utfordringene å få riktig informasjon til å vises basert på de valgene brukeren tok. For eksempel når brukeren velger "klimagass", må det også velges et sted, og deretter skal det kun vises relevante valg for det valgte stedet. Hvis man velger "Norge", skal det kun vises mulig plott som er laget for dette datasettet. Det ble mye arbeid rundt dette med logikk for valg og visning, men jeg er veldig fornøyd med hvordan resultatet ble.

Jeg fikk derimot ikke til å visualisere data gjennom Bokeh-plott i «appen». Disse plottene ville ikke lastes inn og hang seg på "Broks loader". Løsningen her ble å lage tilsvarende visualiseringer med Matplotlib.

Dette er noen av de utfordringene jeg har lært mest av gjennom prosjektet. Jeg har lært mye av å tolke feilene som har oppstått og finne gode løsninger for å fikse det. Noen ganger har jeg måtte fikse tekniske problemer, andre ganger har jeg tilpasse eller endre den opprinnelige planen.

Idunn:

Det var også utfordringer knyttet til datakvaliteten på luftkvaliteten. Vi ønsket å hente ut data fra de siste 50 årene for å kunne sammenligne det med temperaturen, men fant ikke fullstendige målinger fra mer enn 20 år tilbake i tid. Vi brukte NILU, som vi så på som den mest pålitelige kilden, men likevel var det veldig mange manglende verdier og store hull. På bakgrunn av dette valgte vi å filtrere ut alle tomme rader for å sikre at kun gyldige datapunkter ble inkludert i analysene våre. I et stort datasett ble dette flere hundre linjer som ble filtrert ut. Dette har gitt god erfaring med håndtering av manglende verdier, men vi er likevel klar over at dette kan ha vært en betydelig feilkilde.

Underveis i prosjektet har jeg også hatt en del trøbbel med Git-kommandoene i min VSCode, spesielt med pulling og pushing. Det har ført til at programmet har hengt seg ordentlig opp og jeg har slitt med å finne en løsning. Det var spesielt en hendelse der den ene Notebooken hadde hengt seg opp og lå som om jeg hadde gjort en endring, selv om det kun var metadata som hadde hengt seg. Det var spesielt frustrerende siden jeg kom ingen vei og heller ikke fikk jobbet mer med annen kode. Istedenfor å spørre om hjelp begynte jeg da å rote mye med kommandoer jeg egentlig ikke forsto for å prøve å løse det, heller enn å spørre om hjelp for å ikke forstyrre andre. Det gjorde bare vondt verre, og jeg endte opp med å gjøre om på eksisterende kode ved uhell, noe som tok mye lenger tid å rette opp i og skapte enda mer merarbeid for andre på gruppa. Heldigvis fikk jeg god hjelp og det løste seg til slutt, men jeg har lært mye viktig om samarbeid i prosessen, at det er viktig å spille på de ressursene man har og at det er lov (og lurt) å spørre om hjelp.

Victoria:

I den prediktive analysen møtte jeg på endel utfordringer i forhold til det å bare lage en god regresjonsmodell. Den største og egentlig den som gjentakende kom tilbake var utfordringen i å tolke og vurdere om regresjonsmodellen jeg har laget faktisk er god eller ikke. Jeg lagde til sammen tre prediktive analyser, hvor kun én endte opp i det endelige resultatet. Først kom vi fram til at det var vanskelig å lage en ordentlig regresjon med datasettene vi hadde da, da de kun gikk et år tilbake, så dette førte til at vi hentet mer data som gjorde en analyse enklere. Jeg lærte da at jeg måtte være mer kritisk til hvilket data vi bruker i en slik regresjonsanalyse, da det vil være nærmere nytteløst å lage en analyse med for lite data.

En annen utfordring var det å tolke resultatene som kom fram i analysene. Hvis R^2 verdien var for lav, som de var i de to første analysene, så måtte jeg vurdere om det egentlig var en nyttig modell eller om jeg skulle prøve meg på en ny. Endte opp med å gjøre sistnevnte og det ble enklere for hver gang.

Sist skulle analysen i utgangspunktet knyttes opp mot Paris-avtalen til to-graders mål. Med dette måtte jeg da kunne gjøre fremtidige prediksjoner og ta det opp mot hvilket CO₂-ekvivalent-utslippsnivå som var nødvendig. Plutselig måtte jeg også kunne gjøre regresjonsformelen baklengs for å få gjort prediksjonen riktig, så det gikk endel arbeid inn i bare å forstå oppsettet og hvordan de ulike stegene påvirker hverandre.

Jeg har generelt følt på mye mestring når jeg har laget visualiseringer gjennom analysen, og da kunne vist fram hva jeg har klart å komme frem til. Det har vært morsomt og lærerikt. Det å lage den lineære regresjonen har vært arbeid i seg selv, men ikke bydd på mer utfordringer enn de normale som man bare må lære seg.

Refleksjoner over samarbeidet i gruppen

Samarbeidet i gruppen har stort sett fungert godt. En av utfordringene i starten var å finne felles tidspunkter for å møtes, siden ikke alle går samme studieretning. Likevel tok alle ansvar for sine deler og selv om vi ikke alltid fikk jobbet sammen fysisk hadde vi flere møter over nett der vi kunne diskutere prosjektet og eventuelle utfordringer vi hadde.

I del 1 av prosjektet fordelte vi oppgavene slik at hver av oss jobbet med innhenting og rensing av hvert vårt datasett (temperatur, luftkvalitet og klimagassutslipp). Dette gjorde det lett å komme i gang, men det tok tid å sette sammen alle delene til en helhet før innleveringen.

I del 2 endret vi derfor på denne ansvarsfordelingen. I stedet for å jobbe ut fra hvert vårt datasett, fordelte vi ansvaret for ulike oppgaver (oppgave 4, 5 og 6). Dette gjorde det lettere å jobbe parallelt og hjelpe hverandre fordi alle hadde erfaring med de ulike datasettene. Det gjorde at vi fikk en bedre sammenheng i prosjektet.

Det var litt rot i starten når det gjaldt hvordan vi brukte branches og håndterte commits. Hvis vi skulle gjort et lignende prosjekt igjen, ville vi vært mer bevisste på hvordan vi strukturerer Git-repositoriet, hva de ulike branchene brukes til, og at alle commits er tydelige. Dette kunne gjort at samarbeidet ble mer effektivt fra start.

Når det gjelder prosjektledelse, kunne vi også vært flinkere til å lage en tydelig arbeidsplan tidlig i prosjektet. Det tok litt tid før vi kom ordentlig i gang med del 1, og det gjorde at mye måtte gjøres på kort tid før innlevering. Dette ble derimot mye bedre i del 2, hvor vi jobbet mer strukturert og fordelte arbeidet jevnt utover hele perioden.

Vurdering av de endelige resultatene

Vi er veldig fornøyde med det endelige resultatet av prosjektet. Vi fikk hentet inn miljødata fra flere relevante kilder og analysert dem slik at vi fikk oversikt over utvikling og sammenhenger. Visualiseringene vi lagde i notebooken ble oversiktlige, og gjorde det lettere å se mønstre i dataene.

Vi er også fornøyde med hvordan vi klarte å knytte analysene til problemstillingen og målene i oppgaven. Vi fikk brukt statistiske metoder og sentralmål, og fikk vist hvordan forskjellige faktorer kan påvirke miljø og klima.

«Appen» er også veldig stolte av. Her fikk vi knyttet neste all dataen sammen på et sted. Selv om vi hadde noen utfordringer underveis (som krav til API-nøkkel og problemer med enkelte visualiseringer), fikk vi likevel laget en løsning som viser hvordan appen fungerer i praksis, og hvordan den lar brukeren utforske dataene på egen hånd.

Ideer til hvordan prosjektet kan forbedres i fremtiden

Dersom vi skulle videreutviklet prosjektet, er det flere ting vi kunne forbedret og utvidet. For det første kunne vi gjort sanntidsdata-funksjonaliteten i appen bedre. Vi kunne laget en løsning for sikker håndtering av API-nøkler, ved for eksempel bruk av miljøvariabler sammen med et GUI for nøkkelregistrering. Dette ville gjort appen mer fleksibel og aktuell.

Vi kunne også hentet inn data fra andre steder i verden og gjort det mulig for brukeren å hente og sammenligne data fra ulike områder. Da kunne vi laget et kartbasert grensesnitt der brukeren kunne velge sted. Vi har også tenkt at det ville vært bra å lage appen som en webapplikasjon ved hjelp av Dash eller Streamlit, slik at den fungerer utenfor Jupyter Notebook.

En annen forbedring kunne være å bruke mer avanserte prediktive modeller, som random forest eller ridge regression, for å sammenligne ytelse og gjøre analysene bedre. Til slutt kunne vi også hatt flere tester. For eksempel tester som sjekker at visualiseringen og de interaktive elementer fungerer som de skal.

Oppsummering og personlige refleksjoner

Prosjektet har gitt oss mye nyttig erfaring, både faglig og praktisk. Vi har jobbet med mye data, laget løsninger som fungerer, og lært hvordan man jobber sammen på en strukturert måten med kode.

Det viktigste vi sitter igjen med, er følelsen av mestring. Tidligere har programmering vært vanskelig, men i dette prosjektet har vi opplevd at det kan være gøy og meningsfullt. Vi har lært masse, og vet at vi vil få bruk for det videre uansett hva vi ender opp med å jobbe med.