

## Refleksjonsnotat

Ærede sensorer som tar seg tid å lese dette. De siste månedene har vært krevende, men veldig lærerike. Før vi satte i gang med dette prosjektet hadde vi lite erfaring om Git og versjonshåndtering, Pandas, VSCode, Scipy og andre biblioteker. Vi slet i starten med struktur på prosjektet, i tillegg til blant annet SSH-nøkkel. Vi opplevde at «merge»-konflikter oppsto, i tillegg til at vi ikke helt visste hvordan vi skulle gå fram med bruk av flere grener, i stedet for å kun jobbe i main. Dette har endret seg i stor grad.

Vi opprettet hver vår gren som vi jobbet aktivt i for så å oppdatere ved å pulle og merge mellom main og grenene. Vi fordelte oppgavene jevnt mellom oss, og samarbeidet godt med oppgavene. Vår dårlige erfaring med versjonshåndtering førte til at vi i starten «commita» mye oftere enn nødvendig, og det vi commita skulle ikke nødvendigvis blitt commita. For eksempel så sjekket vi inn venv. I tillegg til dette finnes det enda flere forbedringer, da vi for sikkerhets skyld har valgt å sjekke inn både pycache og .env (som har med API-nøkkelen å gjøre, som sagt kunne dette vært gjort bedre) for å være sikker på at koden fungerer (vi er klar over at vi «egentlig» ikke skulle gjort dette, men dette er ikke noe vi får gjort noe med akkurat nå, og får bare ta med oss dette videre som god erfaring). «Ikke fiks noe som ikke er ødelagt».

Vi har i løpet av dette prosjektet lært å for eksempel hente inn data med `requests.get()` for så å behandle det videre som JSON- og CSV-filer med bruk av blant annet Pandas og Pandas SQL. Disse bibliotekene hjalp oss med å blant annet håndtere mangel på verdier, samt opprette en dataframe. På den måten tok vi på oss en større utfordring ved å håndtere to ulike typer datafiler. Det var essensielt å forstå datastrukturen for å jobbe videre med det, derfor lagde vi flere hjelpefunksjoner for dette. Videre utførte vi ulike former for visualiseringer og analyse av dataene.

Til visualiseringene brukte vi i hovedsak Seaborn og Matplotlib. Seaborn ga et godt utvalg av kreative plots, som for eksempel heatmap. En viktig del av analysen er lineær regresjonsanalyse. Vi valgte å vise det på to ulike måter. Den ene var ved hjelp av Scikit-learn, og den andre var ved hjelp av en metode vi nylig har lært om, «minste kvadraters metode». Her var Numpy-biblioteket essensielt, da det hjalp til å utføre multiplikasjon av matriser og vektorer, samt løse matriseligninger.

Når det kommer til selve prosjekt-oppgaven synes vi at det er veldig interessant og nyttig, særlig med tanke på erfaringer rundt det å håndtere store mengder data. På den måten kan vi absolutt få bruk for den tekniske erfaringen som kommer fra dette prosjektet. Vi kan i tillegg

se for oss at denne typen arbeidsoppgave kan hjelpe for framtidig forskning, samt observere mønster mellom ulike målvariabler over tidsperioder. Når det kommer til forbedringer føler vi at prosjektet kunne vært litt mer realfaglig rettet, der man kunne brukt tidligere matematisk erfaring med blant annet derivasjon og integrasjon.

For å konkludere er vi som gruppe alt i alt fornøyd med dette prosjektet. Det har vært svært lærerikt, utfordrende, og givende. Vi har blitt kjent med mange nye metoder som blir nyttig videre. Erfaringene våre fra dette prosjektet kan i tillegg være til god hjelp i framtidige studier og yrkesliv der programmering er nødvendig for å håndtere store mengder informasjon. Til sist kan man bruke koden til å predikere framtidig data, i denne sammenhengen miljødata, men andre typer analyser vil også være mulige å gjennomføre.